Philosophy of Linguistics
Handbook of Philosophy of Science
Volume 14
Philosophy of Linguistics

Edited by

Ruth Kempson
King's College London, UK

Tim Fernando
Trinity College Dublin, Ireland

Nicholas Asher
CNRS Laboratoire IRIT,
Université Paul Sabatier, France

North Holland is an imprint of Elsevier
GENERAL PREFACE

Dov Gabbay, Paul Thagard, and John Woods

Whenever science operates at the cutting edge of what is known, it invariably runs into philosophical issues about the nature of knowledge and reality. Scientific controversies raise such questions as the relation of theory and experiment, the nature of explanation, and the extent to which science can approximate to the truth. Within particular sciences, special concerns arise about what exists and how it can be known, for example in physics about the nature of space and time, and in psychology about the nature of consciousness. Hence the philosophy of science is an essential part of the scientific investigation of the world.

In recent decades, philosophy of science has become an increasingly central part of philosophy in general. Although there are still philosophers who think that theories of knowledge and reality can be developed by pure reflection, much current philosophical work finds it necessary and valuable to take into account relevant scientific findings. For example, the philosophy of mind is now closely tied to empirical psychology, and political theory often intersects with economics. Thus philosophy of science provides a valuable bridge between philosophical and scientific inquiry.

More and more, the philosophy of science concerns itself not just with general issues about the nature and validity of science, but especially with particular issues that arise in specific sciences. Accordingly, we have organized this Handbook into many volumes reflecting the full range of current research in the philosophy of science. We invited volume editors who are fully involved in the specific sciences, and are delighted that they have solicited contributions by scientifically-informed philosophers and (in a few cases) philosophically-informed scientists. The result is the most comprehensive review ever provided of the philosophy of science.

Here are the volumes in the Handbook:

Philosophy of Science: Focal Issues, edited by Theo Kuipers.
Philosophy of Physics, edited by Jeremy Butterfield and John Earman.
Philosophy of Biology, edited by Mohan Matthen and Christopher Stephens.
Philosophy of Mathematics, edited by Andrew Irvine.
Philosophy of Logic, edited by Dale Jacquette.

Philosophy of Information, edited by Pieter Adriaans and Johan van Benthem.

Philosophy of Technology and Engineering Sciences, edited by Anthonie Meijers.

Philosophy of Complex Systems, edited by Cliff Hooker.


Philosophy of Economics, edited by Uskali Mäki.

Philosophy of Linguistics, edited by Ruth Kempson, Tim Fernando and Nicholas Asher.

Philosophy of Anthropology and Sociology, edited by Stephen Turner and Mark Risjord.

Philosophy of Medicine, edited by Fred Gifford.

Details about the contents and publishing schedule of the volumes can be found at http://www.elsevier.com/wps/find/bookdescription.editors/BS_HPHS/description#description

As general editors, we are extremely grateful to the volume editors for arranging such a distinguished array of contributors and for managing their contributions. Production of these volumes has been a huge enterprise, and our warmest thanks go to Jane Spurr and Carol Woods for putting them together. Thanks also to Lauren Schultz and Derek Coleman at Elsevier for their support and direction.
CONTRIBUTORS

Nicholas Asher  
CNRS, IRIT, Université Paul Sabatier, France.  
asher@irit.fr

Emmon Bach  
University of Massachusetts, USA, and SOAS, UK  
ebach@linguist.umas.edu

Giosuè Baggio  
SISSA International School for Advanced Studies, Triest, Italy.  
gbaggio@sissa.it

William O. Beeman  
University of Minnesota, USA.  
wbbeeman@umn.edu

Ronnie Cann  
University of Edinburgh, UK.  
r.cann@ed.ac.uk

Philip Carr  
Université Paul Valéry, France.  
philcarrb@yahoo.com

Wynn Chao  
SOAS, UK.  
wcc@soas.ac.uk

Alexander Clark  
Royal Holloway University of London, UK.  
alexc@cs.rhul.ac.uk

Robin Cooper  
Gothenberg University, Sweden.  
cooper@ling.gu.se

Tim Fernando  
Trinity College Dublin, Eire.  
tim.fernando@cs.tcd.ie

Peter Hagoort  
Radboud Univeristy of Nijmegen and Max Planck Institute for Psycholinguistics, The Netherlands.  
peter.hagoort@fceldoners.ru.nl
Wolfram Hinzen
University of Durham, UK.
wolfram.hinzen@durham.ac.uk

James R. Hurford
University of Edinburgh, UK.
jim@ling.ed.ac.uk

Ruth Kempson
King’s College London, UK.
ruth.kempson@kcl.ac.uk

Michiel van Lambalgen
University of Amsterdam, The Netherlands.
m.vanlambalgen@uva.nl

Shalom Lappin
King’s College London, UK.
shalom.lappin@kcl.ac.uk

Howard Lasnik
University of Maryland, USA.
lasnik@umd.edu

Sally McConnell-Ginet
Cornell University, USA.
smg9@cornell.edu

Glyn Morrill
Universitat Politècnica de Catalunya, Spain.
morrill@lsi.upc.edu

Gerald Penn
University of Toronto, Canada.
gpenn@cs.toronto.edu

Jaroslav Peregrin
Academy of Sciences and University of Hradec Králové, Czech Republic.
jarda@peregrin.cz

Robert van Rooij
University of Amsterdam, The Netherlands.
r.a.m.vanrooij@uva.nl

Juan Uriagereka
University of Maryland, USA.
juan@umd.edu

Daniel Wedgwood
University of Edinburgh, UK.
wedgwood.dan@gmail.com

Contributors
EDITORIAL PREFACE

Ruth Kempson, Tim Fernando, and Nicholas Asher

Ever since the nineteen-sixties, linguistics has been a central discipline of cognitive science, feeding debates within philosophy of language, philosophy of mind, logic, psychology — studies on parsing, production, memory, and acquisition — computational linguistics, anthropology, applied linguistics, and even music. However, one diagnostic attribute of what it takes to be a natural language has been missing from articulation of grammar formalisms. Intrinsic to language is the essential sensitivity of construal of all natural language expressions to the utterance context in which they occur and the interaction with other participants in that same utterance context that this context-relativity makes possible, with rich occasion-specific effects depending on particularities of the individual participants. Given the very considerable hurdles involved in grappling with this core property of language, and the lack of suitable formal tools at the time, it is perhaps not surprising that this diagnostic property of natural languages should have been set aside as peripheral when formal modelling of language took off in the mid-nineteen-sixties. However, the methodology that was then set up, despite the welcome clarity to linguistic investigation that it initially secured, has had the effect of imposing a ceiling on the kind of explanations for what the human capacity for language amounts to.

The justification for setting aside such a core attribute of language was grounded in the point of departure for the methodologies for formal modelling of languages being explored in the fifties and early sixties. Figures such as Harris, Chomsky, Lambek, each in their different ways transformed language theorising by their commitment to articulating formal models of language [Harris, 1951; Chomsky, 1955; Lambek, 1958]. The overwhelming priority at that time was to provide a science of language meeting criteria of empirical verifiability; and context variability was not taken to be relevant to the formal specification of any system meeting such criteria. Rather, grammars were presumed to induce sets of sentences; and the first hurdle was the fact that natural languages allow for infinite variety over and above such context variability, simply because any one sentence can be indefinitely extended. This led to the assumption that the generalisations constituting explanations of language must invariably take the form of a function mapping a finite vocabulary together with a finite set of rules onto an infinite set of (grammatical) sentences. With this requirement in mind, the move made was to take the relatively well-understood formal languages of logics as the pattern to be adapted to the natural language case, since these provided a base for inducing an infinite set of strings from a finite, indeed small number of rules, in so doing assigning structure to each
such string. The so-called Chomskian revolution [Chomsky, 1965] was then to embed such linguistic theorising in a philosophy underpinning the assumptions to be made in advocating such grammars. The capacity for language was then said by Chomsky to be grounded in an ideal speaker/hearer’s competence [Chomsky, 1965], a concept articulated solely with respect to grammars whose empirical content resided in their relative success in inducing all and only the structural descriptions of wellformed strings of the language, in this too following the pattern of formal language grammars. Grammars of natural language were accordingly evaluated solely with reference to judgements of grammaticality by speakers, leaving wholly on one side the dynamics of language as used in interaction between participants. All use of corpus-based generalisations was dismissed as both insufficient and inappropriate. The data of so-called performance were set aside as irrelevant in virtue of the reported disfluency displayed in language performance, its supposed impossibility as a basis for language acquisition, and its obfuscation of the linguistic generalisations that have to be teased out from the intricate intertwining of linguistic principles with grammar-external constraints such as those imposed by memory limitations, processing cost and other constraints determining how linguistic expertise is realisable in language performance in real time.

This methodology was adopted unhesitatingly by the entire research community, irrespective of otherwise fiercely opposed frameworks. This presumption of separation between the competence system and its application in performance was indeed so strongly held that there was condemnation in principle even of defining grammar formalisms in terms relevant to their application in explaining such data. Properties of language were to be explained exclusively in terms of structural properties of grammatical and ungrammatical sentences, independently of any performance-related dynamics of what it means to process language in real time. In consequence, the only characterisation of context-dependence definable was that pertaining to sentence-internal phenomena and without any reference to phenomena not characterisable within the sententialist remit; there was no formulation of how language dependencies may be established in interaction; nor was there any characterisation of construal in terms of how such understanding might be built up in real time. All these were taken to be aspects of discourse modelling, which largely lacked any precise characterisation, or of language performance, leaving the phenomenon of context-dependence at best only partially characterised. Properties of dependency between one expression and another were, equally, taken to be explicable only in so far as these could be defined as a sentence-internal dependency, and accordingly defined structurally, and largely as a phenomenon of syntax.

The divide between the linguistic knowledge requisite for ideal speaker competence and other sources of information potentially applicable to natural language understanding became consolidated in the subsequent establishment of the concept of I-Language [Chomsky, 1986] which refers to an Internal/Individualistic/Intensional(/Innate) body of knowledge, the appropriate object of study for linguistics being taken to be a mental faculty internal to individuals that can be
legitimately studied in isolation from external factors such as communicative context, variation, processing considerations, perceptual abilities etc. An attendant concept in psycholinguistics and the philosophy of language is the *modularity* assumption for the language faculty or the concept of *input system* this being the language module mapping strings of the natural language onto a so-called *language of thought* [Fodor, 1983]. Under this view, the language module, responsible for the structural properties of natural languages, is autonomous and qualitatively different from other cognitive abilities. The crucial ingredients of modularity are *domain specificity* and *information encapsulation* which means that the module is immune from information from other non-linguistic sources.

There were exceptions to this particular variant of the sententialist orthodoxy. The exceptions came from philosophers and the “West Coast” conception of semantics pioneered by Richard Montague in the late 1960s and early 1970s. Montague considered that principles of the semantic interpretation of natural language encoded in the typed lambda calculus should explain certain dependencies,¹ but again only at a sentence internal level. Montague’s work led to the philosopher David Kaplan’s influential treatment of indexicals [Kaplan, 1980], where a device, a Kaplanian context, external to the structural properties of grammar, was responsible for the interpretation of terms like *I, you, here* and *now.* Work by Stalnaker, Thomasson and Lewis provided a semantics of variably strict conditionals according to which the interpretation of the conditional link between antecedent and consequent depended upon an ordering source, a similarity over possible points of evaluation, that was sensitive to the interpretation of the antecedent as well as the current point of evaluation [Stalnaker, 1975; Thomasson and Patel, 1975; Lewis, 1973]. At the same time the seminal work by the philosophers Saul Kripke and Hilary Putnam on proper names and natural kind terms indicated that non-linguistic contextual factors affected interpretation [Kripke, 1980; Putnam, 1975]. Another subversive current that would find its way into semantics in the 1980s in the form of dynamic semantics was happily developing in computer science. Already in the 1970s it was realised that the semantics of programs involved a transition from one machine state to another, since the idea that transitions between machine states is central to the semantics of programs has been known since the Turing machine models, hence from the earliest days of computer science. Vaughan Pratt 1976 was arguably the first to explore the applicability of these notions to logic specifications subsequently leading to dynamic logic, but the importance of transitions between machine states was understood much earlier. In the late 1970s, Hans Kamp [1978; 1981] would discover that the interpretation of indefinites and anaphoric pronouns required the same conception of interpretation: the meaning of a sentence would be no longer simply a function from either its syntactic structure or some correlated language of thought structure onto some articulated concept of truth conditions, but rather a *relation* between one discourse context and another, a concept notably closer to the dynamics of performance. Theo Janssen’s 1981 dissertation

---

¹This last assumption is rejected by Chomsky, who has never advocated the incorporation of a denotational semantics within the grammar.
would make explicit the link between the semantics for programming languages and the semantics for natural languages. Nevertheless, the Chomskian emphasis on syntactic competence and sentence internal studies of language was sustained unquestioningly in all theoretical frameworks for the next several decades.

Over the years, this led to an accumulation of puzzles - syntactic, semantic and pragmatic. A very large proportion of these puzzles coincide on the problem of the endemic context-sensitivity displayed by natural-language expressions, which have in each case to be explained relative to a methodology that is poorly suited to capturing such data. The problem is that context-sensitivity in all its various guises is no respecter of sentence or even utterance boundaries: all context-dependent phenomena can have their interpretations resolvable both within and across sentence and utterance boundaries. And, more strikingly still, all dependencies — even those identified as syntactic and sentence-internal — can be split across participants in conversational dialogue, as individuals have free ability to extend what another person says, to interrupt and take over the articulation of some emergent structure. The puzzles arose because the notions of modularity and the privileged position of the language faculty responsible for the production of grammatical strings and, as extended by Fodor for its interpretation in the language of thought, left no possibility for modelling these sorts of interactions, though exceptions were made for indexical expressions of the sort that Kaplan had studied. Any model taking these principles as basic was poorly adapted to reflecting both context dependency in general, and more particularly the way in which, in conversational exchanges, participants fluently engage in what may be highly interactive modes of communication.

This sentential prejudice has thus left its mark: such models, simply, provide no insight into the nature of context. Putative sub-sentential exemplars of context-dependence in interpretation have been defined in terms of static and global constructs of variable binding to determine fixed construals within a given domain, set by the boundary of a sentence. Supra-sentential exemplars are defined as outside such domains, hence different in kind, indeed complementary. This phenomenon was originally taken to be restricted to anaphora, invariably seen as divided into grammar-internal dependencies vs discourse-level dependencies. But as semanticists developed increasingly sophisticated formal tools for modelling context-relative aspects of nominal construal, of tense, of aspect, of adjectives, of verbs, and of ellipsis, it became apparent that the bifurcation into grammar-internal dependencies and discourse-based dependencies, with each treated as wholly separate from the other, leads to an open-ended set of ambiguities, as the immediate consequence of the sentence-internal remit for formal explications of language; no perspective unifying sentence-internal dependencies and cross-utterance dependencies was expressible therein. Even in the absence of overt expressions, i.e. with ellipsis phenomena, there was the same pattern of bifurcation between what are taken to be sentential forms of ellipsis and discourse forms of ellipsis. Furthermore, the possibility of there being ellipsis which is not expressible without reconstructing it at a sentence level until very recently has not even been envisaged. Rather,
by methodological fiat, the various forms of ellipsis have been analysed as constituting complete sentences at some level of abstraction ([Fiengo and Mary, 1994; Dalrymple et al., 1991] are influential syntactic/semantic exemplars respectively). These prejudices fracture the sub-and super-sentential levels from the sentential level, with only this last understood to be the core of ellipsis for grammatical modelling.

Nonetheless, as the chapters of this volume demonstrate, the Chomskian conception of language as a sentence internal matter is evolving into a more nuanced model in those frameworks concerned with formal articulation of semantics. Dynamic semantics has now for thirty years provided analyses of a variety of phenomena — pronominal anaphora, tense and temporality, presupposition, ellipsis ([Kamp, 1978; 1981; Heim, 1982; 1983; Roberts, 1989; 1996; Kamp and Reyle, 1993; Asher, 1993; Van der Sandt, 1992; Fernando, 2001] to mention a few sources) — and the goal has been to provide an integrated analysis of each phenomenon addressed, without, in general, worrying whether the proposed analysis is commensurate with strict sententialist assumptions. Yet evidence has been accumulating that even explanations of core syntactic phenomena require reference to performance dynamics; and grammatical models are now being explored that reflect aspects of performance to varying degrees and take seriously the need to define a concept of context that is sufficiently structurally rich to express the appropriate means whereby grammar-internal mechanisms and context-bound choices can be seen to interact in principled ways [Steedman, 2000; Hawkins, 2004; Phillips, 2003; Mazzei et al., 2007; Asher and Lascarides, 2003; Ginzburg and Cooper, 2004; Kempson et al., 2001; Cann et al., 2005], with a shift of emphasis that includes exploration of grammars that are able to reflect directly the dynamics of conversational dialogue [Cooper, 2008; Ginzburg, forthcoming; Kempson et al., 2011].

It might seem obvious that an approach which seeks to articulate a much richer concept of interaction between language expertise and its relativity to context for construal is scientific common sense, simply what the facts determine. However, from a methodological point of view, switching to such a perspective had seemed inconceivable. That any such shift has become possible is through the coincidence of two factors: first, the pressure of the continually expanding work of semanticists on context-dependency; second the emergence of formal models of dialogue with the potential to reflect the fine-grained and distributed character of interactions in conversational exchanges. Ever since the advent of dynamic semantics (and more informal but equally “contextualist” approaches to pragmatics: Grice 1975, Sperber and Wilson 1986, Horn and Ward 2000), recognition of the extent of the dependence on context of natural language semantics has been growing exponentially. There are now formal models of the context-relativity of full lexical-content words [Asher and Lascarides, 2003; Pustejovsky, 2005]; there are formal models of the systematic coercive and context-relative shifts available from one type of mean-

---

2In so far as the phenomenon in question spanned sentence boundaries, a default assumption has been that such cases can be analysed as conjunction, in the absence of any other connective.
ing to another (see the chapters by Asher, Cooper); there are models of speech acts and their relativity to context (see the de Rooij chapter), of context-relative factors at the syntax/semantics interface (see the Cann and Kempson chapter, and on the application of probability-based decisions to language processing see the chapters of Penn, and Clark and Lappin). Behind many of these arguments is the issue of whether a level purporting to articulate structural properties of such context-relative interpretations should be advocated as part of the grammar. Moreover, advocacy of wholly model-theoretic forms of interpretation as sustained by upholders of the pure Montagovian paradigm (see [Partee, 1996] and other papers in [van Benthem and ter Meulen, 2010] (2nd edition)), has jostled with advocacy of wholly proof-theoretic formulations (see [Ranta, 1995; Francez, 2007]), with mixed models as well (Cooper this volume, and [Lappin and Fox, 2005; Fernando, 2011]), so there are a range of more or less “syntactic” views even within the articulation of natural-language semantics sui generis. At the turn into this century, this articulation of context-relativity is finally being taken up in syntax, with the development of models of syntax reflecting the incrementality of linguistic performance (Cann et al., this volume). This move had indeed been anticipated in the fifties by Lambek’s categorial grammar (with its “left” and “right” operators: [Lambek, 1958]) but was swept aside by the Chomskian methodology which rapidly became dominant.\footnote{3}{Subsequent developments of categorial grammar incorporated type-raising operations to override the sensitivity to directionality intrinsic to the basic operators.} In the neighbouring discipline of psychology, there has been a parallel vein of research with psycholinguists increasingly questioning the restriction of competence modelling to data of grammaticality allowing no reference to psycholinguistic modelling or testing (Baggio et al., this volume).

Another interesting development has been the integration into mainstream linguistics of decision theoretic and game theoretic models exploiting probability and utility measures (van Rooij, this volume). Lewis [1969] already pioneered the use of game theory with the development of signalling games to model conventions, including linguistic conventions, work that was taken up by economists in the 1980s [Crawford and Sobel, 1982; Farrell, 1993]. Linguists have now used these techniques to model implicatures and other pragmatic phenomena, as does van Rooij, bringing a rich notion of intentional contexts to bear on linguistic phenomena. The use of game theoretic models also brings linguistic research back to the later Wittgenstein’s emphasis on language use and interaction. And in formal language theory, bringing these various strands of research together, theorists are now arguing that the original innateness claims about the unlearnability of a natural language are misplaced [Clark and Lappin, 2011, this volume], and that probabilistically based grammars are viable, contrary to the Chomskian view. The scenario we now face accordingly is that the broad cognitive-science research community is progressively giving recognition to the viability of formal models of language that, in building on these influences, are very much closer to the facts of language performance.

The objective of this handbook is, through its chapters, to set out both the
foundational assumptions set during the second half of the last century and the unfolding shifts in perspective taking place in the turn into this century, in which more functionalist perspectives are explored which nonetheless respect the formalist criteria of adequacy that initiated the extension of the formal grammar methodologies to the natural-language case. Moreover, this shift of perspective is displayed in discussions of syntax, semantics, phonology and cognitive science more generally. The opening chapter lays out the philosophical backgrounds provided variously by Frege, Wittgenstein and others (Peregrin), in preparation for all the papers that follow. A set of syntax papers follow, which consider issues of structure and its characterisation relative to orthodox assumptions of natural-language grammar made by minimalist and categorial grammars (Lasnik and Ugiareka, Morrill), with a philosophical evaluation of the foundationalist underpinnings of the minimalist program (Hinzen). The subsequent chapter (Penn) sets outs how, relative to broadly similar assumptions about the nature of grammar, computational linguistics emerged from under the umbrella of machine translation as a theory-driving discipline in its own right. On the one hand mathematical linguistics took off with the development of Chomsky’s early results on the formal languages hierarchy [Chomsky, 1959]. On the other hand, computational linguistic modelling of language very substantially expanded through highly successful methods adopting Bayesian concepts of probability. This constitutes a major conundrum for conventional assumptions about syntax. Far from progress in the development of automated parsing being driven by linguistic theory as these theories might lead one to expect, parsers based on sententialist grammars have largely been set aside in favour of parsers based on probabilities of expressions co-occurring, these achieving notably greater success rate, a robustly replicated result which is at least suggestive that something is amiss with the orthodox conception of grammar. Of the group of semantics papers, Bach and Chao set the point of departure with a discussion of natural language metaphysics; and van Rooij surveys the background research on context-dependence and the semantics/pragmatics boundary that is problematic for sustaining the established competence performance distinction, arguing nonetheless that the phenomena of content and speech act variability can be expressed without abandoning the semantics pragmatics distinction. The papers of Asher and Cooper then directly address the significance of the challenge of modelling the endemic flexibility of lexical content relative to context for natural language expressions, with Cooper invoking shades of later Wittgenstein in seeking to model language itself as a system in flux. The following papers jointly argue for need of a general shift in perspective. Baggio, van Lambalgen and Hagoort argue for a shift of methodology for cognitive science as a whole into one where language is seen as grounded in perception and action. They urge that the data on which linguists construct their theories should reflect data directly culled from performance, a move which requires radically reshaping the competence-performance distinction. Cann, Kempson and Wedgwood follow this spirit: they argue that syntax is no more than the projection of a representation of some content along a real-time dimension, as displayed in both parsing and production. In the realm of
phonology, Carr argues that the reason why the phenomenon of phonology seems invariably incommensurate with the patterns of syntax/semantics is precisely that it is only the latter which constitute part of the grammar. Clark and Lappin address the challenge of modelling the process of language learning by a child from the conversational dialogue data to which they are exposed, directly countering the influential claims of unlearnability of Gold [1967], which were based on the presumed need to identify learnability of all strings of a language, hence including the worst case scenario. They argue to the contrary from within formal language learning theory assumptions that language learning can indeed be seen as an achievable and formalisable task, if we see the acquisition of linguistic knowledge as taking place within the supportive interactive environment that is provided in ongoing conversation. Hurford explores the expanding horizons in the study of language evolution, arguing for a gradualist, functionally motivated view of language evolution, a view which is at odds with the sharp division of the competence-performance distinction as standardly drawn. McConnell-Ginet argues that the sensitivity to group allegiances displayed by individuals in relation to gender issues is context-relative and context-updating, in the manner of other types of context-dependence; and she concludes that the individualistic approach of conventional methodologies cannot be the sole mode of explanation of types of dependency which natural languages can realise. And the book closes with an overview of anthropological linguistics, explorations of the interface between language and culture, and the overlapping concerns of anthropologists, semanticists and pragmatists, as seen from the anthropological perspective (Beeman).

The bringing together of these chapters has been an extended exercise stretching across two sets of editors, and we have to express our very considerable gratitude to the authors, some of whose patience has been stretched to the limit by the extenuated nature of this process. We have also to thank John Woods and Dov Gabbay for turning to us for help in establishing a philosophy of linguistics handbook. And, most particularly, we have to express our fervent thanks to Jane Spurr for patiently and steadfastly nurturing this process from its outset through the minutiae of the closing editorial stages to its final completion. Jane handles the editorial and publication process as a continuing source of good humour, so that each difficulty becomes eminently hurdleable even when the finishing line threatens to recede yet again.

BIBLIOGRAPHY


LINGUISTICS AND PHILOSOPHY

Jaroslav Peregrin

1 THE INTERACTION BETWEEN LINGUISTICS & PHILOSOPHY

Like so many sciences, linguistics originated from philosophy’s rib. It reached maturity and attained full independence only in the twentieth century (for example, it is a well-known fact that the first linguistics department in the UK was founded in 1944); though research which we would now classify as linguistic (especially leading to generalizations from comparing different languages) was certainly carried out much earlier. The relationship between philosophy and linguistics is perhaps reminiscent of that between an old-fashioned mother and her emancipated daughter, and is certainly asymmetric. And though from philosophy’s rib, empirical investigation methods have ensured that linguistics has evolved (just as in the case of the more famous rib) into something far from resembling the original piece of bone.

Another side of the same asymmetry is that while linguistics focuses exclusively on language (or languages), for philosophy language seems less pervasive — philosophy of language being merely one branch among many. However, during the twentieth century this asymmetry was substantially diminished by the so called linguistic turn, undergone by numerous philosophers — this turn was due to the realization that as language is the universal medium for our grasping and coping with the world, its study may provide the very key for all other philosophical disciplines.

As for the working methods, we could perhaps picture the difference between a philospher of language and a linguist by means of the following simile. Imagine two researchers both asked to investigate an unknown landscape. One hires a helicopter, acquires a birds-eye view of the whole landscape and draws a rough, but comprehensive map. The other takes a camera, a writing pad and various instruments, and walks around, taking pictures and making notes of the kinds of rocks, plants and animals which he finds. Whose way is the more reasonable? Well, one wants to say, neither, for they seem to be complementary. And likewise, contemporary research within philosophy of language and linguistics are similarly complementary: whereas the philosopher resembles the airman (trying to figure out language’s most general principles of functioning, not paying much attention to details), the linguist resembles the walker (paying predominant attention to

\[\text{See [Rorty, 1967]. See also [Hacking, 1975] for a broader philosophical perspective.}\]
details and working a slow and painstaking path towards generalizations). And just as the efforts of the two researchers may eventually converge (if the flyer refines his maps enough and the walker elevates his inquiries to a certain level of generalization), so the linguist and the philosopher may find their respective studies meeting within the realm of empirical, but very general principles of the functioning of language.

Unfortunately though, such meetings are often fraught with mutual misunderstandings. The philosopher is convinced that what is important are principles, not contingent idiosyncrasies of individual languages, and ridicules the linguist for trying to answer such questions as *what is a language?* with empirical generalizations. The linguist, on the other hand, ridicules the philosopher for sitting in an ivory tower and trying to tell us something about languages, the empirical phenomena, without paying due attention to their real natures.

### 2 LINGUISTIC CONCEPTIONS OF THE NATURE OF LANGUAGE

In the nineteenth century, the young science of linguistics was initially preoccupied with comparative studies of various languages. But concurrently it started to seek a subject which it could see as *its own:* is linguistics really to study the multiplicity of languages, or is it to be after something that is invariant across them? And if so, what is it? Similar unclarities arose w.r.t. a single language. What, in fact, *is* a language? Some chunk of mental stuff inside its speakers? Some repertoire of physiological dispositions of the speakers? Some social institution? These questions have subsequently led to fully-fledged conceptions of the nature of language; the most influential of which were tabled by Ferdinand de Saussure (in the end of the nineteenth century) and much later, in the second half of the twentieth century, by Noam Chomsky.

#### 2.1 De Saussure

The Swiss linguist Ferdinand de Saussure, in his posthumously edited lectures published as the *Course of general linguistics* [1916], was the first to provide for linguistics’ standing on its own feet in that he offered an answer to all the above mentioned questions: it is, he argued, a peculiar kind of *structure* that is the essence of each and every language, and the peculiar and exclusive subject matter of linguistics is this very structure. Therefore linguistics basically differs from natural sciences: it does not study the overt order of the tangible world, but a much more abstract and much less overt structure of the most peculiar of human products — language. Studying the psychology, the physiology or the sociology of speakers may be instrumental to linguistics, it is, however, not yet linguistics.

In fact, the conclusion that language is a matter of structure comes quite naturally — in view of the wildness with which the lexical material of different languages often differs. Far more uniformity is displayed by the ways in which the respective materials are sewn together and the traces left by these ways on their
products — complex expressions. But de Saussure claimed not only that grammatical rules and the consequent grammatical structures of complex expressions are more important than the stuff they are applied to; his claim ran much deeper. His claim was that everything which we perceive as “linguistic reality” is a structural matter which is a product of certain binary oppositions. According to him, language is a “system of pure values” which are the result of arrangements of linguistic terms; and hence that language is, through and through, a matter of relations and of the structure these relations add up to.

What exactly is this supposed to mean? What does de Saussure’s term “value” amount to? How is the value of an expression produced by relations among expressions? De Saussure claims that all relevant linguistic relations are induced by what he calls “identities” and what would be, given modern terminology, more adequately called equivalences, which can also be seen as a matter of oppositions (which are, in the prototypical cases, complementary to equivalences). Moreover, he claims, in effect, that values are mere ‘materializations’ of these equivalences resp. oppositions: saying that two elements are equivalent is saying that they have the same value. To use de Saussure’s own example, today’s train going from Geneva to Paris at 8:25 is probably a physical object which is quite different from yesterday’s train from Geneva to Paris at 8:25 — however, the two objects are equivalent in that both are the same 8:25 Geneva-to-Paris train. The abstract object the 8:25 Geneva-to-Paris train is, in this sense, constituted purely by the (functional) equivalence between certain tangible objects; and in the same sense the values of expressions are constituted purely by (functional) equivalences between the expressions.

Moreover, De Saussure saw the equivalences constitutive of ‘linguistic reality’ as resting upon some very simple, binary ones (i.e. such which instigate division into merely two equivalence classes). And these are more instructively seen in terms of the corresponding oppositions — elementary distinctions capable of founding all the distinctions relevant for any system of language whatsoever. (Just as we now know complicated structures can be implemented in terms of bits of information and hence in terms of a single 0-1 opposition.) Hence de Saussure saw the complicated structure of language as entirely emerging from an interaction of various kinds of simple oppositions, like the opposition between a voiced and an unvoiced sound.

De Saussure’s structuralism thus consists first and foremost in seeing language as a system of values induced by elementary oppositions. Moreover, there is no ‘substance’ predating and upholding the oppositions — all items of language, including the most basic ones (“units”), are produced by them. According to de Saussure, language does not come as a set of predelimited signs; it is primarily an amorphous mass, the “units” and other “elements” of which acquire a firm shape only via our creative reflections. It is very misleading, claims de Saussure, to see an expression as the union of a certain sound with a certain concept. Such a view would isolate the expression from the system of its language; it would lead to an unacceptably atomist view that we can start from individual terms and construct
language by putting them together. The contrary is the case: we start from the system and obtain its elements only through analysis.

Hence Saussurean structuralism does not consist merely in the reduction of ‘abstract’ entities to some ‘concrete’ ones (“units”) and their oppositions — it proceeds to reduce also those entities which appear to us, from the viewpoint of the more abstract ones, as ‘concrete units’ or ‘basic building blocks’, to oppositions. “/The characteristics of the unit blend with the unit itself/,” (ibid., p. 168) as de Saussure himself puts it. This means that language is a matter of oppositions alone — “language is a form and not a substance” (ibid., p. 169).

Language, according to de Saussure, has the “striking characteristic” that none of its elements are given to us at the outset; and yet we do not doubt that they exist and that they underlie the functioning of language. This means that although language is primarily an incomprehensible mess or multiplicity, we must take it as a ‘part-whole system’ in order to grasp and understand it. Language thus does not originate from naming ready-made objects — associating potential ‘signifiers’ with potential ‘signifieds’ — for both the signifiers and the signifieds are, in an important sense, constituted only together with the constitution of language as a whole.

All in all, de Saussure’s claim is that besides the ‘natural order’ of things, as studied by natural sciences, there is a different kind of order which is displayed by the products of human activities, especially language, and which is irreducible to the former one. Thus linguistics has its peculiar subject matter — the structure of language.\(^2\)

De Saussure’s insistence that the subject matter of linguistics is essentially ‘unnaturalizable’ — that the structures in question constitute, as it were, an independent stratum of reality, soon became influential not only within linguistics, but across all the humanities. Many partisans of philosophy, anthropology, cultural studies etc. saw this view as a basic weapon for emancipating the humanities from natural science. The resulting movement is now known as structuralism (see [Kurzweil, 1980; Caws, 1988]).

2.2 Chomsky

The other towering figure of linguistics, who has produced a fully-fledged conception of the nature of language which gained a broad influence, is the American linguist Noam Chomsky. His 1957 book *Syntactic Structures* was unprecedented particularly by the extent to which the author proposed supporting linguistics by mathematics. This was unusual: for although the Saussurean picture may — from today’s perspective — have already seemed to invite mathematical means (especially the means of universal algebra, which has come to be understood as the general theory of abstract structures), the invitation was actively suppressed by many of his followers. (Thus Roman Jakobson, an extremely influential post-

\(^2\)For more information about de Saussure’s approach, see [Culler, 1986; Holdcroft, 1991; Harris, 2001].
Saussurean linguistic structuralist, found precisely this aspect of de Saussure’s teaching untenable.) Chomsky based his account of language on the apparatus of *generative and transformational grammars*: of precisely delimited systems of rules capable of producing all and only well-formed sentences of the language in question. These grammars may be, and have been, studied purely mathematically, but their *raison d’être* was that they were intended to be used for the purpose of reconstructing real languages, thus bringing to light their ‘essential structure’. In later years Chomsky upgraded this picture in a number of ways (see [Hinzen, this volume]).

What is important from the viewpoint addressed here, however, is the fact that he turned his attention to the very nature of the covert structure he revealed behind the overt surface of language (see esp. [Chomsky, 1986; 1993; 1995]). And while de Saussure was apparently happy to see the structure as a *sui generis* matter (a matter, that is, of neither the physical world, nor a mental reality — whereby he lay the foundations of structuralism with its own peculiar subject matter), Chomsky takes the order of the day to be naturalism (see 5.2) in the sense of accommodability of any respectable entity within the conceptual framework of natural sciences. Thus he sees no way save to locate the structure of language firmly in the minds of its speakers (while naturalism tells us further that mind and brain cannot but be two sides of the same coin).

Strong empirical support for many of Chomsky’s views came from research into language acquisition. Chomsky noticed that the data an infant adept of language normally has are so sparse that it is almost unbelievable that he/she is able to learn the language, and usually does so rather quickly and effortlessly. Chomsky’s solution is that a great part of language — mostly the structure — is *inborn*. What the infant must truly acquire thus reduces to the vocabulary plus a few parameters of the grammar — everything else is pre-wired up within his/her brain. In this way Chomsky kills two birds with one stone: he solves the problem of the “poverty of the stimulus” concerning language acquisition, and provides a naturalistic explanation of the nature of the structure he reveals within the depths of language.

Chomsky stresses that it is essential to distinguish between that which he calls the E-language and that which he dubs I-language (the letters ‘E’ and ‘I’ standing for ‘external’ and ‘internal’, respectively). Whereas the former consists of all the intersubjective manifestations of language, linguistics is to concentrate on the I-language, which underlies the E-language and which is essentially a matter of the *language faculty*, a specific part of the module of human mind/brain devoted to linguistic skills. Hence there is a sense in which linguistics is, eventually, reducible to a branch of psychology (or even neurophysiology). And the structures envisaged by Chomskyan transformational grammars are ultimately structures founded within this faculty.\(^4\)

---

\(^3\)See [Hopcroft and Ullman, 1979] or [Révész, 1991].

\(^4\)For more about Chomsky and his school see [Pinker, 1994; Cook *et al.*, 1996].
3 PHILOSOPHICAL CONCEPTIONS OF THE NATURE OF LANGUAGE

Philosophers, of course, were interested in language since the dawn of their discipline (probably the first systematic treatise on language was Plato’s dialogue *Cratylus* from around 370 b.c.e.). However, though they took language as an important subject matter, they did not take it as a prominent one. In particular, although studying language was usually regarded as a philosophically important enterprise, it was considered to be secondary to studying thought or the world — for language was usually assumed to be merely an instrument for externalizing thoughts or representing the world.

Some of the modern philosophers having undergone the linguistic turn would claim that the study of language was always totally prominent for philosophy — though the earlier philosophers did not realize this, for they mistook the study of linguistic structures for the study of the structure of thought or the world. Thus Benveniste [1966] famously argued that the categories of Aristotle's metaphysics are in fact nothing else than the categories of Greek grammar; and Carnap's [1934] conviction was that the only genuine philosophical problems that make any sense are linguistic ones in disguise.

Some of the pre-modern philosophers were interested in language not only qua philosophers, but also qua rudimentary scientists. Thus, for example the influential *Port-Royal Grammar*, compiled in 1660 by A. Arnauld and C. Lancelot, was a fairly systematic (though by our current standards rather too speculative) attempt at a general theory of language (though again, it treated language as an entity wholly instrumental to thought). However, it was not until linguistics reached the stage of a fully-fledged science that philosophy could truly be seen as addressing its foundational issues; it was in the twentieth century that philosophers began to pay systematic attention to concepts such as *meaning*, *grammar*, *reference* etc.; and indeed to the very concept of *language*.

3.1 Language as a code

The naive view has it that language is a matter of the interconnection of expressions (sounds/inscriptions) with meanings. Those philosophical conceptions of language which build directly on this intuition attempt to reveal the nature of language by revealing the natures both of the interconnection and of the entities so interconnected with expressions.

Seeking for a paradigmatic example of this kind of interconnection, we are likely to hit upon the interconnection of a proper name and the bearer of this name. This connection appears to be relatively perspicuous: both in how it comes into being (*viz.*, in the typical case, a kind of christening) and in how it is maintained (people forming an association between the name and its bearer, calling the bearer by the name ...). Taking it as the paradigm for semantics, we arrive at what can be called the *code conception of language* (see, e.g., [Dummett, 1993]) or the *semiotic conception of language* [Peregrin, 2001]. According to it, expressions generally
stand for (or name or encode or ...) some extralinguistic entities. The basic idea behind this conception is clearly articulated by Bertrand Russell [1912, Chapter V] — words may get meaning only by coming to represent some entities already encountered by us:

We must attach some meaning to the words we use, if we are to speak significantly and not utter mere noise; and the meaning we attach to our words must be something with which we are acquainted.

However, to make the name-bearer relation into a true paradigm of the expression-meaning relationship, we must indicate how it can be generalized to expressions of categories other than proper names. What could be thought of as named by a common noun, or a sentence (not to mention such grammatical categories as adverbials or prepositions)?

Gottlob Frege [1892a; 1892b] argued that if we take names as generally naming individuals, then there are sound reasons to take (indicative) sentences as naming their truth values (construed as abstract objects — truth and falsity); and he also argued that predicative expressions should be seen as expressing a kind of function (in the mathematical sense of the word), a function assigning the truth value true to those individuals which fall under them and false to those which do not. Equipped with the modern concept of set, we might want to say that predicative expressions (including common nouns) name the sets of objects falling under them.

However Frege (1892) also stressed that these objects cannot be sensibly conceived of as meanings of the words in the intuitive sense (which did not prevent him from calling them, perversely, Bedeutungen, i.e. meanings5). They form only one level of semantics, which must be supplemented by another, which Frege called that of Sinnen, i.e. senses:

Though Frege’s terminology was not found satisfactory, as what is intuitively the meaning of an expression is what he calls its “sense” and what he calls “meaning” was later usually called “referent”, his two-level outline of semantics was to reappear, in various guises, within most of the twentieth century theories of meaning.

5Whether German “Bedeutung” is an exact equivalent of English “meaning” is, of course, open to discussion — but that it is closer to “meaning” than to something like “reference” is beyond doubt.
Here, for example, is the popular semantic triangle from an influential book by Ogden and Richards (1923):

![Semantic Triangle Diagram]

Another elaboration of Frege's ideas was offered by Carnap [1947], who replaced Frege's terms "sense" and "meaning" by the more technical "intension" and "extension", concentrating, in contrast to Frege, on the former (for it is intension, he concluded, which is the counterpart of the intuitive concept of meaning) and thus paving the way for 'intensional semantics' (see 5.3).

From the viewpoint of this two-level semantics, the general idea of meaning as a thing stood for can be developed along at least two very different lines. Namely, we may either claim that it is the relationship expression-referent which constitutes the backbone of semantics, or we may claim that it is rather the relation expression-meaning. Let us deal with the two cases in turn.

3.1.1 Naming things and the relation of reference

According to the one view, the basic task of language is to provide us with tools (and perhaps a 'framework') for giving names to things which surround us. The meaning of a name, if any, is in this way parasitic on its referent (cf. Frege's Sinn as the "way of givenness" of a Bedeutung). Whatever non-name-like expressions and whatever other means a language may possess, they are to be seen as an 'infrastructure' for the crucial enterprise of naming (see, e.g., [Devitt, 1981].)

It is clear that naming objects is one of the things we indeed use our language for. Moreover, it seems that something like naming, based on ostension, plays a central role within language learning. However, we have already noted that in its most straightforward form it is restricted to proper names, which do not appear to form a truly essential part of our language. The importance of the relation may be enhanced by taking not only proper names, but all nominal phrases, as vehicles of reference — the 'improper' names are usually considered as referring in a 'non-rigid' way, namely in dependence on some empirical fact. Thus the phrase "the president of the USA" refers to a person determined by the empirical event of the last presidential election in the USA.

Considerations of these definite descriptions (as they are called since [Russell, 1905]) caused many authors to conclude that most if not all of the names and singular phrases we use are of this kind; and they stimulated various elaborations.
of the logical means of analyzing descriptions (see [Neale, 1990; Bezuidenhout & Reimer, 2003]). Russell’s celebrated analysis led him to assert that, from a logical viewpoint, definite descriptions not only fail to qualify as names, but are not even self-contained phrases; in themselves they refer to nothing, for from the logical viewpoint, they are essentially incomplete. Thus, what a sentence such as *The present king of France is bald* in fact conveys, according to Russell, is not the ascription of baldness to an individual, but the conjunction of the following three propositions: (i) there is an individual which is a king of France; (ii) any individual which is a king of France is identical with this one; and (iii) this individual is bald. Analyzed in this way, the sentence contains no name of a king, but only a predicate expressing the property of being a king of France.

However, it is not difficult to modify the Russellian view in such a way that definite descriptions become self-contained: any description *the P* becomes the name of the single individual falling under *P*, if there is such a single individual; if not, the description names nothing. This would require us to admit nominal phrases without reference, which was impossible within the logic Russell favored, but which appears to be desirable independently; for it seems reasonable to distinguish between saying something false about an existing entity and talking about no entity at all. This became especially urgent within the framework put forward by Strawson [1950] and devised to distinguish between the reference of a nominal phrase in itself and the reference of a specific utterance of the phrase in a context.

The Russellian analysis, moreover, is clearly not applicable to most cases of the usage of the definite article encountered within normal discourse — for it would permit the correct usage of such phrases as “the table” only if there is one and only one table within the whole world. However, the fact seems to be that we very often use the definite article for the purposes of anaphoric reference — for the purpose of referring not to the only relevant thing within the universe, but rather to the only relevant thing among those which are salient within the current context. This led to a semantic analysis of anaphoric uses of the definite article as well as of pronouns and other anaphoric elements based on the assumption that these elements pick up specific elements of the context; and their reference is thus essentially context-dependent [von Heusinger & Egli, 2000; Kamp & Partee, 2004].

These developments fit well with the Fregean two-level notion of semantics: both definite descriptions and anaphoric items have a certain content, and the interplay of this content with some contingent facts (state of the world, context) produces (or fails to produce) their (contemporaneous) referent. However, over recent decades, some philosophers (especially [Kripke, 1972; Putnam, 1975]) have argued vigorously that in many important cases reference can be mediated neither by a description, nor by a Fregean sense or a Carnapian intension. Their claim was that not only proper names, but also terms for the natural kinds (‘water’, ‘gold’, ‘tiger’, ...) obtain their reference through a non-mediated contact with the world. According to this view, the reference of these terms is not derivative to, but rather constitutive of, their content. These considerations initiated what is
sometimes called a *new theory of reference* (see [Humphreys & Fetzer, 1998]).

Can the relation of reference be extended also beyond nominal phrases? Can we see, e.g., the common nouns, such as 'pig' or 'philosopher', as referring to definite entities in a way analogous to that in which ‘Snowball’ refers to a particular pig and ‘Aristotle’ to a particular philosopher? We have already seen that a candidate might be the sets of items falling under the relevant predicates — the set of pigs for ‘pig’ and the set of philosophers for ‘philosopher’. However, we can hardly coin the word ‘pig’ by christening the set of pigs in a way analogous to coining a proper name by christening an individual (if only for the reason that some pigs die and new ones are born every moment and hence the set we would christen would cease to exist almost immediately, leaving the word reference-less again).

Therefore, it might be more plausible to assume that common nouns refer to something like ‘pighood’ or ‘the property of being a pig’. But then it is again unclear how we manage to refer to such entities and what kind of entities they are. (A survey of attempts at formulating systematic theories of properties is given by Bealer and Mönlich, [1989].) Kripke and Putnam tried to force a parallel between proper names and some common nouns (natural kind terms) by claiming that what we christen are *essences* of natural kinds — thus when I point at water and say ‘water!’ I am christening the essence of water and hence making the noun correctly applicable to all and only chunks of water (see [Soames, 2002], for a discussion).

There is perhaps also another option: we might assume that a common noun is a tool of opportunistic referring to this or that individual item falling under it: that we use the noun ‘pig’ to refer to this or another pig, depending on the context. However, this fails to explain the entire role of ‘pig’ — *viz.* such locutions as ‘There is no pig here’.

Hence, although there are ways of extending the concept of reference to expressions other than proper names, they often rob the concept of most of its original appeal: the attraction of grounding language on the reference relation was so attractive especially because reference links a word with a tangible object, which can be pointed at. Moreover, even if we manage to extend reference to names other than proper ones, or perhaps also to sentences or verbs, there will still be a number of grammatical categories whose words cannot be treated as directly vehicles of reference — prepositions, connectives etc. If we want to delimit their roles within the notion of language as basically a means of referring to things, we would have to specify ways in which they aid the other, referring expressions to accomplish their tasks.

### 3.1.2 The semiotic conception of language

Let us now turn to the second way of elaborating the notion of language as a code. Here it is claimed that, just as a proper name *means* its bearer by representing it, so all other expressions, in order to have any meaning, must also represent *some* kind of entity — expressions of kinds different from names perhaps representing objects very different from ‘individuals’. The fact of meaning is necessarily grounded in
a *semiosis* — in the constitution of a *sign* which interconnects a *signifier* with a *signified* and makes it possible for the signifier to act as a proxy for the signified. As Reichenbach [1947, p. 4] puts it:

Language consists of *signs*. ... What makes them signs is the intermediary position they occupy between an object and a sign user, i.e., a person. The person, in the presence of a sign, takes account of an object; the sign therefore appears as the substitute for the object with respect to the sign user.

This way of viewing language leads to the subordination of the category of word under the more general category of *sign*. The general theory of signs was first developed by Charles S. Peirce (see [Hoopes, 1991]). Peirce’s [1932, p. 135] definition of the concept of sign was:

A sign, or representamen, is something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, it creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call the *interpretant* of the first sign. The sign stands for something, its *object*. It stands for that object, not in all respects, but in reference to a sort of idea, which I have sometimes called the *ground* of the representamen.

Peirce classified signs into three categories. The first kind of sign is an *icon*; this is a sign which, in Peirce’s own words, “partakes in the characters of the object”, or, in a more mundane wording, is characterized by a perceptible similarity between the signifier and the signified (thus, a map is an icon of a landscape). The second kind is an *index*, which “is really and in its individual existence connected with the individual object”, i.e. is based on a causal relationship (smoke is an index of fire). The third kind is a *symbol*, which is characterized by “more or less approximate certainty that it will be interpreted as denoting the object, in consequence of a habit”, i.e. by the signifier and the signified being tied together by convention (five circles are the symbol of the Olympic Games). Language is then taken to be simply a collection of symbols.

Charles Morris [1938, p. 3], characterized the process of *semiosis*, in which the two parts of a sign get collated and become the signifier (a *sign vehicle*, in Morris’ term) and the signified (a *designatum* or *denotatum*) as follows:

something takes account of something else mediately, i.e. by means of a third something. Semiosis is accordingly a mediated-taking-account-of. The mediators are *sign vehicles*; the takings-account-of are *interprets*; ...what is taken account of are *designata*.

An outstanding later representative of the semiotic approach to language is Eco [1979; 1986]. According to him, the crucial achievement was “to recognize the *genus* of sign, of which linguistic signs are *species*”. Moreover, as “language was increasingly believed to be the semiotic system which could be analyzed with the
most profit (...) and the system which could serve as a model for all other systems (...), the model of the linguistic sign gradually came to be seen as the semiotic system par excellence” [Eco, 1986, 33]. Hence a certain shift: from presenting and exploiting linguistic sign as subordinate to sign in general, to presenting it as a generally paradigmatic kind of sign. (For more about this kind of semiotic approach to language viz. [Sebeok, 1989].)

The semiotic conception appears to tally with the Saussurean approach; indeed Saussure called his own theory of linguistic signs *semiology* (though he rejected seeing language as a kind of *nomenclature* — as a matter of links between ready-made words and meanings.) For this reason it was readily embraced by many partisans of post-Saussurean structuralism; until it was challenged by what has become known as *poststructuralism* (see 3.3.2).

### 3.2 Language as a toolbox

Meanwhile, other twentieth century philosophers concluded that it was misleading to see language as a system of names. In its stead they proposed seeing it rather as a kind of ‘toolbox’, a kit of tools which we employ as means to various ends. From this viewpoint, the meaning of an expression does not appear to be a thing named by the expression, but rather the capability of the expression to promote particular kinds of ends.

The later Wittgenstein [1969, 67] expresses this view of language in the following way:

> In the tool box there is a hammer, a saw, a rule, a lead, a glue pot and glue. Many of the tools are akin to each other in form and use, and the tools can be roughly divided into groups according to their relationships; but the boundaries between these groups will often be more or less arbitrary and there are various types of relationship that cut across one another.

But already long before this, the American pragmatists, taking language primarily as human activity, had seen linguistic meaning as “primarily a property of behavior” [Dewey, 1925, 179] rather than a represented entity. And recent ‘pragmatist turn’ [Eddington & Sandbothe, 2004], which has rediscovered many of the ideas of classical pragmatism, has resulted in seeing language as not primarily a code, but rather as a means of interaction; and hence in seeing meaning as primarily a matter of the aptitude of an expression to serve a specific purpose, rather than its representing an object.

#### 3.2.1 Speech act theories

In reaction to the theories of language drawn up by those philosophers who, like Russell or Carnap, concentrated especially on language in its capacity of articulating and preserving *knowledge*, different philosophical theories arose which concentrated instead on language as a means of everyday communication. Activities
in this direction were pioneered in particular by the Oxford scholars J. L. Austin, G. Ryle and H. P. Grice, who earned the label of *ordinary language philosophers*. Austin [1961] initiated what has subsequently been called the *speech act theory*. He concentrated not on categories of expressions or sentences, but rather on categories of utterances. His program was to undertake a large-scale ‘catalogization’ of these categories:

Certainly there are a great many uses of language. It’s rather a pity that people are apt to invoke a new use of language whenever they feel so inclined, to help them out of this, that, or the other well-known philosophical tangle; we need more of a framework in which to discuss these uses of language; and also I think we should not despair too easily and talk, as people are apt to do, about the infinite uses of language. Philosophers will do this when they have listed as many, let us say, as seventeen; but even if there were something like ten thousand uses of language, surely we could list them all in time. This, after all, is no larger than the number of species of beetle that entomologists have taken the pains to list.

Austin [1964] distinguished between three kinds of acts which may get superimposed in an act of utterance: the *locutionary act* is “roughly equivalent to uttering a certain sentence with a certain sense and reference”, the *illocutionary act* “such as informing, ordering, warning, undertaking, &c., i.e. utterances which have a certain (conventional) force” and the *perlocutionary act*, which amounts to “what we bring about or achieve by saying something, such as convincing, persuading, deterring, and even, say, surprising or misleading” (109).

Grice [1989] maintained that, over and above the rules of language dealt with by Carnap and others, there are also certain ‘rules of communication’, which he called *conversational maxims*. These are the conventions stating that one usually says things which are not only true, but relevant, substantiated etc. (And these rules are, according to Grice, part and parcel of human rationality just as the rules of logic are.) These rules facilitate that saying something can effect conveying something else: if I ask “Where can I get some petrol here?” and get the answer “There is a garage around the corner”, I assume that the answer is relevant to the question and infer that the message is that the garage sells petrol. The pieces of information a participant of conversation infers like this were called by Grice *conversation implicatures*.

Some recent theoreticians, taking up the thread of addressing language via concentrating on the analysis of discourse and communication [Carston, 2002; Recanati, 2004], deviate from Grice in that they concentrate more on pragmatic than on semantic factors of communication (see 5.1). The notion of what is conveyed by an utterance despite not being explicitly said, is no longer identified with Gricean implicatures: instead a distinction is drawn between an implicature and an *explicature* [Sperber and Wilson, 1986], where the *explicature* amounts to the parts of the message that the hearer gets non-inferentially, despite the fact that they
are not part of the literal meaning of the utterance (a typical example is the information extracted from the context such as the unpacking of the “around the corner” into a pointer to the specific corner determined by the particular situation of utterance).

3.2.2 Pragmatist and neopragmatist approaches to language

The ‘end-oriented’ view of language and meaning suggested itself quite naturally to all kinds of pragmatists, who tend to consider everything as means to human ends. The classical American pragmatists maintained, in Brandom’s [2004] words, that “the contents of beliefs and the meanings of sentences are to be understood in terms of the roles they play in processes of intelligent reciprocal adaptation of organism and environment in which inquiry and goal-pursuit are inextricably intertwined aspects”.

This view of language led to a conception of meaning very different from the view of meaning as that which is “stood for” by the expression in question – to its conception as a kind of capability of serving as a means to peculiar communicative (and possibly other) ends. In an instructive way, this is articulated by G. H. Mead [1934, p. 75–76]:

Meaning arises and lies within the field of the relation between the gesture of a given human organism and the subsequent behavior of this organism as indicated to another human organism by that gesture. If that gesture does so indicate to another organism the subsequent (or resultant) behavior of the given organism, then it has meaning. ... Meaning is thus a development of something objectively there as a relation between certain phases of the social act; it is not a physical addition to that act and it is not an “idea” as traditionally conceived.

It is surely not a coincidence that a very similar standpoint was, by that time, assumed also by the leading figure of American linguistics, Leonard Bloomfield [1933, 27]:

When anything apparently unimportant turns out to be closely connected with more important things, we say that it has after all, a “meaning”; namely it “means” these more important things. Accordingly, we say that speech-utterance, trivial and unimportant in itself, is important because it has meaning: the meaning consists of the important things with which the speech utterance is connected with, namely the practical events [stimuli and reactions].

During the last quarter of the twentieth century, the ideas of pragmatism reappeared in a new guise in the writings of some of the American analytic philosophers, who found it congenial to the kind of naturalism (see 5.2) they wanted to endorse. The initiator of this ‘neopragmatist’ approach to language was Willard Van Orman Quine [1960; 1969; 1974], who proposed seeing language as “a social
art we all acquire on the evidence solely of other people’s overt behavior under
publicly recognizable circumstances” [1969, 26]. Therefore, he concluded, “the
question whether two expressions are alike or unlike in meaning has no determin-
ate answer, known or unknown, except insofar as the answer is settled by people’s
speech dispositions, known or unknown” (ibid., 29). (Obviously a standpoint not
too far from Wittgenstein’s.)

Quine thus concluded that as we cannot find out what a word means otherwise
than by learning how it is used, meaning cannot but consist in some aspects of use.
He claims that though a psychologist can choose to accept or reject behaviorism,
the theorist of language has no such choice: every user of language did learn
language by observing the behavior of his fellow speakers, and hence language
must be simply a matter of this behavior. Resulting from this were various kinds
of ‘use-theories of meaning’.

To throw light on the details of meaning, Quine devised his famous thought
experiment with radical translation. He invites us to imagine a field linguist trying
to decipher an unknown ‘jungle language’, a language which has never before
been translated into any other language. The first cues he gets, claims Quine,
would consist in remarkable co-variations of certain types of utterances with certain
types of events — perhaps the utterances of “gavagai” with the occurrences of
rabbits. Quine takes pain to indicate that the links <sentence → situation> thus
noted cannot be readily transformed into the links <word → object> (<“gavagai”
→ rabbit>), for the same <sentence → situation> link could be transformed
into different <word → object> links (not only <“gavagai” → rabbit>, but also
e.g. <“gavagai” → undetached rabbit part>) and there is no way to single out
the ‘right’ one. Hence, Quine concludes, language cannot rest on a word-object
relation, such as the relation of reference.

From this point of view, there is not more to semantics than the ways speakers
employ words, and hence if we want to talk about meanings (and Quine himself
suggests that we would do well to ewschew this concept altogether, making do
with only such contepts as reference and stimulus-response), we must identify
them with the words’ roles within the ‘language games’ that speakers play. This
Quinean standpoint was further elaborated by a number of philosophers, the most
prominent among them being Donald Davidson [1984; 2005] and Richard Rorty
[1980; 1989].

Davidson compromised Quine’s naturalism and pragmatism by stressing the
centrality of the irreducible (and hence unnaturalizable) concept of truth: “With-
out a grasp of the concept of truth,” he claims, “not only language, but thought
itself, is impossible” [1999, 114]. This means that interpreting somebody as us-
ing language, i.e. uttering meaningful words is impossible without the interpreter
being equipped with the concept of truth, which is irreducible to the conceptual
apparatus of natural sciences. On the other hand, Davidson went even further than
Quine by challenging the very concept of language: “There is no such thing as a
language, not if a language is anything like what many philosophers and linguists
have supposed” [Davidson, 1986]. This means that there is nothing beyond our
‘language games’, the interplays of our communication activities. To see language as a steady, abstract system is a potentially misleading hypostasis.

Rorty, on the other hand, has fully embraced the Quinean (neo)pragmatism, but claims that this view of language leads us if not directly to a form of linguistic relativism, then to its verge. He concluded that the Quinean and Davidsonian view of language implies that there is no comparing languages w.r.t. how they ‘fit the world’; and indeed that there is nothing upon which to base an arbitration between different languages. Hence, Rorty [1989, 80] urges, “Nothing can serve as a criticism of a final vocabulary save another final vocabulary” — we cannot compare what is said with how things really are, for to articulate how things really are we again need words, so we end up by comparing what is said with what is said in other words.

3.2.3 The later Wittgenstein and the problem of rule-following

The concept of language game, of course, was introduced by the later Wittgenstein — he employed it to indicate that the ways we use language are far too varied to be reduced to something like ‘naming things’. Wittgenstein [1953, §23] says:

But how many kinds of sentence are there? Say assertion, question, and command? — There are countless kinds: countless different kinds of use of what we call “symbols”, “words”, “sentences”. And this multiplicity is not something fixed, given once for all; but new types of language, new language-games, as we may say, come into existence, and others become obsolete and get forgotten. ... Here the term “language-game” is meant to bring into prominence the fact that the speaking of language is part of an activity, or of a form of life.

This has led some authors to render Wittgenstein as a relativist and a prophet of postmodernist pluralism (see esp. [Lyotard, 1979]).

However, Wittgenstein did not take the statement of the plurality of language games as a conclusion of his investigations, but rather as a preliminary diagnosis leading him to investigate the specifics of this species of ‘game’ and consequently of the nature of language. Wittgenstein concluded that the concept of language game is inextricable from the concept of rule, and as he was convinced that not all the rules can be explicit (in pain of an infinite regress), he decided that the most basic rules of language must be somehow implicit to the praxis of using language. This has opened up one of the largest philosophical discussions of the second half of the twentieth century — the discussion of what it takes to ‘follow an implicit rule’ (see esp. [Kripke, 1982; Baker & Hacker, 1984; McDowell, 1984]).

The approach to language which stresses the importance of rule-determinedness of the usage of expressions (the distinction between correct and incorrect usage) led to a normative variety of the ‘use-theory of meaning’. In parallel with Wittgenstein, this approach was elaborated by Wilfrid Sellars [1991]. Sellars’ view was that concepts and rules are two sides of the same coin; that having a concept is nothing over and above accepting a cluster of rules. Language, according to him, was
directly a system of rules which gets handed down from generation to generation by initiating new adepts into the rule-following enterprise.

Sellars’ continuator Robert Brandom [1994] then redescribed language as a set of rule-governed games centered around the crucial game of giving and asking for reasons. This game, he claims, is prominent in that it gives language its basic point and it is also constitutive of its semantics. As this game is fuelled by our ability to recognize a statement as a sound reason for another statement and as meaning is constituted by the rule within this very game, meaning comes down to inferential role.

3.3 Continental philosophers on language

The conceptions of language outlined so far have been developed mostly by analytic philosophers, i.e. the philosophers from that side of the philosophical landscape where philosophy borders with science; this approach to philosophy has predominated within the Anglo-American realm as well as in some European countries. But on the other side, there are philosophical lands bordering with literature; and the nature of language has also been addressed by philosophers from these realms, the continental ones, whose center has always been in France and some other European lands. And expectably, the theories of these philosophers are sometimes not really theories in the sense in which the term “theory” is employed by scientists or analytic philosophers, but rather texts of a different genre — in some cases more works of art than of science.

3.3.1 Heidegger

Martin Heidegger, probably the most celebrated representative of continental philosophy of the twentieth century, paid language quite a lot of attention. In his early seminal book Sein und Zeit [1927a], he was concerned with the impossibility of considering language as just one thing among other things of our world. Language — or better, speech, which he maintains is more basic than language as a system — is first and foremost our way of “being within the world”; it is not part of the world, but rather, we can say, its presupposition.

Just like the later Wittgenstein, Heidegger vehemently rejected the code conception of language: “not even the relation of a word-sound to a word-meaning can be understood as a sign-relation” [1927b, 293]. And he insisted that the world we live in is always ‘contaminated’ by the means of our language: “we do not say what we see, but rather the reverse, we see what one says about the matter” [1927a, 75]. Thus Heidegger indicates that language plays a crucial role within the forming of our world.

Speech and language kept assuming an ever more important place in Heidegger’s later writings; and he kept stressing the ‘ineffability’ of language. As Kusch [1989, 202] puts it, he maintained that “we cannot analyze language with the help of any other category, since all categories appear only in language”. He also intensified his pronouncement to the effect of the world-forming capacities of language: “Only
where the word for the thing has been found is the thing a thing. Only thus it is. Accordingly we must stress as follows: no thing is where the word, that is, the name, is lacking” [1959, 164].

In an often quoted passage Heidegger [1947, 145] says:

Language is the house of Being. In its home man dwells. Those who think and those who create with words are the guardians of this home. Their guardianship accomplishes the manifestation of Being insofar as they bring the manifestation to language and maintain it in language through their speech.

In this way he reiterates his conviction that language cannot be seen as merely one of the things within the world, but rather as something more fundamental — not only that it is ‘ ineffable’, but also that it is something we should investigate in a disinterested way, characteristic of science.

3.3.2 The French poststructuralists

In France, de Saussure’s structuralist approach to language led, via generalization, to the philosophy of structuralism and subsequently its poststructuralist revision. Originally, it was based on the generalization of de Saussure’s approach from language to other kinds of ‘systems of signification’; however, it has also brought about new and ambitious philosophical accounts of language.

Michel Foucault [1966; 1971] stressed that the structure of languages and of individual discourses within its framework are man-made and are often tools of wielding power and of oppression. Establishing a vocabulary and standards of a discourse we often establish a social order which favors certain groups whereas it ostracizes others (thus, according to Foucault, calling somebody “mad” is primarily not an empirical description, but rather a normative decision). Therefore, language is a very powerful tool in ‘creating reality’ — it is not a means of describing a ready-made world, but rather a means of production of a world of our own:

The world does not provide us with a legible face, leaving us merely to decipher it; it does not work hand in glove with what we already know ... . We must conceive discourse as a violence that we do to things, or, at all events, as a practice we impose upon them; it is in this practice that the events of discourse find the principle of their regularity.

The most celebrated poststructuralist thinker to deal with language, Jacques Derrida [1967], concentrated especially on the criticism of the “metaphysics of presence.” Meaning, Derrida argues, is usually conceived of as wholly present, as a “transcendental signified”; however, according to him, significance is always a matter of not only presence (of some ‘parts’ of meaning), but also of a necessary absence, of a deference (of other ones). (Hence Derrida’s neologism dif‘érance.)

The failure to see this dialectical nature of any signification, according to Derrida, is closely connected with what he calls the logocentrism of the ordinary
Western philosophy. It was, he says, de Saussure’s failure that he did not utterly repudiate the traditional metaphysical conception of significance, but merely replaced the traditional metaphysics of meanings-objects by the new metaphysics of structures. We must, Derrida urges, see language as lacking any substantial ‘centre’ — hence his views are usually labeled as poststructuralist.

4  KEY CONCEPTS

Aside of the very concept of language, linguistic and philosophical accounts of language usually rest on some fundamental concepts specific to their subject matter. Without aspiring to exhaustivity we list what may be the most crucial of them.

4.1  Grammar

A grammar of a language amounts to the ways in which its expressions add up to more complex expressions. (Sometimes this term is employed so that it applies not only to the expressions themselves, but also to their meanings.) A grammar is usually seen as a system of rules which, thanks to the Chomskyan and post-Chomskyan mathematization of linguistics, can be captured formally in various ways.

Some theoreticians of language, especially logically-minded philosophers, take grammar to be merely ‘in the eye of the beholder’ — i.e. to be just a theoretician’s way of accounting for the apparent ability of the speakers to produce an unrestricted number of utterances. Hence they take the concept of grammar as a not really essential, instrumental matter.

On the other hand, from the perspective of many linguists, it is this very concept which appears as the key concept of the whole theory of language — for grammar, according to this view, is the way in which language is implemented within the human mind/brain. After Chomsky [1957] presented his first mathematical way of capturing grammar, several other attempts (due to himself as well as his followers) followed. This was followed by attempts at addressing semantics in straightforwardly parallel terms [Lakoff, 1971; Katz, 1972]. Also Chomsky himself incorporated semantics into his theory of the “language faculty” as one of its grammatical levels (that of “logical form”).

The concept of grammar is important also because it underlies the much discussed principle of compositionality of meaning [Janssen, 1997; Werning et al., 2005]. This principle states that the meaning of every complex expression is uniquely determined by the meanings of its parts plus the mode of their combination. (Another, equivalent formulation is that to every grammatical rule $R$ there exists a semantical rule $R^*$ so that the meaning of $R(e_1, ..., e_n)$, where $e_1, ..., e_n$ are expressions to which $R$ is applicable, always equals the result of applying $R^*$ to the respective meanings of $e_1, ..., e_n$.) The role of grammar within this principle is essential — taking grammar to be wholly arbitrary trivializes it (for then every
language becomes compositional); so we can have a nontrivial concept of compositionality only if we rely on some substantial concept of grammar [Westerståhl, 1998].

4.2 Meaning

The study of meaning is, of course, a natural part of the study of language; and it was a linguist, Michel Bréal [1897] who coined the word *semantics*. However, the study of meaning within linguistics was always hindered by the fact that the linguists were not quite sure *what exactly* to study under the heading of meaning. Even de Saussure, who proposed the structuralist foundations of linguistics, did not give a clear answer to this question; and Chomsky explicitly denied that we need any such things as meanings to account for linguistic communication. (“As for communication,” he claims [1993, p. 21], “it does not require shared ‘public meanings’ any more than it requires ‘public pronunciations’.”).

However, as a matter of fact, we often do speak about meaning: we say that words acquire, change or lose their meanings, we distinguish between words or expressions which do have meaning and those which do not etc. This made many philosophers contemplate the question *what kind of entity* (if any) *is meaning?* For the answer there are four basic kinds of candidates:

1. Meaning is a ‘tangible’ object, i.e. an object of the physical world. This answer suggests itself if we take proper names as our paradigm of meaningful expressions (see 3.1). However, if we insist that each ‘meaningful’ expression should have a meaning, then there are clearly not enough suitable entities of this kind to fulfill the task. What would be, e.g., the tangible meaning of ‘pig’? We have already seen that it can be neither a particular pig; nor the collection of all existing pigs (unless we want to allow the word to change its meaning all the time). Therefore probably no one would want to explicate the concept of meaning in this way — though these considerations may lead to a view of language in which the concept of meaning is superseded by the concept of reference (see 3.1.1).

2. Meaning is a mental entity. This explication avoids the problem of the previous one, as the mental realms appear to be inexhaustibly rich. However, it faces another kind of problem: it would seem that meaning, by its very nature, must be something that can be *shared* by various speakers and hence cannot be locked within the head of any of them. Nevertheless, there is little doubt that meaningful language is closely connected with mental content; and hence psychologist theories of semantics flourish (see [Shiffer, 1972; 1987; Fodor, 1987; 1998]).

3. Those who think that meaning is an object and admit that it can be neither physical, nor mental are forced to maintain that it must be an entity of a ‘third realm’ (beyond those of the physical and the mental). This was
the conclusion from Frege [1918/9], who initiated a host of semantic theories grappling with meaning using the means of mathematics or logic. The semantics of the formal languages of logic was then elaborated especially by Tarski [1939]; but this still provided no suitable framework for natural language analysis. Only after Chomsky’s revolution in linguistics did the methods of ‘formal semantics’ come to be applied to natural language; the first to do this quite systematically was Montague [1974].

4. A large number of philosophers and linguists put up with the conclusion that there is no such object as meaning, that the meaning talk is a mere façon de parler. This does not mean that there is no distinction between meaningful and meaningless expressions; but rather that meaningfulness should be seen as a property of an expression rather than as an object attached to it. Typically, to have such and such meaning is explicated as to play such and such role within a language game.

Aside of the questions concerning the ‘substantial’ nature of meaning, we can investigate also its ‘structural’ nature. This is to say that there are some determinants of meaning which hold whatever kind of stuff meanings may be made of. An example of such a principle is the principle of compositionality (see 4.1), or else the principle stating that if two sentences differ in truth values, then they are bound to differ in meanings [Cresswell, 1982]. Structuralism with respect to meaning can then be characterized as the standpoint denying meaningfulness of the ‘substantial’ questions and concentrating on the ‘structural’ ones. In the spirit of this standpoint Lewis [1972, p. 173] claimed that “in order to say what a meaning is, we may first ask what a meaning does and then find something which does that.”

4.3 Reference

The paradigm of the relation of reference is the link between a singular term, such as “the king of Jordan” and the object within the real world that is ‘picked up’ by the term — the actual king. Some theoreticians of language argue that this is the relationship constitutive of language, for they see the whole point of language in referring to things (see 3.1.1).

On the other extreme, there are theories which deny reference any important place at all. An example of such an approach is Quine’s, resulting into the doctrine of the indeterminacy of reference (see 3.2.2), which, according to Davidson [1979, pp. 233-234], must lead us to the conclusion that “any claim about reference, however many times relativized, will be as meaningless as ‘Socrates is taller than’.”

From the viewpoint of the two-level semantics (see 3.1), the level of reference (Frege’s level of Bedeutung, Carnap’s level of extension) is considered important also because it appears to be just on this level that truth emerges (indeed, according to both Frege and Carnap, the reference of a sentence directly is its truth value). However, Carnap’s considerations indicated that this level is not ‘self-sustaining’:
that the extension of many complex expressions, and consequently truth values of many sentences, are a matter of more than just the extensions of its parts (in other ways, extensions are not compositional — see [Peregrin, 2007]).

4.4 Truth

One of the most crucial questions related to the working of language was always the question how does language “hook on the world”. And it was often taken for granted that it is the concept of truth which plays an important role here — for is it not truth which is the mark of a successful “hooking”? Do we not call a sentence or an utterance true just when it says things within the world are just the way they really are?

Viewed in this way, truth appears to be something like the measure of the success of the contact between our linguistic pronouncements or theories and reality; and hence appears as one of the indispensable concepts of any account of language. This construal of truth as a matter of correspondence between the content of what is said and the facts of the matter is almost as old as the interest in language itself — thus, Aristotle [IV 7, 1011b25-28] writes

To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true.

However, the construal of truth as a correspondence has been often challenged on the grounds that the idea of comparing two such different entities as a (content of a) linguistic expression and a (part of the) world does not make any understandable sense — what can be compared, claim the critiques, is always a statement with another statement, a belief with another belief, or a proposition with another proposition. This led to an alternative, coherence theory of truth, which maintains that truth amounts to a coherence between a statement (or a belief) and a body of other statements (beliefs). The trouble with this construal of truth is that the concept of coherence has never been made sufficiently clear.

During the first half of the twentieth century, the logician Alfred Tarski [1933; 1944] tried to provide a theory of truth in the spirit of contemporary axiomatic theories of other general concepts (e.g. set or natural number). And though some of the consequences of his achievement are still under discussion, their influence on almost all subsequent theoreticians of truth has been overwhelming. Tarski concluded that what we should accept as the determinants of the theory of truth are all statements of the form

The sentence ... is true iff ...

where the three dots are replaced by a name of a sentence and the three dashes by the very sentence. Thus, an instance of the scheme is, for example,

The sentence ‘Snow is white’ is true iff snow is white.
Tarski showed that to find a finite number of axiom entailing the ensuing infinite number of statements requires underpinning the concept of truth with the semantic notion of satisfaction (this holds for languages of the shape of predicate logic, on which he concentrated; for natural languages it might possibly be a concept such as designation — cf. Carnap, 1942). Some of Tarski’s followers have taken this as indicating that Tarski’s theory is a species of the correspondence theory; others have taken it to be sui generis (the semantic conception).

Today, we can distinguish several competing answers to the question about the nature of truth (see [Kirkham, 1992; Künne, 2005], for more details). Besides various elaborations of the correspondence theory (see [Davidson, 1969; Armstrong, 2004]) and the coherence theory [Rescher, 1973], we can also encounter various neopragmatic approaches, taking truth as a form of utility [Rorty, 1991], approaches taking truth as a kind of ideal justifiability [Dummett, 1978], ‘minimalist’ or ‘deflationist’ theories based on the conviction that the role of the truth-predicate within language is purely grammatical and hence that there is really no concept of truth [Horwich, 1998], and also theories which, contrary to this, hold the concept of truth for so fundamental that it is incapable of being explained [Davidson, 1999].

5 METHODOLOGICAL ISSUES

It is clear that linguistics is partly carried out in accordance with the relatively clear methodological canons of empirical science. However, we saw that the closer we are to such abstract questions as what is meaning? and what is language?, the less clear its methodological tenets are. Should we answer these questions by comparative investigations of various languages; or should we resort to some kind of ‘philosophical’ or ‘a priori’ analysis?

Let us survey some of the most discussed problems concerning the ways to study language.

5.1 Syntax, semantics and pragmatics

The study of language is usually subdivided into various subdisciplines. The most common division, cannonized by Morris [1938], distinguishes between

syntax, which deals with the relations between expressions;

semantics, which addresses the relations between expressions and what they stand for;

and

pragmatics, which examines the relations between expressions and those who use it.

This delimitation has been widely accepted, but is also subject to quarrel. Philosophers usually do not question the boundary between syntax and semantics (though
within some linguistic frameworks, in which semantics looks very much like an ‘inner syntax’, even this boundary may get blurred), but they often dispute the one between semantics and pragmatics (see [Turner, 1999]).

The boundary is clear only when we stick to the code conception of language: within this framework an expression comes to literally stand for its meaning (or its referent) and we may say that pragmatics concerns various ‘side-issues’ of this standing for. Pragmatics thus appears as entirely parasitic upon semantics. On the other hand, from the viewpoint of the toolbox conception it looks as if, on the contrary, semantics were parasitic upon pragmatics: the meaning of an expression appears to be simply the most central part of the employment of the expression by its users. Hence semantics comes to appear as a (rather arbitrarily delimited) core of part of pragmatics.

5.2 Naturalism

What kind of idiom should we use to account for language and meaning? What kind of reality do we refer to when we say that an expression means thus and so?

Modern science tends to take for granted that everything there really is is capturable by the conceptual means of natural sciences and consequently perhaps of physics, to which the other natural sciences are thought to be principally reducible. This kind of naturalism seems to suggest that if the talk about language and meaning is to be understood as contentful at all, then it too must in principle be translatable into the language of physics. So how can we so translate a statement to the effect that some expression means thus and so? In general, there seem to be three possibilities:

1. We can try to reduce the concept of meaning to the concept of reference and explain reference physicalistically — usually in terms of a causal connection [Field, 1972] or a co-occurrence [Dretske, 1981; Fodor, 1998].

2. We can claim that we do not need the concept of meaning at all and all we have to do is to describe the way we use language and/or the way our brains back up this usage [Quine, 1960; Chomsky, 1995].

3. Posit some irreducible non-physicalist concepts. The most popular options appear to be the concept of intentionality between mental contents, and consequently expressions expressing them, and things in the world [Searle, 1983]; and the normative mode of speech rendering the meaning talk as a normative talk (explicate E means thus and so roughly as E should be used thus and so — [Brandom, 1994]).

5.3 Formal models

When Chomsky bridged the gulf which traditionally separated linguistics from mathematics, the study of language became receptive to the ‘mathematization’
which many natural sciences had undergone earlier. Language as an empirical phenomenon (just like many other empirical phenomena) is described in mathematical terms to obtain a ‘model’, which is investigated using mathematical means and the results are then projected back on the phenomenon. (We can also understand this mathematization as a matter of extracting the structure of the phenomenon in the form of a mathematical object.)

In his first book, Chomsky [1957] often talked about “models of language”; however, later he has ever more tended to see the rules he was studying as not a matter of a model, but as directly engraved within the “language faculty” of the human mind/brain. Formal models of language, however, started to flourish within the context of the so called formal semantics (a movement on the borders of logic, linguistics, philosophy and computer science) which used mathematical, and especially mathematico-logical means to model meaning.

This enterprise was based on the idea of taking meanings-objects at face value and hence modeling language as an algebra of expressions, compositionally (and that means: homomorphically) mapped on an algebra of denotations, which were usually set-theoretical objects. As this amounted to applying the methods of model theory, developed within logic (see, e.g. [Hodges, 1993]), to natural language, this enterprise is sometimes also referred to as model-theoretic semantics). The first models of language of this kind were the intensional ones of Montague [1974], Cresswell [1973] and others; and various modified and elaborated versions followed (see [van Benthem & ter Meulen, 1997], for an overview).

Some of the exponents of formal semantics see their enterprise as the underwriting of the code conception of language, seeing the relationship between an expression of the formal model and its set-theoretical denotation as a direct depiction of the relationship between a factual expression and its factual meaning. This, however, is not necessary; for the relation of such models to real languages can be understood in a less direct way — for example the set-theoretical denotations can be seen as explication of inferential roles of expressions (see [Peregrin, 2001]).

5.4 Linguistic universals and linguistic relativism

One of the tasks often assigned to a theory of language is the search for ‘linguistic universals’, for features of individual languages which appear to be constant across them. The study of such universals is then considered as the study of ‘language as such’ — of a type whose tokens are the individual natural (and possibly also some artificial) languages. Theoreticians of language often differ in their views of the ratio of the universal vs. idiosyncratic components of an individual language.

At one extreme, there are ‘universalist’ theories according to which all languages are mere minor variations of a general scheme. Thus, Wierzbicka [1980] proposed that there is a minimal, generally human conceptual base such that every possible language is merely its elaboration. Also, Chomsky suggests that the most important inborn linguistic structures are the same for every individual — learning only delivers vocabulary and fixes a few free parameters of this universal structure.
At the other extreme, there are those who doubt that there are any important linguistic universals at all. These ‘linguistic relativists’ claim that, at least as for semantics, individual languages may well be (and sometimes indeed are) so ‘incommensurable’ that their respective speakers can not even be conceived as living within the same world. The idea of such relativism goes back to Wilhelm von Humboldt, and within the last century it was defended both by linguists [Sapir, 1921; Whorf, 1956] and by philosophers [Cassirer, 1923; Goodman, 1978].

6 PROSPECTS

It is clear that a language, being both a ‘thing’ among other things of our world and a prism which is related to the way we perceive the world with all its things, has one aspect which makes it a subject of scientific study and another which makes it an important subject matter for philosophical considerations. Hence, linguistics and philosophy (of language) are destined to cooperate. However, the fruitfulness of their cooperation largely depend on the way they manage to divide their ‘spheres of influence’ within the realm of language and on building a suitable interface between their ‘spheres’. Fortunately, the host of scholars who study language disregarding barriers between disciplines continually increases.

The list of questions situated along the border of linguistics and philosophy, the answers to which are far from univocally accepted, is long; without pretending to exhaustivity, let me indicate at least some of the most important:

- **the nature of language**: Should we see language primarily as a communal institution; or rather as a matter of individual psychologies of its speakers; or rather as an abstract object addressable in mathematical terms?

- **the nature of meaning**: Should we see meaning as an abstract object, as a mental entity or rather as kind of role?

- **the nature of reference**: What is the tie between an expression and the thing it is usually taken to ‘refer to’? Is its nature causal, is it mediated by some non-causal powers of human mind (‘intentionality’), or is it perhaps a matter of ‘rules’ or ‘conventions’?

- **language vs. languages**: does it make sense to ponder language as such, or should we investigate only individual languages (making at most empirical generalizations)? How big is the ‘common denominator’ of all possible languages? Can there exist languages untranslatable into each other?

- **the ‘implementation’ of language**: what is the relationship between public language and the states of the minds/brains of its speakers (Chomsky’s E-language and I-language)? Is the former only a kind of statistical aggregation of the manifestations of the former, or does it rather exist in some more ‘independent’ way, perhaps even conversely influencing people’s minds/brains?
• the nature of a theory of language: What conceptual resources should we use to account for language and meaning? Are we to make do with the terms we use to account for the non-human world, or are we to avail ourselves of some additional concept of a different kind? And if so, what kind?

ACKNOWLEDGEMENTS

Work on this text was supported by the research grant No. 401/04/0117 of the Grant Agency of the Czech Republic. I am grateful to Martin Stokhof and Vladimír Svoboda for helpful comments on previous versions of the text.

BIBLIOGRAPHY


A. Tarski. O ugruntowaniu naukowej semantyki, Przegląd Filozoficzny, 1939, 50-57; English translation The establishment of scientific semantics in [Tarski, 1956, pp. 401-408].


STRUCTURE

Howard Lasnik and Juan Uriagereka

1 INTRODUCTION

The twentieth century saw the birth of structural linguistics, a discipline rooted in Ferdinand de Saussure’s posthumous *Cours de linguistique générale* (1916). The *Cours* stressed the need to conceive language as separate from what it is used for. It also concentrated on how language is, not how it changes, as had been the practice in historical linguistics (the dominant approach during the nineteenth century). A system of interconnected values, language as such is pure structure, independent even from the meaning it may convey. Now the notion ‘structure’ itself is hard to pin down, particularly if the set of patterns that any given system is supposed to characterize is not finite. Therein lies the fundamental difference between classical structuralism and modern generative grammar. Earlier models focused on those levels of language that are patently finite (phonology and possibly morphology), leaving linguistic creativity to the realm of language use. Great progress was made in terms, for instance, of the paradigmatic inventory underlying phonemics and the very notion of linguistic feature. However, those levels of language that deal with its recursive nature (syntax) involve a more subtle conception of structure, whose understanding has consequences for many scientific disciplines. Interestingly, also, a treatment of structure in these terms returns to some of the fundamental results in mathematical logic (the axiomatic-deductive method).

In this review of the notion structure we will concentrate on that conception, particularly its computational version. The central challenge of generative grammar is to propose a finite system of procedures to describe an infinite output. Although the milieu within which these ideas were developed is interesting (see [Tomalin, 2006; Scholz and Pullum, 2007]), for the most part we will concentrate only on the ideas themselves, concretely the so-called Chomsky Hierarchy and related notions. This is a systematization and generalization of proposals that, through the work of Bar-Hillel [1953], connects to Ajdukiewicz [1935]. Chomsky’s work was influenced by Rosenbloom [1950] (connecting to [Post, 1947; Thue, 1914]) and Davis [1958] (which circulated in manuscript form during the early fifties). These works developed a version of recursive function theory based on unrestricted rewriting systems and progressively more intricate conditions imposed on them. Chomsky’s take on these matters in the fifties, however, and explicitly in *Aspects of the Theory of Syntax* (1965), is unique, and in fact different from
both the logico-mathematical tradition just alluded to and even the structuralist approach (which Chomsky inherited from [Harris, 1951]).

Chomsky certainly benefited from all these lines of research (as well as the search for simplicity in theorizing, emphasized by Goodman [1951]), but took a naturalistic stand that had been absent from modern linguistic studies. This is not just in terms of seeking the mental reality of linguistic structures, a reaction to behaviorism (as in his 1959 review of [Skinner, 1957]). More generally Chomsky was interested in the neuro-biological nature of linguistic structure (a concern that only Hockett [1955] shared), and to this effect he teamed up with Morris Halle, George Miller, Eric Lenneberg and several younger colleagues at Harvard and MIT, to create what was eventually to be known as the ‘biolinguistics’ approach. This decision deeply affects what to make of the notion ‘linguistic structure’.

The matter can be clarified, using terminology from [Chomsky, 1965], in terms of distinguishing the so-called weak generative capacity of a computational system and its strong generative capacity, notions defined for a formal language. One such object consists of an alphabet (a set of symbols) and some combinatorial rules, typically concatenations into strings of elements of the relevant alphabet. Given a formal grammar, a well-formed formula (abbreviated wff) is a string that is derived by that formal grammar. In this view, a given grammar \( G \) is said to generate a formal language \( L \), a set of expressions. We can then say that a string \( S \) is a wff (with respect to \( G \)) if \( S \) belongs to the set \( L(G) \), and concomitantly that \( G \) weakly generates \( L(G) \). While this is all quite relevant for specifying axiomatic systems in general (e.g. to describe the structure of a mathematical proof), it is an empirical matter how the notion of a formal language (and weakly generating it) relates to the biolinguistic notion of a natural language. One could certainly assume that the relevant alphabet is a mental lexicon and the corresponding rules happen to be the combinatorial procedures that determine linguistic structure in a human mind. The ensuing object in that instance, for a grammar like the one presumed relevant for a language like English, generates a set of expressions, each a wff.

Chomsky has systematically rejected the idea that what human speakers ‘carry in their minds’ is aptly described that way. This is, in part, because speakers appear to have intuitions even about unacceptable expressions (e.g., they can find the subject in the ungrammatical *who did John die?*). As Chomsky [1986] argued (developing points already made in [1965]), this suggests that wffs are somewhat orthogonal to what constitutes knowledge of a language. It is of course senseless to declare who did John die a member of the set of grammatical English expressions; but assuming otherwise, from that particular view point, entails denying the factuality of speakers’ intuitions about this expression. In a nutshell, native speakers do not seem to mentally store anything like this English-as-a-set-of-strings.

In a biolinguistic sense, the notion ‘grammaticality judgment’ is as hard to pin down as the notion ‘sickness’.\(^1\) Several factors are involved in each instance, and

\(^1\)Strictly, there are no “grammaticality judgments”. Linguists use judgments of acceptability from native speakers of a given language. Each of those judgments is a small psycho-linguistic
it doesn’t help our understanding of either grammaticality or disease to resolve it by definition. Some sentences ‘sound bad’ to the natives, but degrees of badness arise:

1. (a) Who does John love?
   (b) Who does John love Mary?
   (c) Who do John love Mary?
   (d) Who John love Mary?
   (e) Who John Mary love?

The expressions in (1) degrade as we go down the list, and declaring them (or not) part of ‘the set called English’ doesn’t help us characterize this fact. All objects in (1) except (1a) are not in the relevant set. However speakers have robust intuitions about many structural aspects of all of these expressions. None of that real knowledge can be characterized by an ‘extensional’ view of English. The analogy with a sick animal again drives the point home: declaring it sick, and therefore not a member of some (healthy) species, may help improve breeding methods — but it hardly tells us anything about the structure of the sick animal and, ultimately, its species.

One may be tempted to object that, if we are to proceed in that biolinguistic fashion, it is pointless to use a computational system to describe human language; after all computational systems do, in point of fact, work with wffs. Ultimately that may be true, and nothing in the biolinguistics program would collapse if it were: all it would mean is that the characterization of the relevant structure is to be sought beyond computations. But given our present understanding of these matters, there is no powerful reason to abandon a computational approach to linguistics. All that is necessary is a slight shift in perspective. Intuitively, in (1b) fewer ‘grammatical features’ are violated than in (1c). But then we need a computational system that generates not just the ‘terminal strings’ corresponding to the sentences in (1), but all the nuanced properties of those strings (their agreement marks, configurational relations, and so on). In technical parlance, again from Chomsky [1965], we need to mind the strong generative capacity of the system (the set of strings generated and the structure assigned to those strings)

That distinction is central. In mathematics, it doesn’t make a difference whether we generate, say, the set of negative numbers by inverting addition operations in the way shown, or by using a Turing machine to generate the numbers. Nowadays judgments can be tabulated in careful conditions involving small populations, or be “directly” observed in terms of techniques to measure brain reactions to experimentally modified specifications. But ultimately a theorist must hypothetically decide whether to interpret a given observation in terms of “grammaticality” (as opposed to a host of orthogonal factors pertaining to a speaker’s performance, for example).

As is well-known, the Turing architecture involves a countable number of steps, but there is an uncountable number of sets of natural numbers (i.e. of subsets of \( \mathbb{N} \)). (Basing his approach on Cantor’s famous Uncountability Theorem, Gödel [1934] establishes the basic proof that eventually leads to the familiar computational architecture). Langendoen and Postal [1985] attempted to build one such argument for English, but their sketch of a proof was artificial. The matter remains open.
the set of natural numbers or, rather, by multiplying natural numbers by negative one. The relevant set is (weakly) generated both ways. However, in human language the most basic distinctions are captured only if close attention is paid not just to the output string, but also to the process that took us there. Furthermore, at least in a naturalistic program, the only thing that exists in the mind/brain is the generative grammar, which yields structured expressions, and only yields structureless strings by extra operations (like associativity, or wiping out hierarchy). So weak generation is a derivative notion, based on strong generation.

Consider for example Groucho Marx’s famous comment: “I shot an elephant in my pajamas — what it was doing in my pajamas, I’ll never know”. Close analysis of this ‘structurally ambiguous’ sentence clarifies the joke.

1. *I shot an elephant in my pajamas.*

The structures in (2) are identical at the terminal string. However, the structures also differ when we realize that in (2a) the prepositional-phrase (PP) *in my pajamas* is attached to the verb-phrase (VP) (thereby modifying the event of shooting denoted by the verbal expression), while in (2b) the same phrase is attached to the direct object noun-phrase (NP) (entailing modification of the elephant referred to by that nominal expression). So there are two ways of strongly generating the output string, which is the approach linguists give to the phenomenon of structural ambiguity. A sentence like *I shot an elephant in my pajamas* is structurally ambiguous if it can be generated in non-equivalent ways, as in (2). The joke works because pragmatic considerations favor the structuring in (2a), but deadpanning that the elephant had managed to enter Groucho’s pajamas forces the unexpected reading.
We may call a ‘generative procedure’ not just the results of a derivation (its terminal string) but, rather, the detailed process that took the system there. The complex system constituting the generative procedure may be technically called an I-language, the ‘I’ standing for ‘intensional’ or ‘internal’. From a biolinguistic or I-language perspective, really only those results in Formal Language Theory matter that pertain to strong generative capacity — and these are hard to come by in purely formal terms (see the Appendix). All of that said, however, we feel it is important not to throw out the baby with the bath water: there is reason for optimism in the study of the formal structures that underlie language, even if much work lies ahead. Our approach in this review can largely be seen in this light, as we know of no worked out alternative to the computational theory of mind, even when it comes to I-language as just described.

2 THE CHOMSKY HIERARCHY

Structural linguistics has shown how such phenomena as ‘a phrase’, or even deceptively simple notions like ‘a phoneme’, are far from simple-minded, directly observable conditions that one could easily abstract from, say, a synthesized speech stream. The problem starts with a drastic oversimplification that this very sentence highlights. It has twelve words. This is easy to ascertain in our familiar writing system. Such a system appears so natural to us that we are bound to forget that it involves several theoretical abstractions including one at the level of words, and another at the level of the letters that comprise their orthographic representations. While both of these are important, neither has a privileged status in syntactic theory. In fact, their status within the model turns out to be quite hard to pin down, as we proceed to show in several steps.

Any student of linguistics knows that words group themselves into constituents. While such nuances are part of the training of linguists, no writing system fully represents them. What gets written as words and letters is already a folk theoretical proposal about language. English reminds us how limited that proposal is in the case of letters. If one wanted to remain close to observables, one would have to turn to a phonetic transcription;\(^3\) then the very first issue would be to identify (then abstract) words out of such a speech stream. Writing gives us a shorthand method to achieve this task, although of course it has little to say about phrases, and literally nothing about more complex structural conditions of the sort we discuss below. These ideas were clearly overlooked in the logical tradition that was alluded to in the previous section, and which still informs mainstream computational linguistics, through the very notion of a formal language.

To see the centrality of structure to language, recall how Chomsky began the discussion of grammar in *Syntactic Structures* (1957) by presenting ‘an elementary linguistic theory’ consisting solely of finite state Markov processes (of much use

\(^3\)In fact, even that is already abstract: it presupposes the invariants of universal phonetic transcription.
in thermodynamics and statistical mechanics, mathematical biology, economics, and, in fact, behavior generally). These are mathematical models for a system’s progress, for which a given state depends solely on itself (or a finite collection of preceding states). In other words, Markovian systems are devoid of useful memory. Chomsky offered an often cited argument for the inadequacy of these processes as general models of human languages: ( Portions of) such languages have non-Markovian properties. As presented, the argument was expressed in terms of sets of wffs and the (weak) generative capacity of the grammars generating them. But between the lines of Chomsky [1965, 60-61] is another, equally powerful and more straightforward, argument: Sentences of human languages are structured objects, whereas Markov processes just provide linear strings of symbols.

2.1 The Upper Limits of Finite-state Description

For illustrative purposes, we begin with some simple examples of finite state formal languages (the first finite, the second infinite), from Chomsky [1957], and graphic representations of the finite state Markov processes generating them:

3. The man comes / The men come

4. In the fifties, very elementary Markov processes were assumed to be essentially the whole story not just for language but for behavior generally, in large part because of their role in information.

5. Although technically this is not accurate, as a Markov source generates an intricately structured object, which is converted into a string by assuming associativity among relevant elements. In other words, the object does have structure — but it is the wrong kind of structure for language. It is also worth mentioning that Markovian systems can express the transition between some state and the next in terms of transitional probabilities. We will abstract away from this point here, as it doesn’t affect the reasoning.
5. The man comes / The old man comes / The old old man comes / ...

6.

Alongside these, Chomsky introduces some non finite state context-free languages. We present these now, and context-free grammars generating them. Chomsky calls these $(\Sigma, F)$ grammars, as $\Sigma$ is a finite set of initial strings and $F$ a finite set of Post-style instruction formulas, each rewriting a single symbol (rewrite rules):

7. $ab, aabb, aaabbb, \ldots$, and in general, all sentences consisting of $n$ occurrences of $a$ followed by $n$ occurrences of $b$ and only these.

8. $\Sigma : S$

   $F : S \rightarrow aSb$
   $S \rightarrow ab$

9. $aa, bb, abba, abaa, baaa, aabb, aabaa, abbaa, \ldots$, and in general, all sentences consisting of a string $X$ followed by the ‘mirror image’ of $X$ (i.e., $X$ in reverse), and only these

10. $\Sigma : S$

    $F : S \rightarrow aSa$
    $S \rightarrow bSb$
    $S \rightarrow aa$
    $S \rightarrow bb$

Chomsky shows how English (portions) cannot be described in finite state terms:

11. (a) If $S_1$, then $S_2$
    (b) Either $S3$, or $S4$
(c) The man who said that \(S5\), is arriving today

The crucial property of these examples is not merely that there can be a string of unlimited length between the dependent items (\(if\)-\(then\), \(either\)-\(or\), \(man\)-\(is\)). There can also be a string of unlimited length between the \(the\) and \(man\) in the finite state language (6). But in (11) we have recursion, while in (6) we merely had iteration. As he puts it (p.22):

In [(11)a], we cannot have ‘or’ in place of ‘then’; in [(11)b], we cannot have ‘then’ in place of ‘or’; in [(11)c], we cannot have ‘are’ instead of ‘is’. In each of these cases there is a dependency between words on opposite sides of the comma (i.e., ‘\(if\)’-‘\(then\)’, ‘\(either\)’-‘\(or\)’, ‘\(man\)’-‘\(is\)’). But between the interdependent words, in each case, we can insert a declarative sentence \(S_1, S_3, S_5\), and this declarative sentence may in fact be one of [(11)a-c] ... It is clear, then, that in English we can find a sequence \(a+S_1+b\), where there is a dependency between \(a\) and \(b\), and we can select as \(S_1\) another sequence containing \(c+S_2+d\), where there is a dependency between \(c\) and \(d\), then select as \(S_2\) another sequence of this form, etc. A set of sentences that is constructed in this way ... will have all of the mirror image properties of [(9)] which exclude [(9)] from the set of finite state languages.

\(\Sigma, F\) grammars are capable of handling languages with the properties of those in (7) and (9). Further, they can easily generate all finite state (formal) languages as well, thus yielding a set-theoretic picture as in Figure 1:

![Figure 1. Chomsky Hierarchy up to context-free languages](image-url)

At this point in his presentation, Chomsky simply abandons finite state description.

Abandoning a description because an alternative is \textit{more inclusive} (in the sense of Figure 1) is an argument about the system’s weak generative capacity; i.e.,
an extensional set-theoretic characterization. Chomsky later coined the term E-language (where the $E$ stands for ‘extensional’ and ‘external’) to denote this conception of language, basically anything other than I-language (implying that there is no utility to this notion). He opposed the concept to the linguistically more relevant I-language, mentioned in the Introduction. From the biolinguistic perspective, the linguist’s task is to formulate feasible hypotheses about I-language, to test them against reality (in describing acceptable expressions, how children acquire variants, the specifics of language use, how brains may represent them, and so on). It is an interesting empirical question whether, in I-language terms, a ‘more inclusive’ description entails abandoning a ‘less inclusive’ one, when the meaning of ‘inclusiveness’ is less obvious in terms of a generative procedure.

The descriptive advantage of Post-style PS grammars, as compared to finite state grammars, is that PS grammars can pair up things that are indefinitely far apart, and separated by dependencies without limit. The way they do that is by introducing symbols that are never physically manifested: the non-terminals. That is, PS grammars introduce structure, as graphically represented in the tree diagram of a sentence from language (7), $aaabbb$, where $a$, $b$ are the terminal symbols and $S$ is a non-terminal symbol:

Although we return to this fundamental consideration, we want to emphasize that there is no dispute within generative grammar with regards to the significance of this sort of structure, and arguments abound to demonstrate its reality. We will mention just four.

Consider, first, the contrasts in (13):

13. (a) (Usually,) cats chase mice.
(b) Cats chase mice (, usually).
(c) Cats (usually) chase mice.
(d) Cats chase (*usually) mice.

The question is why adverbs like *usually* can be placed in all the positions in (13) except in between the verb and the direct object. An explanation of the contrasts is possible if an adverb must associate to a phrasal constituent, not just a single word. If there is a constituent formed by *chase* and *mice* (a verb-phrase, VP), then the modification by the adverb in (13c) is as straightforward as in (13a) or (13b), involving even more complex constituents (entire sentences) that get modified. (13d) fails because no constituent is formed by *cats* and *chased*, and therefore the adverb has nothing it can associate to.

Confirmation of the reality of the abstract VP structure stems from the fact that it can be displaced as a unit, which the facts below directly show:

14. (a) They say cats chase mice, and chase mice, I've surely seen they can!
(b) They say cats chase mice, * and cats can, I've surely seen chase mice!

So-called VP fronting is a colloquial way of emphasizing this sort of expression, as in (14a), where the entire constituent *chase mice* is displaced. A parallel fronting, involving the subject *cats* and the verb *can* — though logically imaginable as a source of emphasis of the semantic dependency between cats and their abilities — is unavailable, as (14b) shows. This follows if only phrases can displace and *cats can* is not a phrase. The issue is purely structural. Had language presented ‘subject phrases’ (including the subject and the auxiliary verb) as opposed to verb phrases, the paradigms above would reverse.

Asymmetries between subjects and predicates and what they contain are easy to find, and they provide yet another argument for structure. Thus consider (15), which involves an anaphor *each other* (that is, an element whose antecedent for referential purposes must be grammatically determined, in a sense we are about to investigate):

15. (a) Jack and Jill [kissed each other].
(b) *Each other [kissed Jack and Jill].

Whereas an anaphor in object position can take the subject as its antecedent, the reverse is not true. This is not just a fact about anaphors; asymmetries remain with pronouns

16. (a) Jack and Jill said that [someone [kissed them]].
(b) They said that [someone [kissed Jack and Jill]].

In (16a) the object pronoun can take the names in subject position as antecedent; in contrast, in (16b), with the reverse order, *they* (now in subject position) must
not refer to Jack and Jill in object position. The matter cannot be one of simple precedence in the presentation of names, pronouns and anaphors, for (17), which is very similar to (16b), is in fact fine with their referring (forward) to Jack and Jill:

17. [Their teacher] said that [someone [kissed Jack and Jill]].

The difference between these two sentences is standardly described in terms of structure too: their is buried inside the phrase their teacher in (17), the subject of the main clause, while this is not true for they in (16b), which is itself the subject of the sentence. As it turns out, anaphors, pronouns and names are sensitive to whether or not there is a direct path between them and their antecedent, which goes by the name of c-command.\(^6\) That notion is totally dependent on a precise phrasal description.

Although the divide between subject and predicate is fundamental in determining phrasal asymmetries, it is not the only one. Anaphoric facts of the abstract sort discussed in [Barss and Lasnik, 1986] directly show asymmetries internal to the verb phrase:

18. (a) Jack [showed Jill to herself (in the mirror)].
   (b) *Jack [showed herself to Jill (in the mirror)].

Such asymmetries are prevalent, and become manifest in all sorts of circumstances. For example, while compounds are possible involving a direct object and a verb, as in (20a), whose import is that of (19), they are generally impossible with any other verbal dependent (indirect object, subject, etc.) or with so-called (circumstantial) adjuncts:

19. Mailmen carry letters (for people) (on weekdays).

20. (a) Mailmen are letter-carriers (for people) (on weekdays).
   (b) *Mailmen are people-carriers (letters).
   (c) *Letters are mailman-carrier/carried (for people).
   (d) *Mailmen are weekday-carriers (letters) (for people).

2.2 The Upper (and Lower) Limits of Phrasal Description

Having shown the centrality of phrasal description in language, we want to signal a certain irony that arises at this point. While human language fragments abstractly like (7) and (9) were presented by Chomsky as motivating the move from finite state description to \(\Sigma, F\) description, he did not actually use the crucial relevant

\(^6\)We say that \(A\) c-commands \(B\) if all nodes (crucially in a phrase-marker) which dominate \(A\) also dominate \(B\). This is perhaps the most important structural notion within generative grammar, and it shows up across domains. See Uriagereka [forthcoming, chapter 3] for current discussion and references.
property of these grammars in his linguistic work at the time. That relevant property is unbounded self-embedding. However, the theory of Chomsky [1955], LSLT, assumed in [Chomsky, 1957], has no ‘recursion in the base’. Instead, it is the transformational component that accounts for the infinitude of language. This point is only hinted at in [Chomsky, 1957, p.80], but had been fully developed in [Chomsky, 1955/1975, pp. 518 and 526].

Chomsky proceeded to develop a $\Sigma, F$ grammar for a small fragment of English, and then used the English auxiliary verb system as part of a demonstration of the inadequacy of even $\Sigma, F$ description. In a nutshell, he showed that basic parts of the system have cross-serial dependencies, as in the partial structure of *have been writing*:

21.

```
  have  be  en write ing
```

Phrase structure rules cannot in general deal with these sorts of dependencies; they work at characterizing nested dependencies, as we saw in (8), but not cross-serial ones. This is one of Chomsky’s arguments for a more powerful device — so-called transformations.

The new sort of structuring has the quality of being ‘context-sensitive’, as opposed to ‘context-free’. This is because of the way in which transformational rules are supposed to work. Contrary to what we saw for Post-style rewrite rules (which only care about rewriting a given non-terminal symbol regardless of the context in which it appears) context-sensitive operations are so-called because they do care about the neighborhood in which a given symbol has been generated. Context here simply means a (partial) record of the derivational history. Context-free operations have no use for such a record, but context-sensitive operations are a modification of pre-existing representations, and therefore must know what sort of object to manipulate (the ‘structural description’) in order to produce a related, but different, output (the ‘structural change’).

Now the auxiliary portion of English is finite, so strictly speaking that fragment of English is not beyond the bounds of $\Sigma, F$, or even finite state description. A list is, after all, essentially a trivial finite state grammar. But that is really beside the point. A lower level account would be clumsy and uninformative. As Chomsky put it: ‘... we see that significant simplification of the grammar is possible if we are permitted to formulate rules of a more complex type than those that correspond to a system of immediate constituent analysis’ (p.41). It is important to emphasize

---

7It is not unfair to refer to these formal objects as ‘new’, since they were Chomsky’s genuine contribution. They relate intuitively to objects discussed in [Harris, 1951], and more formally to constructs explored by Thue [1914]. But their specific import within language description (for example to capture auxiliary dependencies or the phenomenon of ‘voice’) is due to Chomsky.
that this simplification is not just a formal matter, the sort of concern that Chom-
sky inherited from Goodman [1951]. As noted in the Introduction, Chomsky was
already deeply interested in biolinguistic issues, as is apparent on the very first
page of Chapter I of [Chomsky, 1955/1975, 6]:

A speaker of a language has observed a certain limited set of utterances
in his language. On the basis of this finite linguistic experience he can
produce an indefinite number of new utterances which are immediately
acceptable to other members of his speech community. He can also
distinguish a certain set of ‘grammatical’ utterances, among utterances
that he has never heard and might never produce. He thus projects his
past linguistic experience to include certain new strings while excluding
others.

*Syntactic Structures* [Chomsky, 1957, 15] also briefly summarizes this point of
view:

... a grammar mirrors the behavior of the speaker who, on the basis of
a finite and accidental experience with language, can produce or under-
stand an indefinite number of new sentences. Indeed, any explication
of the notion ‘grammatical in \( L \)’ (i.e., any characterization of ‘gram-
matical in \( L \)’ in terms of ‘observed utterance of \( L \)’) can be thought
of as offering an explanation for this fundamental aspect of linguistic
behavior.

Interestingly, Chomsky’s conclusion in the case of cross-serial dependencies dif-
fers in a deep respect from his conclusion in the case of inadequacy of finite state
description. In the latter case, he completely rejected the lower level description,
constructing a theory in which Markov processes play no role. But in the one
now under discussion, his proposal was to keep both the lower level (\( \Sigma, F \) descrip-
tion) and a higher level (transformations). He writes on pp. 41-42 of *Syntactic
Structures*:

It appears to be the case that the notions of phrase structure are quite
adequate for a small part of the language and that the rest of the lan-
guage can be derived by repeated application of a rather simple set of
transformations to the strings given by the phrase structure grammar.
If we were to attempt to extend phrase structure grammar to cover
the entire language directly, we would lose the simplicity of the limited
phrase structure grammar and of the transformational development.

We speculate that the different kinds of rejection of the two lower level modes
of description stem from the fact that human language sentences are highly struc-
tured objects, while Markov productions are ‘flat’. Thus, it presumably seemed
at the time that there was no role at all in the grammar for Markov processes. Be
that as it may, it is important to emphasize, once again, that Chomsky’s reasoning
makes sense only if the strong generative capacity of the system is what is at stake.
This is because, as we saw in the related situation in Figure 1 — for the relation between Markovian and phrase-structure conditions — in set-theoretic terms, a new and more inclusive set has now been added, as in Figure 2:

![Figure 2. Chomsky Hierarchy up to context-sensitive languages](image)

In terms of the system's weak generative capacity, one could ask why the context-free conditions should be identified if the system is already at the more inclusive context-sensitive level. Nonetheless, from the point of view of strong generative capacity, the way in which structuring happens at a 'more inclusive' level is radically different from how it happens at the 'less inclusive' one, a matter we return to below. In those terms, while mechanisms of the context-sensitive sort presuppose mechanisms of the context-free sort, it is not true that context-sensitivity makes context-freeness superfluous. A biological analogy with locomotion may clarify this point: running presupposes walking, but the more difficult condition doesn’t make the simpler one superfluous (even if the set of distances a runner runs would cover the set of distances a walker walks).

Curiously — and related to these concerns separating the system's weak and strong generative capacity — there are certain natural language phenomena that plausibly involve a degree of flatness, as discussed by Chomsky [1961, p.15] and Chomsky and Miller [1963, p.298]. One is what Chomsky called 'true coordination' as in (5), repeated here:

8There is a certain reconstruction of history in interpreting the introduction of transformational rules as posing issues of context-sensitivity. In point of fact these rules were introduced for purely linguistic reasons – that is, attempting to find the best explanation for linguistic facts. In other words, context sensitivity never arose at the time, although it arises in automata-theoretic inquiry into formal languages.
22. The man comes / The old man comes / The old old man comes / ...

Chomsky states, for this and certain other cases, that ‘[i]mmediate constituent analysis has been sharply and, I think, correctly criticized as in general imposing too much structure on sentences.’ That is, there is no evident syntactic, semantic, or phonological motivation for a structure in which, say, each *old* modifies the remaining sequence of *olds* plus *man*, as in (23), or some such (with irrelevant details omitted).

23.

![Diagram](image)

Preferable might be something like:

24.

![Diagram](image)

Chomsky [1961] mentions (25) and says (p. 15): ‘The only correct P-marker would assign no internal structure at all within the sequence of coordinated items. But a constituent structure grammar can accommodate this possibility only with an infinite number of rules; that is, it must necessarily impose further structure, in quite an arbitrary way.’

25. The man was old, tired, tall ..., but friendly.

Chomsky and Miller [1963, p.298] present a very similar argument: ‘... a constituent-structure grammar necessarily imposes too rich an analysis on sentences because of features inherent in the way *P[hrase]*-markers are defined for such sentences.’ With respect to an example identical to (25) in all relevant respects, they say:
In order to generate such strings, a constituent-structure grammar must either impose some arbitrary structure (e.g., using a right recursive rule), in which case an incorrect structural description is generated, or it must contain an infinite number of rules. Clearly, in the case of true coordination, by the very meaning of the term, no internal structure should be assigned at all within the sequence of coordinate items.

2.3 On the Inadequacy of Powerful Solutions to Simple Structuring

The conclusion of Chomsky and of Chomsky and Miller about the issue in the previous section was that we need to go beyond the power of \( \Sigma, F \) description to adequately describe natural languages. In particular, the model is augmented by a transformational component. We would like to show in this section that this particular approach does not ultimately address the concern raised in the previous section.

Chomsky [1955] and [1957] had already shown how transformations can provide natural accounts of phenomena that can only be described in cumbersome and unrevealing ways (if at all) by \( \Sigma, F \) grammars. But he had little to say there about the ‘too much structure’ problem we are now considering. Chomsky [1961] and Chomsky and Miller [1963] don’t have much to say either, beyond the implication that transformations will somehow solve the problem. That is, we need to move up the power hierarchy as in Figure 2. In fact, as already mentioned in the previous section, Chomsky [1955] had already claimed that there is no recursion in the \( \Sigma, F \) component, the transformational component (generalized transformations (GTs)) being responsible in toto for infinitude.\(^9\)

Chomsky discussed several aspects of the coordination process, though without giving a formulation of the relevant transformation(s). It is interesting to note that all the examples discussed in Chomsky [1955] involve coordination of two items, as in (26).

\begin{equation}
\text{John was sad and tired}
\end{equation}

For such cases, it is straightforward to formulate a GT, even if, as claimed by Chomsky [1961, p.134], these are strictly binary operations. As an example, he gives \textit{John is old and sad}, from \textit{John is old, John is sad}, with resulting structure (27).

\begin{equation}
\text{John is old and sad}
\end{equation}

\(^9\)A \textit{generalized} transformation maps separate phrase-markers \( K \) and \( L \) into a single phrase-marker \( M \). In this they differ from \textit{singular} transformations, which map a phrase-marker \( K \) into a modified version \( K' \).
Chomsky and Miller also seem to assume binarity, at least in one place in their discussion: ‘The basic recursive devices in the grammar are the generalized transformations that produce a string from a pair [emphasis ours] of underlying strings’ [p. 304]. It is not entirely clear, however, what is supposed to happen when we have multiple items coordinated, as in the phenomena under discussion here, or in, e.g.:

28. John was old and sad and tired.

One possibility is that we would preserve the structure of old and sad in (28), and create a higher structure incorporating and tired.

29.

Or, somewhat revising (27):
Another possibility is a right-branching analogue:

31.

\[
\begin{array}{c}
\text{A} \\
\text{A} \quad \text{and} \quad \text{A} \\
\text{A} \quad \text{and} \quad \text{A} \\
\text{and} \\
\text{tired} \\
\text{old} \\
\text{sad}
\end{array}
\]

But any of these would run afoul of Chomsky’s argument: We do not always want that extra structure. That is not to say the extra structure should never be available. Rather, we must be able to distinguish between the situations where the structure is motivated and those where it is not. For example, a person who is tired for someone being old and sad (with a compositional structure plausibly corresponding to (30)) may not be the same as a person who is old for someone who is sad and tired (31). And both of these might differ from someone who merely has all three properties with no particular relations among one another. But we are lacking a representation for this final case.

Reciprocals bring that out even more clearly. Consider:

32. John and Mary and Susan criticized each other

Given a sentence like this, relevantly different structures may be as in (33) and (34):
A situation making (32) true might involve John and Mary criticizing Susan, and vice-versa, which is naturally expressed as in (33). But it might also be possible, for (32) to be true, if John criticized Mary and Susan, and vice-versa, whose compositional structure is naturally expressed as in (34). Now here is the crucial situation for our purposes: What if each of the three criticized each of the others? This is certainly a possible, indeed even plausible, scenario that would make (32) true. The most natural structure to correspond to such a semantics would seem to be the unavailable flat structure.

A formal possibility that might yield the desired result arises if we relax the binarity requirement altogether. Chomsky and Miller [1963] seemingly countenance this option in at least one place in their discussion: ‘We now add to the grammar a set of operations called grammatical transformations, each of which maps an n-tuple [emphasis ours] of P-markers \((n \geq 1)\) into a new P-marker.’ [p.299] Then a GT could be formulated to coordinate three items (alongside the GT coordinating two items). But, as already noted, there really is no limit on the number of items that can be coordinated — Chomsky’s original point. So this solution merely replaces one untenable situation with another: In place of an infinite number of phrase structure rules, one for each number of coordinated items, we now have an
infinite number of generalized transformations.\textsuperscript{10}

Thus, moving up the power hierarchy ultimately does not help in this instance. In a manner of speaking, what we really want to do is move down the hierarchy, so that, at least in relevant instances, we give ourselves the possibility of dealing with flat structure. Again, finite state Markov processes give flat objects, as they impose no structure. But unfortunately that is not quite the answer either. While it would work fine for coordination of terminal symbols, phrases can also be coordinated, and, again, with no upper bound. Alongside (35), we find (36).

35. John and Mary and Susan criticized each other.

36. The old man and the young man and the boy and ...

So we need a sort of ‘higher order’ flatness to handle this flatness in full generality.

\section*{2.4 Regaining Flatness without Losing Structuring}

The last sentence in the previous section can hardly be meaningful if one’s notion of language is insensitive to structure – an abstract construct that is independent of wffs and the weak generative capacity of a system. It is only if one attempts to capture the nuances of linguistic phenomena via structural dependencies that such abstractions make sense. To start with, the situation in (36) highlights the need for ‘cyclic’ computations, in the sense of Chomsky \textit{et al.} [1956]. Otherwise, it would make little sense to have chunks of structure that have achieved phrasal status (e.g. the old man) concatenate one to the next as if they were words. So what we need should be, as it were, ‘dynamic flatness’. But this is the sort of concept that sounds incomprehensible in a classical computational view, while making sense to those for whom syntactic computations are psychologically real.

A familiar concept treats language in terms of concentric diagrams of the sort in Figure 3, indicating proper subset relations as we have discussed above, which extends to the entire class of denumerable constructs. This is the so-called Chomsky Hierarchy.

Now while we find this concept insightful, it is necessary to be cautious in any attempt to compare any of these diagrams to each other as models of human language competence, as has always been understood, at least within linguistics.

First, to insist on a point we have now raised several times, the proper subsets concern solely weak generative capacity — languages as sets of strings of symbols, or E-languages. Yet, as amply discussed already and forcefully observed by Chomsky [1965, p. 60], ‘the study of weak generative capacity is of rather marginal linguistic interest... Presumably, discussion of weak generative capacity marks

\textsuperscript{10}It is interesting to note that the sort of artificial ‘sentence’ constructed by Langendoen and Postal [1985] (alluded to in fn. 2 for the purposes of challenging the denumerability of linguistic objects) involves coordination in its attempt to construct an infinitely long linguistic token.
only a very early and primitive stage of the study of generative grammar.\textsuperscript{11}

Second, all of the sets in the hierarchy include finite languages, trivially (in fact, all finite languages). But human languages seem to be invariably infinite,\textsuperscript{12} so much of what these diagrams describe squarely falls outside of the language faculty of humans (though it is of course possible, indeed likely, that other animals’ cognition or communication systems can be thus described).

Related to that concern is a third one that has been implicit in our discussion: going up in the hierarchy does not entail that simpler types of structures should be discarded. If human language were some simple formal object, a more powerful description carrying the system to a formal domain where it can describe the

\textsuperscript{11}With the perspective of history, a more guarded way to say this is that, historically, the study of weak generative capacity arose in parallel to the study of generative grammar, and conceptually it constitutes an extension of the study of generative grammar that may not be relevant to the biolinguistic perspective.

\textsuperscript{12}Everett (2005) famously claims otherwise, citing as evidence his own study of the Pirah\={a} language in the Amazon. According to Everett, the syntax of this language presents no recursion. For skepticism regarding such a claim see [Nevins et al., 2009].
complex and the simpler wffs should indeed make us discard the less powerful description, as redundant. But human language does not just worry about a set of strings of words, generated in any imaginable way. Indeed, the very ways of generating the strings of words is what seems central. In the case discussed above, the same string of words — or even phrases — may have different derivational histories and corresponding phonological or semantic properties. That said, the systems we use to generate the observable words, say in finite-state or phrase-structure terms, are no longer redundant.

For example (33) and (34) and a hypothetical flat structure with the very same words simply do not constitute a natural class, despite appearances to the contrary. The good news is that formal languages, in the traditional sense, are systems whose mathematical properties are well understood, at least at the level of their weak generative capacity. The bad news, that for the most part this is probably irrelevant to the concerns of biolinguists. But, although the task ahead seems daunting in formal terms, we find no reason to doubt that the Chomsky Hierarchy is real in a broad sense that ought to matter even to biolinguists. Granted, so far as we know no one has come up with a precise treatment of the relevant formal objects that is sensitive to what really matters in natural language (for the most part, inner structure, not outer word strings). But then again, the intuition is clear enough, and has resisted the fads and controversies that fuel scientific progress: that levels of complexity matter in syntactic structuring.

Thus, for instance, a constituent is not the same as a ‘discontinuous constituent’ (which is usually called a displacement), or for that matter a flatter dependency. That much is agreed upon by every linguist, even if different schools of thought build complexity in different fashions: For some, the syntactic structures themselves yield the relevant structural abstractions needed to yield observable patterns; for others, abstraction should be built into the system’s semantics, by allowing for formal objects of a sort that do not seek their dependencies in more or less local and standard terms, but instead satisfy such needs at a distance. At the level of abstraction that we are talking about, however, these are probably notational variants, though this is hard to prove.

---

13 Similarly, there is no fully formalized theory of ‘computational biology’, a discipline concerned with such problems as the relation between zygote and individual or, more specifically, topological problems of the protein-folding sort. But lack of formalization does not prevent active research in the field.

14 For example, Weir and Joshi [1988] observe that there is a close relation between so-called linear indexed rules and the combinatory rules of categorial grammar. In formalisms of this sort, and even others, it is possible to get linguistic information to pass along in a phrase-marker, by coding it in terms of a higher-order symbol whose purpose is basically to store non-local information. At the relevant level of abstraction, however, a system with lower-order symbols and non-local operations is equivalent to a system with higher-order symbols and local operations. This is not to say, of course, that the systems are totally equivalent. A much more difficult task is to decide whether the differences in structuring and concomitant semantic assumptions in fact correspond to observables in the (bio)linguistic system. We know of no decisive argument one way or the other.
Thus conceived, much work lies ahead, beyond the need for a detailed formulation. For example, pursuing ideas that go back to [Berwick and Weinberg, 1984], Boeckx and Uriagereka [2011] observe that core linguistic structures are extensions of binary constructs, albeit at various levels within the Chomsky Hierarchy. They argue in particular that most finite-state structures (of the infinitely many that could exist) are extensions of reduplicative patterns of the *da-da* sort (as in *very very simple*). In turn, they suggest that projections along phrasal stems are, in some sense, super-reduplicative patterns, stated on non-terminal expressions, not words (so a path going from X to XP is a relation between two X-X labels likened to *da-da* at the morphological level).\(^{15}\) They finally extend this idea from abstract projections to abstract ‘subjacent’ domains, understood as super-projections among systemic phases within the derivation.\(^{16}\) In all these instances, infinitely many other abstract dependencies would be formally possible, but in fact natural language restricts itself to extensions of a very trivial reduplicative pattern, at various levels of formal complexity. This move makes sense only if the Chomsky Hierarchy is taken seriously, making generalizations about abstract structures as opposed to observable symbols.

3 CONCLUSIONS

Here we have taken the view that structure reduces to computational structuring because this is the paradigm we work with, and which we consider productive. But in fairness to the discussion, particularly in the context of an ultimately philosophical analysis, the matter is far from obvious. In point of fact structure, broadly construed, has to be more than (standard Turing) computational structure, or there would be no structure to non-denumerable spaces within mathematics. Now whether that has a bearing on specifically linguistic structuring is something we cannot afford to go into here.

It seems to us that, strictly within the confines of (standard Turing) computational structuring, generative grammar has provided a rather interesting hypothesis about the structure of the human language faculty, as a biological capacity. In this regard, what seems crucial is not to confuse the (technically) weak and strong generative capacities of a grammar, understood as a computational procedure. Somewhat surprisingly, human language appears to be sensitive to the richer notion of structure, to the extent that it seems to deploy radically different forms of structuring of the same superficial observables, and to make use of these differences in very nuanced semantic ways.

\(^{15}\)They are attempting to capture the familiar ‘headedness’ of phrasal projections; that is, the fact that a verb-phrase must contain a verb, and noun-phrase a noun, and so on.

\(^{16}\)The Subjacency Condition, introduced in [Chomsky, 1973], basically states that a long-distance dependency is possible across certain domains (classically called ‘bounding nodes’, nowadays referred to as ‘phases’ in the derivation) only if it proceeds in ‘successive cyclic’ stages, from one such domain to the one immediately containing it. The term ‘subjacent’ refers to this form of internal adjacency between wholes and their parts.
The task ahead for the biolinguistic project is to test that hypothesis in terms of the findings stemming from the neuro-biological and related sciences. If the notion of structure investigated in this chapter is even remotely on track, a very concrete program lies ahead, in terms of what it might mean for relevant representations to ultimately be embodied in the human brain. In the end, this is the most exciting aspect of the generative enterprise. Future understanding, both in empirical and in formal terms, may bring us closer to a viable characterization of linguistic structure within a human mind. But when this structure is postulated as explicitly and flexibly as we have attempted to depict it here, the question of whether its essentials (primitive symbol and process, memory capacity, etc.) have a concrete place in the neuro-biology of brains — while certainly difficult — does not seem senseless at all.

APPENDIX

In this Appendix we want to return to the issue of strong generative capacity in a computational system, concretely studying conditions arising within context-free systems of two closely related sorts. Our objective is to show in detail how decisions about structure in this regard can have far-reaching consequences for the system.

Context-free PS grammars (Σ, F grammars in Chomsky’s terminology) consist of:

1. A designated initial symbol (or a set thereof) (Σ);
2. Rewrite rules (F), which consist of a single symbol on the left, followed by an arrow, followed by at least one symbol.

A derivation consists of a series of lines such that the first line is one of the designated initial symbols, and to proceed from one line to the next we replace exactly one symbol by the sequence of symbols it can be rewritten as until there are no more symbols that can be rewritten. The last line is the sentence (weakly) generated.

Here is a toy example:

i. (a) Designated initial symbol (Σ) : S
   (b) Rewrite Rules (F):
       \[ S \rightarrow NP \, VP \]
       \[ NP \rightarrow N \]
       \[ VP \rightarrow V \]
       \[ N \rightarrow John \]
       \[ V \rightarrow laughs \]

And a derivation using this PS grammar:
PS grammars capture constituent structure by introducing non-terminal symbols, symbols that are part of the grammar, but that do not appear in the sentences generated. Suppose we take (ii), then connect each symbol with the symbol(s) that it had been rewritten as. In this way we can trace back units of structure.

After joining the symbols we can represent the derivation in the form of a tree as in (iii). Getting rid of the symbols that are mere repetitions, we end up with (iv), a ‘collapsed derivation tree’, a familiar way of representing structure.

Interestingly, though, Chomsky’s formalization of phrase structure was not a graph theoretic one like that. Rather, it was set theoretic. A ‘phrase marker’ for a terminal string is the set of all strings occurring in any of the equivalent derivations of that string, where two PS derivations are equivalent if and only if they involve the same rules the same number of times, but not necessarily in the same order.

Additional equivalent derivations for John laughs are the following:
Given the \((\Sigma, F)\) grammar in (i), the representation of the phrase structure (the ‘P-marker’) of \(\text{John laughs}\) is then:

\[
\{ \text{S, NP VP, NP V, N V, N VP, John VP, John V, NP laughs, N laughs, John laughs} \}
\]

Notice that equivalent derivations provide the same information about the constituent structure of the sentence. The information is that \(\text{John left}\) is an \(\text{S}\), and \(\text{John}\) is an \(\text{N}\), and \(\text{John}\) is an \(\text{NP}\), and \(\text{left}\) is a \(\text{V}\), and finally that \(\text{left}\) is a \(\text{VP}\). That, Chomsky claimed, is everything linguistic theory needs to know about the structure of this sentence. That is, we need to be able to determine, for each portion of the terminal string, whether that portion comprises a constituent or not, and, when it comprises a constituent, what the ‘name’ of that constituent is.

To recapitulate, Chomsky’s empirical claim is that all and only what we want a PM to do is to tell us the ‘is a’ relations between portions of the terminal strings and non-terminal symbols. Syntactic, semantic and phonological operations need that information, and no more. Anything that tells us those and only those is an adequate PM; anything that does not is inadequate as a PM. Familiar trees, as in (iii) or (iv), actually provide more information than that, coding further aspects of the histories of the derivations. If Chomsky is right that the additional information is not needed, then his (‘minimalist’) conclusion follows that this information should not be in the representation.

Note now that we don’t even need the entire P-marker to determine the ‘is a’ relations. Some members of the set have the property that they have exactly one non-terminal symbol and any number of terminal symbols. Let us call them ‘monostrings’ (they are italicized below):

\[
\{ \text{S, NP VP, NP V, N V, N VP, John VP, John V, NP laughs, N laughs, John laughs} \}
\]

By comparing the monostrings with the terminal string one by one, one can compute all the ‘is a’ relations in the following fashion.

Compare ‘\(\text{John laughs}\)’ with ‘\(\text{John VP}\)’:

\[
\text{John laughs} \quad \text{John VP}
\]
From (xi) we deduce that \textit{laughs} is a VP. Now, compare ‘\textit{John laughs}’ with ‘\textit{John V}’:

xii. \textit{John laughs}

\begin{center}
\textit{John} V
\end{center}

From (xii), that \textit{laughs} is a V. Next, compare ‘\textit{John laughs}’ with ‘NP \textit{laughs}’:

xiii. \textit{John VP}

\begin{center}
\textit{NP left}
\end{center}

We can conclude that \textit{John} is an NP. And so on.

If all we are trying to do is determine all and only the ‘is a’ relations, we have a straightforward algorithm for doing that: to compare the terminal string and the monostrings. But in general, a set-theoretic PM will contain far more than the terminal string and the monostrings. The monostrings constitute a small percentage of all of the strings in the PM. The question is whether we need the ‘extra’ strings. Lasnik and Kupin [1977] argued that we do not. Since to determine the ‘is a’ relations we only need the terminal string and the monostrings, they proposed a construct called ‘reduced phrase marker’ (RPM), which only includes the terminal string and the monostrings.

In order to construct an RPM we could construct a phrase marker in Chomsky’s sense and then ‘knock out’ everything except the terminal string and the monostrings. But Lasnik and Kupin, instead, built RPMs from scratch and they asked: What is a PM \textit{in a set-theoretic sense}? It is a set of strings. So we can stipulate that \textit{any} set of strings is an RPM so long as it meets some conditions imposed on it. Lasnik and Kupin formalized these conditions. For example, ‘completeness’ (when to determine that all and only the strings in a given set ‘fit’ into an RPM), ‘consistency’ (what specific strings are such that elements within them either dominate or precede all other elements in the RPM), and so on. Operationally, this is very different from Chomsky’s model. In Lasnik and Kupin’s model of the RPM, it is not necessary to go through all the equivalent derivations as seen above, or, in fact, \textit{any} derivations at all.

It is curious to note that Lasnik and Kupin’s was a theory of PS, but it was not based on PS rules at all, unlike the classical theory. The work that PS rules do is really in the way of constructing equivalent derivations, but Lasnik and Kupin did not need those equivalent derivations. So the question is: Does it make sense to have a theory of phrase structure \textit{without PS rules}? A few years later, an answer emerged, most explicitly in [Stowell, 1981]. Here it was argued that PS rules are redundant, duplicating information that must be available in other ways regardless. There is some discussion of this in Chapter 1 of Lasnik and Uriagereka [1988]. That conclusion strengthens the particular model that Lasnik and Kupin explored, as argued in [Martin and Uriagereka, 2000].

The methodological point to bear in mind is that this discussion is all about the \textit{strong} generative capacity of the system. (It should be easy to see that a grammar based on PMs is weakly equivalent to one based on RPMs.) Interestingly, taking
such matters earnestly leads to a different theory, with important biolinguistic consequences.

Nowadays PS is typically formulated in the operational ways suggested in [Chomsky, 1994] (see [Uriagereka, 1998, Appendix]) for an explicit formalization in collaboration with Jairo Nunes and Ellen Thompson). It is not clear to us that this particular instantiation of phrasal conditions, which is constructed ‘bottom-up’ (from lexical items all the way up to the present-day equivalent of $S$ in the discussion above), is an improvement over the Lasnik and Kupin formalism.

ACKNOWLEDGEMENTS

We would like to thank Ruth Kempson for her patience and encouragement, and Norbert Hornstein, Terje Lohndal, an anonymous reviewer, and, especially, Noam Chomsky for numerous helpful suggestions.

BIBLIOGRAPHY


LOGICAL GRAMMAR

Glyn Morrill

1 FORMAL GRAMMAR

The canonical linguistic process is the cycle of the speech-circuit [Saussure, 1915]. A speaker expresses a psychological idea by means of a physiological articulation. The signal is transmitted through the medium by a physical process incident on a hearer who from the consequent physiological impression recovers the psychological idea. The hearer may then reply, swapping the roles of speaker and hearer, and so the circuit cycles.

For communication to be successful speakers and hearers must have shared associations between forms (signifiers) and meanings (signifieds). De Saussure called such a pairing of signifier and signified a sign. The relation is one-to-many (ambiguity) and many-to-one (paraphrase). Let us call a stable totality of such associations a language. It would be arbitrary to propose that there is a longest expression (where would we propose to cut off I know that you know that I know that you know . . .?) therefore language is an infinite abstraction over the finite number of acts of communication that can ever occur.

The program of formal syntax [Chomsky, 1957] is to define the set of all and only the strings of words which are well-formed sentences of a natural language. Such a system would provide a map of the space of expression of linguistic cognition. The methodological idealisations the program requires are not unproblematic. How do we define ‘words’? Speaker judgements of well-formededness vary. Nevertheless there are extensive domains of uncontroversial and robust data to work with. The greater scientific prize held out is to realize this program ‘in the same way’ that it is done psychologically, i.e. to discover principles and laws of the language faculty of the mind/brain. Awkwardly, Chomskyan linguistics has disowned formalisation as a means towards such higher goals.

The program of formal semantics [Montague, 1974] is to associate the meaningful expressions of a natural language with their logical semantics. Such a system would be a characterisation of the range and means of expression of human communication. Again there are methodological difficulties. Where is the boundary between linguistic (dictionary) and world (encyclopedic) knowledge? Speaker judgements of readings and entailments vary. The program holds out the promise of elucidating the mental domain of linguistic ideas, thoughts and concepts and relating it to the physical domain of linguistic articulation. That is, it addresses a massive, pervasive and ubiquitous mind/body phenomenon.
It could be argued that since the program of formal syntax is hard enough in itself, its pursuit should be modularised from the further challenges of formal semantics. That is, that syntax should be pursued autonomously from semantics. On the other hand, attention to semantic criteria may help guide our path through the jungle of syntactic possibilities. Since the raison d’être of language is to express and communicate, i.e. to have meaning, it seems more reasonable to posit the syntactic reality of a syntactic theory if it supports a semantics. On this view, it is desirable to pursue formal syntax and formal semantics in a single integrated program of formal grammar.

We may speak of syntax, semantics or grammar as being logical in a weak sense when we mean that they are being systematically studied in a methodologically rational inquiry or scientific (hypothetico-deductive) fashion. But when the formal systems of syntax resemble deductive systems, we may speak of logical syntax in a strong sense. Likewise, when formal semantics models in particular the logical semantics of natural language, we may speak of logical semantics in a strong sense. Formal grammar as comprising a syntax which is logical or a semantics which is logical may then inherit the attribute logical, especially if it is logical in both of the respects.

In Section 2 of this article we recall some relevant logical tools: predicate logic, sequent calculus, natural deduction, typed lambda calculus and the Lambek calculus. In Section 3 we comment on transformational grammar as formal syntax and Montague grammar as formal semantics. In Section 4 we take a tour through some grammatical frameworks: Lexical-Functional Grammar, Generalized Phrase Structure Grammar, Head-driven Phrase Structure Grammar, Combinatory Categorial Grammar and Type Logical Categorial Grammar. There are many other worthy approaches and no excuses for their omission here will seem adequate to their proponents, but reference to these formalisms will enable us to steer towards what we take to be the ‘logical conclusion’ of logical grammar.

2 LOGICAL TOOLS

2.1 Predicate logic

Logic advanced little in the two millennia since Aristotle. The next giant step was Frege’s [1879] *Begriffsschrift* (‘idea writing’ or ‘ideography’). Frege was concerned to provide a formal foundation for arithmetic and to this end he introduced quantificational logic. Peano called Frege’s theory of quantification ‘abstruse’ and at the end of his life Frege considered that he had failed in his project; in a sense it was proved shortly afterwards in Gödel’s incompleteness theorem that the project could not succeed. But Frege had laid the foundations for modern logic and already in the *Begriffsschrift* had in effect defined a system of predicate calculus that would turn out to be complete. Frege used a graphical notation; in the textual notation that has come to be standard the language of first-order logic is as follows:
Logical Grammar

Figure 1. Semantics of first-order logic

(1) **Definition (language of first-order logic)**

Let there be a set $C$ of (individual) constants, a denumerably infinite set $V$ of (individual) variables, a set $F^i$ of function letters of arity $i$ for each $i > 0$, and a set $P^i$ of predicate letters of arity $i$ for each $i \geq 0$. The set $T$ of first-order terms and the set $F$ of first-order formulas are defined recursively as follows:

$$
T ::= C \mid V \mid F^i(T_1, \ldots, T_i), i > 0 \\
F ::= P^iT_1 \ldots T_i, i \geq 0 \\
\mid \neg F \mid (F \land F) \mid (F \lor F) \mid (F \rightarrow F) \mid \forall VT \mid \exists VT
$$

The standard semantics of first-order logic was given by Tarski [1935]; here we use $\{\emptyset\}$ and $\emptyset$ for the truth values true and false respectively, so that the connectives are interpreted by set-theoretic operations. An interpretation of first-order logic is a structure $(D, F)$ where domain $D$ is a non-empty set (of individuals) and interpretation function $F$ is a function mapping each individual constant to an individual in $D$, each function letter of arity $i > 0$ to an $i$-ary operation in $D^{D^i}$, and each predicate letter of arity $i \geq 0$ to an $i$-ary relation in $P(D^i)$. An assignment function $g$ is a function mapping each individual variable to an individual in $D$. Each term or formula $\phi$ receives a semantic value $[\phi]^g$ relative to an interpretation $(D, F)$ and an assignment $g$ as shown in Figure 1.

A formula $A$ entails a formula $B$, or $B$ is a logical consequence of $A$, if and only if $[A]^g \subseteq [B]^g$ in every interpretation and assignment. Clearly the entailment relation inherits from the subset relation the properties of reflexivity ($A$ entails $A$) and transitivity (if $A$ entails $B$ and $B$ entails $C$, then $A$ entails $C$).
2.2 Sequent calculus

First-order entailment is an infinitary semantic notion since it appeals to the class of all interpretations. Proof theory aims to capture such semantic notions as entailment in finitary syntactic formal systems. Frege’s original proof calculus had proofs as sequences of formulas (what are often termed Hilbert systems). Such systems have axiom schemata (that may relate several connectives) and rules that are sufficient to capture the properties of entailment. However, Gentzen [1934] provided a great improvement by inventing calculi, both sequent calculus and natural deduction, which aspire to deal with single occurrences of single connectives at a time, and which thus identify in a modular way the pure inferential properties of each connective.

A classical sequent $\Gamma \Rightarrow \Delta$ comprises an antecedent $\Gamma$ and a succedent $\Delta$ which are finite, possibly empty, sequences of formulas. A sequent is read as asserting that the conjunction of the antecedent formulas (where the empty sequence is the conjunctive unit true) entails the disjunction of the succedent formulas (where the empty sequence is the disjunctive unit false). A sequent is called valid if and only if this assertion is true; otherwise it is called invalid. The sequent calculus for the propositional part of classical logic can be presented as shown in Figure 2. Each rule has the form $\Sigma_1, \ldots, \Sigma_n, n \geq 0$ where the $\Sigma_i$ are sequent schemata; $\Sigma_1, \ldots, \Sigma_n$ are referred to as the premises, and $\Sigma_0$ as the conclusion.

The identity axiom $id$ and the Cut rule are referred to as the identity group; they reflect the reflexivity and transitivity respectively of entailment. All the other rules are left ($L$) rules, involving active formulas on the left (antecedent) of the conclusion, or right ($R$) rules, involving active formulas on the right (succedent) of the conclusion.

The rules $W$ (weakening), $C$ (contraction) and $P$ (permutation) are referred to as structural rules; they apply to properties of all formulas with respect to the metalinguistic comma (conjunction in the antecedent, disjunction in the succedent). Weakening corresponds to the monotonicity of classical logic: that conjoining premises, or disjoining conclusions, preserves validity. Contraction and weakening, and permutation correspond to the idempotency and commutativity of conjunction in the antecedent and disjunction in the succedent. They permit each side of a sequent to be read, if we wish, as a set rather than a list, of formulas.

Then there are the logical rules, dealing with the connectives themselves. For each connective there is a left rule and a right rule introducing single principal connective occurrences in the active formula in the antecedent ($L$) or the succedent ($R$) of the conclusion respectively.

A sequent which has a proof is a theorem. The sequent calculus is sound (every theorem is a valid sequent) and complete (every valid sequent is a theorem).

All the rules except $Cut$ have the property that all the formulas in the premises are either in the conclusion (the side-formulas in the contexts $\Gamma_{(i)}/\Delta_{(i)}$, and the active formulas of structural rules), or else are the (immediate) subformulas of the active formula (in the logical rules). In the Cut rule, the Cut formula $A$ is a
Figure 2. Sequent calculus for classical propositional logic
new unknown reading from conclusion to premises. Gentzen proved as his \textit{Hauptsatz} (main clause) that every proof has a Cut-free equivalent (Cut-elimination). Gentzen’s Cut-elimination theorem has as a corollary that every theorem has a proof containing only its subformulas (the \textit{subformula property}), namely any of its Cut-free proofs.

Computationally, the contraction rule is potentially problematic since it (as well as Cut) introduces material in backward-chaining proof search reading from conclusion to premises. But such Cut-free proof search becomes a decision procedure for classical propositional logic when antecedents and succedents are treated as sets. First-order classical logic is not decidable however.

2.3 Natural deduction

\textit{Intuitionistic} sequent calculus is obtained from classical sequent calculus by restricting succedents to be non-plural. Observe for example that the following derivation of the law of excluded middle is then blocked, since the intermediate sequent has two formulas in its succedent: $A \implies A \implies A, \neg A \implies A \lor \neg A$. Indeed, the law of excluded middle is not derivable at all in intuitionistic logic, the theorems of which are a proper subset of those of classical logic.

\textit{Natural deduction} is a single-conclusioned proof format particularly suited to intuitionistic logic. A natural deduction proof is a tree of formulas with some coin-indexing of leaves with dominating nodes. The leaf formulas are called \textit{hypotheses}: \textit{open} if not indexed, \textit{closed} if indexed. The root of the tree is the \textit{conclusion}: a natural deduction proof asserts that the conjunction of its open hypotheses entails its conclusion. A trivial tree consisting of a single formula is a proof (from itself, as open hypothesis, to itself, as conclusion, corresponding to the identity axiom of sequent calculus). Then the proofs of $\{\to, \land, \lor\}$-intuitionistic logic are those further generated by the rules in Figure 3. Hypotheses become indexed (closed) when the dominating inference occurs, and any number of hypotheses (including zero) can be indexed/closed in one step, cf. the interactive effects of weakening and contraction.

2.4 Typed lambda calculus

The untyped lambda calculus was introduced as a model of computation by Alonzo Church. It uses a variable binding operator (the $\lambda$) to name functions, and forms the basis of functional programming languages such as LISP. It was proved equivalent to Turing machines, hence the name Church-Turing Thesis for the notion that Turing machines (and untyped lambda calculus) capture the concept of algorithm.

Church \cite{Church1940} defined the simply, i.e. just functionally, typed lambda calculus, and by including logical constants, higher-order logic. Here we add also Cartesian product and disjoint union types.
Figure 3. Natural deduction rules for \{\to, \land, \lor\}-intuitionistic logic

(2) Definition (types)

The $T$ set of types is defined on the basis of a set $\delta$ of basic types as follows:

$$T ::= \delta \mid T \to T \mid T \land T \mid T \lor T$$

(3) Definition (type domains)

The type domain $D_\tau$ of each type $\tau$ is defined on the basis of an assignment $d$ of non-empty sets (basic type domains) to the set $\delta$ of basic types as follows:

$$D_\tau = d(\tau) \quad \text{for } \tau \in \delta$$

$$D_{\to} = D_{\tau_1 \to \tau_2}^{D_{\tau_1} \to D_{\tau_2}} \quad \text{i.e. the set of all functions from } D_{\tau_1} \text{ to } D_{\tau_2}$$

$$D_{\land} = D_{\tau_1 \land \tau_2} = D_{\tau_1} \times D_{\tau_2} \quad \text{i.e. } \{(m_1, m_2) \mid m_1 \in D_{\tau_1} \land m_2 \in D_{\tau_2}\}$$

$$D_{\lor} = D_{\tau_1 \lor \tau_2} = D_{\tau_1} \lor D_{\tau_2} \quad \text{i.e. } \{\{1\} \times D_{\tau_1}\} \cup \{\{2\} \times D_{\tau_2}\}$$

(4) Definition (terms)

The sets $\Phi_\tau$ of terms of type $\tau$ for each type $\tau$ are defined on the basis of a set $C_\tau$ of constants of type $\tau$ and an denumerably infinite set $V_\tau$ of
\([c]^g = f(c)\) for \(c \in C_\tau\)
\([x]^g = g(x)\) for \(x \in V_\tau\)
\(\[(\phi \psi)]^g = [\phi]^g([\psi]^g)\)
\(\[\pi_1 \phi\]^g = \text{the first projection of } [\phi]^g\)
\(\[\pi_2 \phi\]^g = \text{the second projection of } [\phi]^g\)
\(\[\phi \rightarrow y.\psi ; z.\chi\]^g = \begin{cases} 
[\psi]^g \cup \{(y,g)\} & \text{if } [\phi]^g = \langle 1, d \rangle \\
[\chi]^g \cup \{(z,g(x))\} & \text{if } [\phi]^g = \langle 2, d \rangle 
\end{cases}\)
\(\[\lambda x.\tau \phi\]^g = D_\tau \ni d \mapsto [\phi]^g\)
\(\[\phi, \psi\]^g = \langle [\phi]^g, [\psi]^g \rangle\)
\(\[\iota_1 \phi\]^g = \langle 1, [\phi]^g \rangle\)
\(\[\iota_2 \phi\]^g = \langle 2, [\phi]^g \rangle\)

Figure 4. Semantics of typed lambda calculus

variables of type \(\tau\) for each type \(\tau\) as follows:

\[
\Phi_\tau ::= C_\tau \mid V_\tau \\
| (\Phi_{\tau\rightarrow\tau} \Phi_{\tau'}) \quad \text{functional application} \\
| \pi_1 \Phi_{\tau\&\tau'} \mid \pi_2 \Phi_{\tau'\&\tau} \quad \text{projection} \\
| (\Phi_{\tau_1 + \tau_2} \rightarrow V_{\tau_1} \Phi_{\tau}; V_{\tau_2} \Phi_{\tau}) \quad \text{case statement}
\]

Each term \(\phi \in \Phi_\tau\) receives a semantic value \([\phi]^g \in D_\tau\) with respect to a valuation \(f\) which is a mapping sending each constant in \(C_\tau\) to an element in \(D_\tau\), and an assignment \(g\) which is a mapping sending each variable in \(V_\tau\) to an element in \(D_\tau\), as shown in Figure 4.

An occurrence of a variable \(x\) in a term is called \textit{free} if and only if it does not fall within any part of the term of the form \(\lambda x.\cdot\) or \(x.\cdot\); otherwise it is \textit{bound} (by the closest variable binding operator within the scope of which it falls). The result \(\phi\{\psi/x\}\) of substituting term \(\psi\) (of type \(\tau\)) for variable \(x\) (of type \(\tau\)) in a term \(\phi\) is the result of replacing by \(\psi\) every free occurrence of \(x\) in \(\phi\). The application of the substitution is \textit{free} if and only if no variable free in \(\psi\) becomes bound in its new location. Manipulations can be pathological if substitution is not free. The laws of lambda conversion in Figure 5 obtain (we omit the so-called commuting conversions for the case statement \(\cdot \rightarrow x.\cdot ; y.\cdot\)).

The Curry-Howard correspondence [Girard et al., 1989] is that intuitionistic natural deduction and typed lambda calculus are isomorphic. This formulas-as-types and proofs-as-programs correspondence takes place at the following three levels:
Logical Grammar

$$\lambda y \phi = \lambda x (\phi[x/y])$$

if $x$ is not free in $\phi$ and $\phi[x/y]$ is free

$$\phi \to y.\psi; z.\chi = \phi \to x.(\psi[x/y]); z.\chi$$

if $x$ is not free in $\psi$ and $\psi[x/y]$ is free

$$\phi \to y.\psi; z.\chi = \phi \to y.\psi; x.(\chi[x/z])$$

if $x$ is not free in $\chi$ and $\chi[x/z]$ is free

$\alpha$-conversion

$$\lambda x \phi \psi = \phi\{\psi/x\}$$

if $\phi\{\psi/x\}$ is free

$$\pi_1(\phi, \psi) = \phi$$

$$\pi_2(\phi, \psi) = \psi$$

$$\iota_1 \phi \to y.\psi; z.\chi = \psi\{\phi/y\}$$

if $\psi\{\phi/y\}$ is free

$$\iota_2 \phi \to y.\psi; z.\chi = \chi\{\phi/z\}$$

if $\chi\{\phi/z\}$ is free

$\beta$-conversion

$$\lambda x (\phi x) = \phi$$

if $x$ is not free in $\phi$

$$\iota \phi \to \iota \phi \psi = \phi$$

$\eta$-conversion

Figure 5. Laws of lambda conversion

<table>
<thead>
<tr>
<th>intuitionistic natural deduction</th>
<th>typed lambda calculus</th>
</tr>
</thead>
<tbody>
<tr>
<td>formulas</td>
<td>types</td>
</tr>
<tr>
<td>proofs</td>
<td>terms</td>
</tr>
<tr>
<td>proof normalisation</td>
<td>lambda reduction</td>
</tr>
</tbody>
</table>

Overall, the laws of lambda reduction are the same as the natural deduction proof normalisations (elimination of detours) of Prawitz [1965]. For the calculi we have given we have formulas-as-types correspondence $\to \equiv \to \land \equiv \land \lor\equiv\lor$. By way of illustration, the $\beta$- and $\eta$-proof reductions for conjunction are as shown in Figures 6 and 7 respectively.

In contrast to the untyped lambda calculus, the normalisation of terms (evaluation of ‘programs’) of our typed lambda calculus is **terminating**: every term reduces to a normal form in a finite number of steps.

2.5 The Lambek calculus

The Lambek calculus [Lambek, 1958] is a predecessor of linear logic [Girard, 1987]. It can be presented as a sequent calculus without structural rules and with single
formulas (types) in the succedents. It is retrospectively identifiable as the multiplicative fragment of non-commutative intuitionistic linear logic without empty antecedents.

(6) **Definition (types of the Lambek calculus)**

The set $F$ of types of the Lambek calculus is defined on the basis of a set $P$ of primitive types as follows:

$$F ::= P | F \cdot F | F \backslash F | F / F$$

The connective $\cdot$ is called product, $\backslash$ is called under, and $/$ is called over.

(7) **Definition (standard interpretation of the Lambek calculus)**

A standard interpretation of the Lambek calculus comprises a semigroup $(L, +)$ and a function $[[\cdot]]$ mapping each type $A \in F$ into a subset of $L$ such that:

$$[[A\backslash C]] = \{s_2 | \forall s_1 \in [[A]], s_1 + s_2 \in [[C]]\}$$

$$[[C/B]] = \{s_1 | \forall s_2 \in [[B]], s_1 + s_2 \in [[C]]\}$$

$$[[A\cdot B]] = \{s_1 + s_2 | s_1 \in [[A]] \& s_2 \in [[B]]\}$$

A sequent $\Gamma \Rightarrow A$ of the Lambek calculus comprises a finite non-empty antecedent sequence of types (configuration) $\Gamma$ and a succedent type $A$. We extend the standard interpretation of types to include configurations as follows:

$$[[\Gamma_1, \Gamma_2]] = \{s_1 + s_2 | s_1 \in [[\Gamma_1]] \& s_2 \in [[\Gamma_2]]\}$$
A sequent $\Gamma \Rightarrow A$ is valid iff $[[\Gamma]] \subseteq [[A]]$ in every standard interpretation. The Lambek sequent calculus is as shown in Figure 8 where $\Delta(\Gamma)$ indicates a configuration $\Delta$ with a distinguished subconfiguration $\Gamma$. Observe that for each connective there is a left (L) rule introducing it in the antecedent, and a right (R) rule introducing it in the succedent. Like the sequent calculus for classical logic, the sequent calculus for the Lambek calculus fully modularises the inferential properties of connectives: it deals with a single occurrence of a single connective at a time.

(8) **Proposition** (*soundness of the Lambek calculus*)

In the Lambek calculus, every theorem is valid.

**Proof.** By induction on the length of proofs. ■

(9) **Theorem** (*completeness of the Lambek calculus*)

In the Lambek calculus, every valid sequent is a theorem.

**Proof.** [Buszkowski, 1986]. ■

Soundness and completeness mean that the Lambek calculus is satisfactory as a logical theory.

(10) **Theorem** (*Cut-elimination for the Lambek calculus*)

In the Lambek calculus, every theorem has a Cut-free proof.

**Proof.** [Lambek, 1958]. ■
In the Lambek calculus, every theorem has a proof containing only its subformulas.

**Proof.** Every rule except Cut has the property that all the types in the premises are either in the conclusion (side formulas) or are the immediate subtypes of the active formula, and Cut itself is eliminable. ■

(12) **Corollary** (**decidability of the Lambek calculus**) 

In the Lambek calculus, it is decidable whether a sequent is a theorem.

**Proof.** By backward-chaining in the finite Cut-free sequent search space. ■

### 3 FORMAL SYNTAX AND FORMAL SEMANTICS

#### 3.1 Transformational grammar

Noam Chomsky’s short book *Syntactic Structures* published in 1957 revolutionised linguistics. It argued that the grammar of natural languages could be characterised by formal systems, so-called generative grammars, as models of the human capacity to produce and comprehend unboundedly many sentences, regarded as strings. There, and in subsequent articles, he defined a hierarchy of grammatical production/rewrite systems, the Chomsky hierarchy, comprising type 3 (regular), type 2 (context-free), type 1 (context-sensitive) and type 0 (unrestricted/Turing powerful) grammars. He argued formally that regular grammars cannot capture the structure of English, and informally that context-free grammars, even if they could in principle define the string-set of say English, could not do so in a scientifically satisfactory manner. Instead he forwarded *transformational grammar* in which a deep structure phrase-structure base component feeds a system of ‘transformations’ to deliver surface syntactic structures.

To emphasize the link with logical formal systems, we describe here a ‘proto-transformational grammar’ like sequent calculus in which base component rules are axiomatic rules and transformational rules are structural rules.

Let there be *modes* $n$ (nominal), $v$ (verbal), $a$ (adjectival) and $p$ (prepositional). Let there be *types* $PN$ (proper name), $NP$ (noun phrase), $VP$ (verb phrase), $TV$ (transitive verb), $COP$ (copula), $TPSP$ (transitive past participle), $Pby$ (preposition by), $CN$ (count noun) and $DET$ (determiner). Let a *configuration* be an ordered tree the leaves of which are labelled by types and the mothers of which are labelled by modes. Then we may have base component rules:

\[
\begin{align*}
[\, vTV,NP \,] & \Rightarrow VP \\
[\, vNP,VP \,] & \Rightarrow S \\
[\, nDET,CN \,] & \Rightarrow NP \\
[\, nPN \,] & \Rightarrow NP
\end{align*}
\]
There may be the following agentive passive transformational rule:

\[
[v[n\Gamma_1], [vTV, [n\Gamma_2]]] \Rightarrow S \quad \text{Agpass}
\]

Then the sentence form for *The book was read by John* is derived as shown in Figure 9. This assumes lexical insertion after derivation whereas transformational grammar had lexical insertion in the base component, but the proto-transformational formulation shows how transformations could have been seen as structural rules of sequent calculus.

### 3.2 Montague grammar

Montague [1970b; 1970a; 1973] were three papers defining and illustrating a framework for grammar assigning logical semantics. The contribution was revolutionary because the general belief at the time was that the semantics of natural language was beyond the reaches of formalisation.

‘Universal Grammar’ (UG) formulated syntax and semantics as algebras, with compositionality a homomorphism from the former to the latter. The semantic algebra consisted of a hierarchy of function spaces built over truth values, entities, and possible worlds.

‘English as a Formal Language’ (EFL) gave a denotational semantics to a fragment of English according to this design. Since denotation was to be defined by induction on syntactic structure in accordance with compositionality as homomorphism, syntax was made an absolutely free algebra using various kinds of brackets, with a ‘(dis)ambiguating relation’ erasing the brackets and relating these to ambiguous forms.

‘The Proper Treatment of Quantification’ (PTQ) relaxed the architecture to generate directly ambiguous forms, allowing itself to assume a semantic representation language known as (Montague’s higher order) Intensional Logic (IL) and including an ingenious rule of term insertion (S14) for quantification (and pronoun binding) which is presumably the origin of the paper’s title.
4 GRAMMATICAL FRAMEWORKS

4.1 Lexical-Functional Grammar

The formal theory of Lexical-Functional Grammar, LFG, [Kaplan and Bresnan, 1982; Bresnan, 2001] is a framework which takes as primitive the grammatical functions of traditional grammar (subject, object, ...). It separates, amongst other levels of representation, constituent-structure (c-structure) which represents category and ordering information, and functional-structure (f-structure) which represents grammatical functions and which feeds semantic interpretation.

The phrase-structural c-structure rules are productions with regular expressions on their right-hand side, and which have ‘functional annotations’ defining the correspondence between c-structure nodes and their f-structure counterparts, which are attribute-value matrices providing the solution to the c-structure constraints. The functional annotations, which also appear in lexical entries, are equations containing ↑ meaning my mother’s f-structure and ↓ meaning my own f-structure:

\[(15)\]

\[\begin{align*}
\text{a. } & \text{hit} : V, \ (↑ TENSE) = PAST \\
& \quad (↑ PRED) = 'hit(⟨SUBJ,OBJ⟩)' \\
\text{b. } & S \rightarrow NP \quad (↑ SUBJ) = \downarrow \quad ↑=↓ \quad \text{VP} \rightarrow V \quad NP \quad (↑ OBJ) = \downarrow
\end{align*}\]

Then *Felix hit Max* receives the c-structure and f-structure in Figures 10 and 11 respectively.

One of the first LFG analyses was the lexical treatment of passive in Bresnan [1982]. She argued against its treatment in syntax, as of Chomsky [1957]. Since
around 1980 there has been a multiplication of grammar formalisms also treating other local constructions such as control by lexical rule. More recently Bresnan’s LFG treatment of lexical rules such as passive have been refined under ‘lexical mapping theory’ with a view to universality.

Kaplan and Zaenen [1989] propose to treat unbounded dependencies in LFG by means of functional annotations extended with regular expressions: so-called functional uncertainty. Consider an example of topicalization:

(16) Mary John claimed that Bill said that Henry telephoned.

They propose to introduce the topic Mary and establish the relation between this and telephoned by a rule such as the following:

(17) \( S' \rightarrow \left( \uparrow TOPIC \right) = \downarrow S \\
\left( \uparrow TOPIC \right) = (\uparrow COMP^* OBJ) \)

Here, * is the Kleene star operator, meaning an indefinite number of iterations.

To deliver logical semantics in LFG, Dalrymple [1999] adopts linear logic as a ‘glue language’ to map f-structure to semantic-structure (s-structure), for example to compute alternative quantifier scopings under Curry-Howard proofs-as-programs. The multistratality of the c/f/s-structure of LFG is seen by its proponents as a strength in that it posits a level of f(unctional)-structure in relation to which universalities can be posited. But consider the non-standard constituent conjuncts and coordination in say right node raising (RNR):

(18) John likes and Mary dislikes London.

It seems that in view of its traditional c(onstituent)-structure LFG could not characterise such a construction without treating likes in c-structure as an intransitive verb. How could this be avoided?

4.2 Generalized Phrase Structure Grammar

Generalized Phrase Structure Grammar (GPSG; [Gazdar, 1981; Gazdar et al., 1985]) aimed to develop a congenial phrase structure formalism without exceeding
context-free generative power.

Let there be a basic context-free grammar:

(19) \[ S \rightarrow NP \ VP \]
    \[ VP \rightarrow TV \ NP \]
    \[ VP \rightarrow SV \ CP \]
    \[ CP \rightarrow C \ S \]

(20)  Bill := NP
      claimed := SV
      Henry := NP
      John := NP
      Mary := NP
      said := SV
      telephoned := TV
      that := C

To treat unbounded dependencies, Gazdar [1981] proposed to extend categories with ‘slash’ categories \( B/A \) signifying a \( B \) ‘missing’ an \( A \). Then further rules may be derived from basic rules by metarules such as the following:\(^1\)

\[
\begin{align*}
B & \rightarrow \Gamma A \\
B/A & \rightarrow \Gamma \\
C & \rightarrow \Gamma B \\
C/A & \rightarrow \Gamma B/A
\end{align*}
\]

(21) slash introduction
     slash propagation

Then assuming also a topicalization rule (23), left extraction such as (22) is derived as shown in Figure 12.

(22) Mary John claimed that Henry telephoned.

(23) \( S' \rightarrow XP \ S/XP \)

The phrase structure schema (24) will generate standard constituent coordination.

(24) \( X \rightarrow X \ CRD \ X \)

But furthermore, if we assume the slash elimination rule (25), non-standard constituent RNR coordination such as (18) is also generated; see Figure 13.

(25) \( B \rightarrow B/A \ A \)

However, if GPSG needs to structure categories with slashes to deal with extraction and coordination, why not structure categories also to express subcategorization valencies?

\(^1\)Gazdar et al. [1985] delegated slash propagation to principles of feature percolation, but the effect is the same.
Figure 12. Left extraction in GPSG
Figure 13. Right node raising in GPSG
4.3 **Head-driven Phrase Structure Grammar**

The framework of Head-driven Phrase Structure Grammar (HPSG; [Pollard and Sag, 1987; Pollard and Sag, 1994]) represents all linguistic objects as attribute-value matrices: labelled directed (but acyclic) graphs. Like LFG and GPSG, HPSG is a unification grammar, meaning that the matching of formal and actual parameters is not required to be strict identity but merely compatibility, that is unifiability.

The form (signifier) associated with a sign is represented as the value of a PHON(LOGY) attribute and the meaning (signified) associated with a sign as the value of a CONTENT attribute. Subcategorization is projected from a lexical stack of valencies on heads: the stack-valued SUBCAT(EGORIZATION) feature (there are additional stack-valued features such as SLASH, for gaps). Thus there is a subcategorization principle:

\[(26) \ H[\text{SUBCAT } \langle \ldots \rangle] \rightarrow H[\text{SUBCAT } \langle X, \ldots \rangle], X\]

where the phonological order is to be encoded by linear precedence rules, or by reentrancy in PHON attributes. See Figure 14. HPSG is entirely encoded as typed feature logic [Kasper and Rounds, 1990; Johnson, 1991; Carpenter, 1992]. The grammar is a system of constraints, and the signs in the language model defined are those which satisfy all the constraints.

HPSG can treat left extraction and right node raising much as in GPSG, but what about left node raising (LNR) non-standard constituent coordination such as the following?

(27) Mary gave John a book and Sue a record.

Since it is the head which is left node raised out of the coordinate structure in LNR it is unclear how to categorize the conjuncts and derive them as constituents in Head-driven Phrase Structure Grammar.

4.4 **Combinatory Categorial Grammar**

Combinatory Categorial Grammar (CCG; [Steedman, 1987; Steedman, 2000]) extends the categorial grammar of Adjukiewicz [1935] and Bar-Hillel [1953] with a small number of additional combinatory schemata. Let there be forward- and backward-looking types \(B/A\) and \(A\backslash B\) defined recursively as in the Lambek calculus.\(^2\) Then the classical cancellation schemata are:

\[(28) \quad >: B/A, A \Rightarrow B\]
\[<: A, A\backslash B \Rightarrow B\]

Thus:

\(^2\)CCG writes \(B\backslash A\) to mean "looks for an \(A\) to the left to form a \(B\)", but we keep to the original Lambek notation here.
Figure 14. HPSG derivation of *Felix hit Max*. 
John claimed that Henry telephoned.

Figure 15. Left extraction in CCG

Mary dislikes London.

Figure 16. Right node raising in CCG

CCG adds combinatory schemata such as the following:

\( T : A \Rightarrow B/(A\backslash B) \) type raising
\( B : C/B, B/A \Rightarrow C/A \) composition

(The combinator names define the associated semantics: \( T = \lambda x\lambda y(y\ x) \); \( B = \lambda x\lambda y\lambda z(x\ (y\ z)) \).) This allows left extraction and right node raising to be derived as shown in Figures 15 and 16 [Steedman, 1987].

Dowty [1988] observes that backward counterparts of (30) derive left node raising; see Figure 17.

However, multiple right node raising will require additional type shifts:
Figure 17. Left node raising in CCG
Likewise, combined left and right node raising:

(33) John gave Mary a book and Bill a record about bird song.
\[ N,N/PP \Rightarrow (((N\backslash S)/N)/N)\backslash(N\backslash S)/PP \]

It seems unfortunate to have to posit new combinatory schemata adhoc on an example-by-example basis. All the above type shifts are derivable in the Lambek calculus, and type logical categorial grammar takes that as its basis.

4.5 Type Logical Categorial Grammar

The framework of Type Logical Categorial Grammar (TLCG; [van Benthem, 1991; Morrill, 1994; 2010; Moortgat, 1997]) is an enrichment of Lambek calculus with additional connectives, preserving the character of the latter as a non-commutative intuitionistic linear logic. For our illustration here, let the set \( \mathcal{F} \) of syntactic types be defined on the basis of a set \( \mathcal{A} \) of primitive syntactic types as follows:

(34) \[ \mathcal{F} := \mathcal{A} | [\ ]^{-1}\mathcal{F} | \{\}\mathcal{F} | \mathcal{F} \land \mathcal{F} | \mathcal{F} \lor \mathcal{F} | \mathcal{F}\backslash\mathcal{F} | \mathcal{F}/\mathcal{F} | \mathcal{F}•\mathcal{F} | \Delta \mathcal{F} \]

We define sequent antecedents as well-bracketed sequences of types; neither sequents nor brackets may be empty. The sequent calculus is as shown in Figure 18.

The connectives \( \{\} \) and \( [\ ]^{-1} \) are bracket operators [Morrill, 1994; Moortgat, 1995]. They may be used to project bracketed domains; in our examples these will be domains which are islands to extraction. We refer to this as structural inhibition since the brackets may block association and permutation. \( \land \) and \( \lor \) are additives in the terminology of linear logic. They can express polymorphism [Lambek, 1961; Morrill, 1990]. The Lambek connectives \( \backslash, \bullet, / \) are multiplicatives. The structural operator or modality \( \Delta \) [Barry et al., 1991] licenses the structural rule of permutation and is inspired by the exponentials of linear logic.

Consider a mapping as follows from our TLCG syntactic types to the types of the lambda calculus of Section 2.4:

(35) \[ T(\{\}A) = T(A) \]
\[ T([\ ]^{-1}A) = T(A) \]
\[ T(A \land B) = T(A) \land T(B) \]
\[ T(A \lor B) = T(A) + T(B) \]
\[ T(A \bullet B) = T(A) \land T(B) \]
\[ T(A \backslash C) = T(A) \rightarrow T(C) \]
\[ T(C/B) = T(B) \rightarrow T(C) \]
\[ T(\Delta A) = T(A) \]
\[
\begin{align*}
\frac{A \Rightarrow A}{id} & \quad \frac{\Gamma \Rightarrow A}{\Delta(\Gamma) \Rightarrow B} \quad \frac{\Delta(A) \Rightarrow B}{\text{Cut}} \\
\Delta(A) \Rightarrow C & \quad \frac{\Delta([\cdot]^{-1} A)) \Rightarrow C}{[\cdot]^{-1} L} \quad \frac{\Gamma \Rightarrow A}{\Gamma \Rightarrow \cdot^{-1} A} \quad \frac{\cdot^{-1} R}{A} \\
\Delta([\cdot]) \Rightarrow C & \quad \frac{\Delta(\cdot A) \Rightarrow C}{\cdot L} \quad \frac{\Gamma \Rightarrow \cdot A}{\Gamma \Rightarrow \cdot A} \quad \frac{\cdot R}{C} \\
\Delta(A \land B) \Rightarrow C & \quad \frac{\Delta(A \land B) \Rightarrow C}{\land L} \quad \frac{\Delta \Rightarrow A \quad \Delta \Rightarrow B}{\land R} \\
\Delta(A \lor B) \Rightarrow C & \quad \frac{\Delta(A \lor B) \Rightarrow C}{\lor L} \quad \frac{\Delta \Rightarrow A \quad \Delta \Rightarrow B}{\lor R} \\
\frac{\Gamma \Rightarrow A \quad \Delta(C) \Rightarrow D}{\Delta(\Gamma, A \setminus C) \Rightarrow D} \quad \frac{A, \Gamma \Rightarrow C}{\Gamma \Rightarrow A \setminus C} \quad \frac{\Gamma \Rightarrow B \quad \Delta(C) \Rightarrow D}{\Delta(C/B, \Gamma) \Rightarrow D} \quad \frac{\Gamma, B \Rightarrow C}{\Gamma \Rightarrow C/B} \\
\frac{\Delta(A, B) \Rightarrow D}{\Delta(A \bullet B) \Rightarrow D} \quad \frac{\Gamma \Rightarrow A \quad \Delta \Rightarrow B}{\Gamma, \Delta \Rightarrow A \bullet B} \quad \frac{\Delta(A) \Rightarrow B}{\Delta(\Delta A) \Rightarrow B} \quad \frac{\Delta \Delta \Rightarrow B}{\Delta \Delta \Rightarrow \cdot B} \quad \frac{\Delta(A, B) \Rightarrow C}{\Delta(B, A) \Rightarrow C} \quad \frac{\Delta P, A \text{ or } B \Delta-ed}{\Delta B, A} \\
\end{align*}
\]
and \[ \lambda x \lambda y[y \land x] := (S\backslash[]^{-1}S)/S \]

annoy \[ \rightarrow \] 
\[ \lambda x \lambda y[y \land x] := ((\lambda)(CP\backslash)S)/N \]

felix \[ \rightarrow \] 
\[ f := N \]

from \[ \rightarrow \] 
\[ \lambda x((\text{from}_{\text{adv}} \ x), (\text{from}_{\text{adv}} \ x)) := ((CN\backslash CN) \land ((N\backslash S)(N\backslash S)))/N \]

hit \[ \rightarrow \] 
\[ := (N\backslash S)/N \]

is \[ \rightarrow \] 
\[ \lambda x\lambda y(x \rightarrow z, [y = z]; w.((w \lambda u[ u = y]) y)) := (N\backslash S)(N \lor (CN\backslash CN)) \]

max \[ \rightarrow \] 
\[ m := N \]

that \[ \rightarrow \] 
\[ \lambda x\lambda y\lambda z([y \land z] \land (x \land z)) := (CN\backslash CN)/(S/\Delta N) \]

that \[ \rightarrow \] 
\[ \lambda xx := CP/S \]

Figure 19. TLCG lexicon

Under this mapping, every TLCG proof has a reading as a proof in \{\rightarrow, \land, \lor\}-intuitionistic logic. This categorial semantics is called Curry-Howard type-logical semantics. Lambda-term lexical semantics is substituted into the lambda reading of a syntactic proof/derivation to deliver the semantics of derived expressions.

Let there be the lexicon in Figure 19. Then \textit{Felix hit Max} is derived as follows with semantics \textit{(hit max felix)}:

\[
\begin{array}{c}
\frac{N \Rightarrow N}{N \Rightarrow N} \frac{S \Rightarrow S}{\backslash L} \\
\frac{N \Rightarrow N}{N, (N\backslash S)/N, N \Rightarrow S} / L
\end{array}
\]

Left extraction such as \textit{man that John thinks Mary loves} is derived as shown in Figure 20 with semantics \textit{\lambda z[\{\text{man} \ z\} \land \{\text{think} \ (\text{love} \ z \ m) \ j\}]}.

The role of the permutation modality is to allow medial extraction such as \textit{man that Mary met today} as follows, where ADN and ADV abbreviate \textit{CN\backslash CN} and \textit{(N\backslash S)(N\backslash S)} respectively:

\[
\begin{array}{c}
\frac{N, (N\backslash S)/N, N, ADV \Rightarrow S}{\Delta L} \\
\frac{N, (N\backslash S)/N, \Delta N, ADV \Rightarrow S}{\Delta P} \\
\frac{N, (N\backslash S)/N, ADV, \Delta N \Rightarrow S}{R} \\
\frac{N, (N\backslash S)/N, ADV \Rightarrow S/\Delta N}{ADN \Rightarrow ADN} / L
\end{array}
\]

\[ ADN/(S/\Delta N), N, (N\backslash S)/N, ADV \Rightarrow ADN \] / L
The use of the bracket operators in Figure 19 marks coordinate structures and sentential subjects as islands:

(38)  a. *man that John likes Suzy and Mary loves
b. *man who that Mary likes annoys Bill

First, note how bracketed domains are induced. For, say, Mary walks and Suzy talks:

(39)  
\[
\begin{align*}
S & \Rightarrow S \\
N \Rightarrow N \quad S \Rightarrow S \\
\Delta N & \Rightarrow N \\
\Delta L & \Rightarrow N \\
N \Rightarrow N \\
N, (N\backslash S)/S, S \Rightarrow S \\
N, (N\backslash S)/S, S \Rightarrow S \\
S \Rightarrow S \\
\end{align*}
\]

And for, say, That Mary talks annoys Bill:

(40)  
\[
\begin{align*}
CP/S, N, N\backslash S & \Rightarrow CP \\
\langle \rangle CP/S, N, N\backslash S & \Rightarrow \langle \rangle CP/S \Rightarrow S \\
N & \Rightarrow N \\
\end{align*}
\]

Second, observe that the coordinator type \((S\backslash [\_1^S])/S\) and the sentential subject verb type \((\langle \rangle CP\backslash S)/N\) will block the overgeneration in (38) because the brackets projected will block the conditionalised gap subtype from associating and permitting into the islands.

5 WHY MIGHT GRAMMAR AND PROCESSING BE LOGICAL?

The formalisms we have considered have particular empirical and/or technical characteristic features. LFG: grammatical functions; GPSG: context-freeness;
HPSG: heads and feature logic; CCG: combinators; TLCG: type logic. We have traced a path leading from each to the next. Young science does not readily renounce treasured key concepts, but our own ‘logical conclusion’ of logical grammar, indeed formal grammar, is enrichment of non-commutative intuitionistic linear logic. This latter was already in existence at the time of Syntactic Structures in the form of the Lambek calculus.

One may question whether formal grammar is a good linguistic program at all. All grammars leak, and logical semantics has little to say about allegory, metaphor, or poetry. But that is not to say that grammaticality and truth conditions are not real. It seems to me that formal grammar has been tried but not really tested: after an initial euphoria, the going got heavy. But we have an opportunity to develop linguistic formalism in the paradigm of modern mathematical logic.

We conclude by considering why it might have been expected that grammar would take the form of a logic and processing would take the form of deduction. We consider the engineering perspective of language engineering and the scientific perspective of cognitive science.

On the engineering perspective, linguistic formalisms can be seen as construction kits for building formal languages which are like, or resemble, fragments of natural language. The charting of natural language syntax and semantics is then a massive information engineering task. It seems likely that logic would be a helpful tool/organisational principle for this. Indeed, if the mapping strategy were not logical, on what basis could it succeed?

Automated language processing divides mainly into parsing (computing meanings/signifieds from forms/signifiers) and generation (computing forms/signifiers from meanings/signifieds). When grammar is a logic, these computational tasks take the form of parsing-as-deduction and generation-as-deduction. The setting up of grammar as logic and processing as the corresponding deduction seems to augur well for verification: the transparency of the correctness of processing with respect to grammar.

We know something of the macroscopic and microscopic physiology of the brain, and where the language faculty is normally located; and it is usual to view cognitive processes as computations, or at least unconscious and automatic cognition such as human language processing. We want to express our cognitive theories in terms of algorithms, representations and processes eventually implemented neuronally. But there is a huge gap in our knowledge of these concepts at the level at which we want to theorise. We do not know how to define algorithms, representations or processes except in ways dependent on arbitrary features of models of computation like neural nets, RAMs, or Turing machines which we have no basis to posit as characteristic of the levels of the higher cognitive functions of our psychological theories.

Surely an eventual understanding of such concepts will come at least partly from logic. As well as with knowledge and semantics, logic has deep relations with computation (Cut-elimination, logic programming, resolution, computation as proof-search, functional programming, computation as proof normalisation). A
natural theory of algorithms, representations and processes would be one akin to logic. Pending such theory it seems reasonable to express our models of knowledge of language —grammar— at a logical level of type formulas and proof terms.

As cognitive phenomena, parsing and generation are termed comprehension and production. In TLCG syntactic structures are proofs (of grammaticality) and semantic structures are also proofs: meanings are the way in which grammaticality is proved. So interpreted psychologically, TLCG models production and comprehension as synthesis and analysis of proofs. Not just manipulation of arbitrary or language-specific structures and representations, but the resonance of logic in the dynamics of words and ideas: grammar and processing as reasoning.

ACKNOWLEDGEMENTS

This work was partially funded by the DGICYT project TIN2008–06582–C03–01 (SESAAME-BAR). Thanks to Hiroyuki Uchida for comments. All errors are my own.

BIBLIOGRAPHY


Logical Grammar


MINIMALISM

Wolfram Hinzen

OVERVIEW

Minimalism, like earlier incarnations of generative linguistic theory, is a quest for linguistic explanation. Even more than its predecessors, it pursues this overarching concern with an emphasis on explanatory factors that were largely outside the scope of the earlier theories, and that have only recently acquired a measure of promise and feasibility for the particular domain of language: general explanatory principles such as computational efficiency and related economy principles that are not specific to language or humans and may not even be specific to the organic world as such. This chapter introduces the minimalist framework with an emphasis on linguistic explanation. Section 2 characterizes Minimalism’s rationale, its vision of the structural architecture of the language faculty, its place in the generative tradition, and its descriptive apparatus and methodological basis, illustrating all of these with some linguistic analyses. Section 3 turns to philosophical aspects, relating specifically to Minimalism’s underlying philosophy of science and its idea of a ‘mind science’. Section 4 presents the standard Minimalist account of the computational system of language, elaborating on some potential problems with this account. Section 5 concludes. Throughout, I emphasize the sense in which Minimalism is, more and differently than its predecessors, also intrinsically an approach to the evolution of language.

1 MINIMALISM AS A MODE OF INQUIRY

1.1 Description versus explanation

Linguistic theory is nowhere near complete. The precise characterization of basic construction types such as passives, islands, existentials or possessives even within single languages is wide open, and there can be no doubt cross-linguistic descriptive work will continue for a long time to come. All that is in the absence of an agreement of what the overall descriptive and explanatory framework of linguistic theory should be, and in fact relatively little discussion on the issue of frameworks
among theoretical linguists, who usually simply think of themselves as belonging to one or another particular school. Despite all that — and perhaps because of it — Minimalism centrally aims to transgress descriptive work in linguistics in favour of a form of explanation that is, in a sense to be clarified, ‘principled’ and that makes us understand why the apparent laws of language are the ones they are – in short, why things fall into the cross-linguistic patterns that they seem to do. Any such attempt will naturally involve a scrutiny of the question of what these principles have been taken to be, and it will also likely lead to a re-conceptualization of many of them, hence also to new descriptive work in the study of languages. Yet, it is worth emphasizing that the attempt is to ‘rationalize’ language more than to describe it.¹

Minimalism, in short, is not a study of the facts of languages, but why they should obtain. So it is one thing, for example, to find empirically that human clauses, architecturally, fall into roughly three layers: the verbal layer (VP), the tense layer (TP), and the Complementizer layer (CP, forming part of the ‘left periphery’ of the clause); or that sentences demand subjects, an apparently universal fact of language, still largely opaque, which is captured under the name of the ‘EPP-principle’; or that locality is a crucial and universal constraint on grammatical operations. But it is a completely different question why all this should be so. This latter question crucially includes the question of how these and other structural facts are similar to those operative in other cognitive domains or else whether they are special to language, and why a language of this specific type evolved, as opposed to communication systems of many other imaginable kinds. It is in this way that Minimalism is intrinsically a project in comparative cognition and language evolution as well, in a way that no earlier incarnations of the generative project in linguistics has been.

Expectedly, if we engage in such an explanatory quest, in the beginning most apparent facts of language will simply not make any deeper sense to us. They are just that: facts. That might end the Minimalist quest, but then consider that we might ask what the facts of language should be, for them to make more sense. For example, we might conclude that, given the assumption that the computational system of language essentially links sound and meaning, there shouldn’t be any levels of representations in it that are superfluous with respect to this specific task of linking. Now suppose that, nonetheless, we had empirical reasons for the existence of such levels. Then a puzzle arises: why is what we find — the facts of language — not what we would rationally expect? That may then lead to either suspect that other, extraneous factors were involved to yield an apparently sub-optimal design, or else to conclude that perhaps we misdescribed the earlier ‘facts’. In the latter case, the task arises to re-describe them in a way that they come to fit into our picture of what the facts should rationally be.

This sort of objective has been central to minimalist practice, and in some ways it is part of rational scientific practice as such. To the extent that we suc-

¹Even though it should be conceded that description and theorizing are never strictly separable.
ceed in relevant re-descriptions, we will have arrived at a more rational account of language that goes some way to explaining why it is the way it is. None of this necessarily requires us to deny the empirical data or the explanatory principles on which mature versions of the earlier so-called ‘Principles and Parameters’ framework were based (see [Haegeman, 1994; Chomsky and Lasnik, 1993]). It is more that we will have gone some way towards deriving the earlier explanations from independently needed and more principled ones, eliminating merely technical solutions and descriptive artifacts in our theory of language.

Thus characterized, Minimalism reflects a choice of a mode of inquiry and a program of research that as such could in principle be applied to any given theoretical framework in linguistic theory (see e.g. [Culicover and Jackendoff, 2005], for a minimalist agenda in a non-Chomskyan framework), or even in other domains of naturalistic inquiry such as biology. Even though the point just made has been repeated ad nauseam, it bears emphasis once again: Minimalism isn’t itself a theory of the language faculty that as such would or could compete with other such theories. No matter one’s theoretical persuasion, a minimalist strategy of linguistic explanation is something one can choose to be interested in or not. Hence it is also nothing one could refute (even though one may well have doubts about its feasibility or point). It would thus be a serious and unproductive mistake to inextricably link Minimalism to some or other specific theoretical assumption: for everything in Minimalism is at stake.

I will now review some of the philosophical assumptions that underlie the mainstream minimalist agenda.

1.2 Language as a natural object

Even though the first paper in the generative tradition addressing an explicitly ‘minimalist’ concern dates back almost two decades [Chomsky, 1993], a concern with explanation has been central to generative grammar virtually from its inception. It is closely linked to what is perhaps its defining feature: to aim for an account of language that makes sense of it as a natural object, as opposed to a purely formal object of the sort studied in logic, or else a wholly social or conventional one. In both of the latter cases, methods of naturalistic inquiry would not obviously apply. That is, even though language is of course (used as) a communication system that as such essentially involves social and communicative aspects, language if viewed in this way would be essentially accessible through the methods of the social sciences.\footnote{The natural sciences are powerful because of the self-imposed restrictions on the domains they can deal with: only very simple domains are accessible to the methods of the natural sciences. Where domains become more complex, as in biology, we see another scientific paradigm. But domains quickly become inaccessible even to the methods of biology, say in economics or psychology. Philosophy we apply to what we understand least. Human interaction and communication is clearly at the very top of the scale of complexity, hence an unlikely start for an attempt at a naturalistic science of language. Any such attempt will have to restrict its domain and isolate out an idealized system amenable to its methodology.} That language is not only accessible by these methods but
is a subject matter of the physical sciences too, is a basic assumption of Minimalism and the generative or biolinguistic program at large. This assumption does not beg the question against those who view language as essentially a social object. It rather reflects little more than a choice of perspective: one can choose to look at anything with naturalistic eyes. This perspective implies no essentialist stipulation of what language is, and should be judged by its fruits. It is only if the view of language as a communication system is dogmatically declared to be the essential or only one that a Minimalist approach to language would seem necessarily unfeasible. So, against the claims of some Wittgensteinians that language is intrinsically social, the Minimalist maintains that while this is surely correct, systems of cognitive competence enter into the social use of language which are not as such social themselves: they reflect an inherent aspect of our mental organization as human beings. Or, against the claims of some functionalists that all positing of linguistic structures or forms must be premised by claims about linguistic functions, the Minimalist maintains the suggestion – essentially a suggestion of methodological caution — that a tight form-function relationship should not be adopted at the outset of inquiry, if at all. Form and function are conceptually distinct aspects of language that must be studied partially independently. This is what evolutionary biology independently teaches, where we see that the relationship between (organic) form and function is never one-to-one: the same form or structure will typically support different functions in its evolution and history, and the same function may be supported by different structures.

A naturalistic approach primarily focuses on the computational system of language — a system of rules and principles governing the generation of an unbounded set of complex expressions. The nature of this computational system is logically independent of its use as a communication system: though used for communication, there is no logical contradiction in imagining it to be used differently. Although communication is one function of this system, an account of the function or purpose of some natural object is not in itself (is logically distinct from) an account of the forms or the mechanisms that enable this function and its correlated adaptive benefits. We may be lucky that positing certain functions will lead us to find the mechanisms that are actually operative, but since claims about functions and claims about mechanisms are logically independent, there is no necessity to that. Moreover, given that we seem to have found apparent universals of language irrespective of certain claims about their functional rationales, and that the principles of syntax as depicted in current standard textbook accounts do not make a great deal of functional sense, there is not even a likelihood to that. Language should be regarded as logically independent of its use as a communication system also because communication does not in any way require language: no evolved communication system but the human one is a linguistic one. Communication systems could evolve and have evolved for millions of years which do not involve language in anything like the human sense. So what’s special to language is not that it is a communication system, but that it is a linguistic one. If so, the study of communication as such will not tell us much about what is special about language.
As for the view of language as a formal system, a language will then typically be identified with an infinite set of well-formed expressions, and is thus a mathematical construct. Such objects are not the subject of naturalistic inquiry and have not been the object of study in generative grammar, which instead traditionally aims to characterize a state of knowledge underlying the use of such expressions: a state of knowing (‘cognizing’) a language. Characterizing this state requires one to come up with empirical hypotheses about principles that our brain uses to generate the expressions of a language: a so-called generative procedure. The characterization of a language as a set of well-formed expressions does not determine any such procedure: necessarily, many such procedures will yield the same output (be ‘extensionally’ or ‘weakly’ equivalent). Hence the fact that a procedure yields all and only the elements of a particular designated set of expressions as output is as such of no particular linguistic significance, if linguistics is conceived as above, as characterizing a generative procedure, hence aiming at what is called ‘strong generation’. Neither does that procedure as it is implemented in human brains, whatever it turns out to be, uniquely characterize a set of expressions in the extensional sense. There is no notion of well-formed expression that applies to its output. There is a notion of the ‘last line’ of a syntactic derivation — the step-wise construction of a linguistic expression. The expression as constructed at this end of the derivation (or before that) may violate certain grammatical constraints. But even if a constraint is violated, the result is an output of the generative procedure in question. Crucially, our knowledge of language extends to expressions that diverge from some normative notion of what makes them ‘well-formed’ (e.g., they may be mispronounced, semantically deviant, syntactically deviant but assigned a coherent interpretation, pragmatically infelicitous, etc.). In human languages native speaker judgements about the grammaticality status of specific expressions are moreover typically graded: they are judgements like ‘this sounds odd’, ‘that sounds slightly better in this context’, ‘here I don’t know what to say’, etc. There is thus standardly no notion of ‘grammaticality’ of the sort that exists for formal languages in logic, and which speakers could use as a basis for such judgements. Speakers have intuitions about grammaticality for sure, but grammaticality is ultimately not an intuitive but a theoretical notion that changes as our inquiry into the nature of the human language faculty proceeds.

A formal-language approach to human linguistic competence is thus fundamentally different from a generative one, in that (i) for the latter an explanation of even some of the observed data would count for more than a mere ‘weak’ generation of ‘all and only’ the data of some designated set, and (ii) languages are inappropriately characterized as such sets. Generative grammar is not the attempt to provide rules for generating a set of all and only the well-formed expressions of

3Typically, at least, in logically inclined linguistics. From a logical point of view, nothing requires a formal approach to language to be wedded to any such construct as a set of well-formed expressions (intuitionistic logic is a good example). Generative linguistics denies the existence of languages in the sense of such a set, whose status as an empirical object and subject to naturalistic inquiry is dubious.
a language, because there isn’t such a thing as a well-formed expression within a
naturalistic inquiry into language.

Since philosophy and formal semantics has been dominated by the Fregean
formal-language approach, it is to be expected that from a philosophical point of
view, Minimalism implies an even greater change of perspective than the generative
grammar project as such. In particular, Minimalism strives, over and above
strong generation of language, for what we might call ‘natural adequacy’: explana-
tory principles should make good sense in the light of a perspective on language
as a natural object that is subject to biological and physical forces. Also, Min-
imalism aims to see language as a ‘perfect’ system, whereas, in the philosophy
of language, human language is standardly seen as highly imperfect: language is
there often viewed, not as a natural object, but functionally, as a tool. As such it
could have been certainly designed much better: a language expressly designed for
purposes of efficient communication would presumably look quite different (e.g. it
would probably lack morphological in addition to syntactic organization, and lack
transformations and ambiguity). As Frege notoriously argued, linguistic structure
is logic seen through the deficiencies of the human mind. Adopting this ‘anti-
psychologist’s stance, Russell would make a notion of ‘logical form’ a basis for
philosophical logic that was crucially non-linguistic. Wittgenstein’s philosophy is
similarly a fight against the way in which natural language ‘bedevils’ our minds.
It includes the recommendation to give up any strivings for logical perfection and
simply put up with whatever language happens to be good for. Linguistic com-
plexity and its treatment in the emerging generative tradition accordingly played
very little role in 20th century post-war analytic philosophy, not e.g. in Carnap’s
system, where logic alone (and not, in particular, grammar) was declared apriori
and an intrinsic aspect of the mind. The behaviorist tradition departing from
Quine and Davidson could make no better sense of the generative naturalistic
treatment of language. Quine’s and Davidson’s verdicts against the possibility of
‘analytic’ knowledge — knowledge depending on the language faculty alone, as
opposed to world knowledge — in particular stand in a stark contrast to many
decades of concrete empirical suggestions in generative grammar for what is ana-
lytic, i.e. what are the specifically linguistic and innate principles that guide the
organization of meaning (see [Chomsky, 2000], for comments on this issue).

Given the above, adjudicating between the views of language as a public medium
and views of language as a computational system seems wrongheaded. The per-
spectives and goals involved seem too different. Adjudicating between language
as based on ‘Universal Grammar’ (UG) and so-called ‘usage-based’ (UB) accounts
[Tomasello, 2003] may well be wrongheaded too. UG is not a theory but an area
of inquiry, in which, again, one can choose to be interested in or not. That area is
the initial state of the language faculty in a child’s brain at the onset of language
acquisition, whatever it is.4 There clearly is such state at a time when the child

\[^{4}\text{Is it question-begging to call this initial state a state of a ‘language faculty’? Surely the mechanisms underlying island violations, split infinitives, unergatives, or other such standard objects of linguistic inquiry have been described in a vocabulary proprietary to that particular}\]
doesn’t yet know a language, while having the structural preconditions to acquire one. Moreover, it is universal (and hence these structural preconditions are), because any child can learn any language, with equal ease. The issue thus can only be *how rich* the initial state, and hence its theory (UG) is, but not whether it exists, or whether language acquisition is based on it. Few would endorse the view today that it is an empty, ‘tabula rasa’ sort of structure. It rather appears to be a richly structured cognitive state whose mechanisms at least appear to be domain-specific and specific to humans (for crucial qualifications see later). It provides testable restrictions on what a possible (natively acquirable) human language is; within these restrictions, linguistic cultures differ.

Accordingly, some account of UG such as Government and Binding (GB) [Haege-man, 1994] can, like any other theory, be wrong about the specific universal principles it posits (and indeed Minimalism says it is). But it is not clear where in its naturalistic methodological assumptions or basic concepts it might be methodologically flawed, or where principled questions of legitimacy arise. A UG-account of language, moreover, though not conceived as such by proponents of ‘UB’, is a ‘UB’-account, in the sense that its theory is expressly designed as an account of how grammar can be acquired from experience (i.e., usage): structure in the initial state is posited on the very basis of considerations of what the data of language are, what the final competence is like, and how the child can possibly and plausibly bridge the gap between the two. The difference between GB and Minimalism in this respect is only that while GB asked *how rich* the initial state must be for this gap to be bridged (and has generally stressed how rich it is), Minimalism asks the same question with a different emphasis: with *how little* structure in the initial state we get away with in the light of what the child’s input is and what cognitive state has to be attained [Chomsky, 2008b]. The less structure was needed, the more the question of language evolution will become a potentially tractable research program.

In naturalistic inquiry, we will quite generally expect a shift of focus away from the attempt to cover data to an attempt to find laws that explain them, and beyond that to criteria of simplicity and elegance in the choice of some such explanatory theory. These criteria form a central part in the minimalist incarnation of generative grammar, but they are not foreign to earlier versions of that program. Striving for more restrictive theories that meet demands of learnability and, above that, demands of elegance and parsimony, are in particular not unfamiliar from its predecessor, the Principles & Parameters (P&P) framework. What distinguishes the two frameworks is that while P&P aims at the characterization of a state of knowledge, and thus wants linguistics to associate with psychology and ultimately biology, Minimalism’s basic explanatory principles – economy principles with a least effort and minimality flavor – have a much more natural interpretation as
principles that we take to guide the physical world more generally.\(^5\)

In the so-called ‘biolinguistic’ program as a whole [Lenneberg, 1967; Jenkins, 2000], linguistics is taken to be continuous with biology, but the question what biology (and life) precisely is, and how it relates to physics, is of course quite open. Indeed, language, as a subject matter for biology, is a highly unusual and isolated object in the living world, and this may well make us suspect that principles may govern its origin and structure that are not characteristic of the main explanatory principles used in the ‘Neo-Darwinian Synthesis’. This is the synthesis of the Darwinian theory of evolution with the early 20\(^{th}\) century science of molecular genetics. It reduces the study of evolution to the molecular-genetic and population-genetic study of how organisms gradually change their structure over time in response to the adaptive pressures they are exposed to in a competitive environment. Given the peculiarities of language as a natural object, it is a widely open question whether this sort of study is likely to illuminate it. Independently of that, it is a virtual truism that natural selection selects adaptive variants against the constraints that a physico-chemical channel provides for what variants are possible. It seems not only possible but plausible that principles of this origin should be instrumental for the emergence and structure of language, a point to which I return.

In the light of the essential continuity of the generative tradition with regard to its basic aims, which I have stressed here, it may be surprising to note a persistent criticism of the generative enterprise, which appeals to its lack of continuity and repeated revolutions.\(^6\) Such criticisms reveal a misunderstanding of the generative tradition, and as far as the present chapter is concerned, it is important to note that there really isn’t any revolution to be reported here (see, on the contrary, [Searle, 2003]). The generative project at large is stunning for the persistence with which it has given shape to a basic rationalist vision in the analysis of human knowledge, and for the way in which it has sought to integrate it into visions of explanation in biology and natural science at large. The quest is to characterize language as a natural object and as an aspect of human nature in something like Hume’s [1739] sense, a central query being which principles are special to this cognitive domain and uniquely human [Chomsky, 1957; 1959, Hauser et al., 2002].

\(^5\)P&P in turn was based on a rejection of the view that the theory of grammar can be thought of as a science like engineering (or computational linguistics): the formulation of the aim of generative linguistics, to generate ‘all and only the well-formed expressions of a language’, is misunderstood as a task of engineering. Success in this generation task is of no great interest to the project of generative linguistics, which is about generative procedures, not sets of expressions, and which does not operate with a notion of a language as a set of well-formed expressions.

\(^6\)See [Lasnik, 2000] for a particularly telling textbook account of the generative project that starts with *Syntactic structures* [Chomsky, 1955] and ends on a continuous path with a discussion of Chomsky [1993].
1.3 The rise of economy principles

P&P itself was born in the early 1970s with concrete suggestions for how to move, from language-specific rule systems generating particular construction types of single languages, to general and more abstract overarching principles (so-called ‘conditions on rules’, [Chomsky, 1973]) that would constrain possible rule systems for any human language. Their interaction and parameterization would then explain (allow us to derive) the construction types in question, without further language- and construction-specific rule-systems needed. In other words, construction types — from passives to raising constructions, interrogatives, or sentential complements — exist (it’s not that P&P would explain them away), but they are emergent phenomena following from something else: more general underlying principles such as the option of ‘displacing’ a constituent to another position in the sentence, an option that cuts across the construction types in question (thus, in particular, all of questions, topicalizations, raising-constructions, and passives are essentially derived by displacement or ‘movement’). The basic problem with generative rule systems had been noted already much earlier: the enormous complexity of such systems raised a principled problem of how such systems could actually be learned in a finite human time. True, with the human brain assumed to harbor a Turing-machine architecture, any computable linguistic construction can in principle be arrived at by some rule system or other: achieving ‘descriptive adequacy’ in this sense is a technical problem, a problem of computer engineering. But achieving it is therefore also something that is obtainable, as it were, too cheaply. Any such technical solution is without explanatory scope if the basic problem remains unsolved, the problem of telling how the child arrives at the right grammar — in the effortless and unfailing way it does. The answer was to abstract out overarching principles from the construction- and language-specific rule systems and to view them as a set of constraints that are part of the biological toolkit that the child brings to bear on the task of language acquisition. That task is then primarily one of figuring out the remaining language-specific features of a language, by setting the values of a number of universal parameters, which are defined points of variation that interact with language-universal principles to yield the properties of a particular language. Ideally, there are then no construction- and language-specific principles in the sense of primitives at all. Language acquisition can be viewed as an environment’s selecting from or triggering parametric options that UG freely provides. In a nutshell, the child doesn’t learn language, but which language is spoken in its environment (see [Yang, 2002], for an updated version of essentially this perspective).

Note that the shift from rules to principles, which is continued in mainstream Minimalism, directly reflects the generative tradition’s basic interest in explanation. Yet again, Minimalism is not wedded to this shift. Indeed, what was wrong with rules was not that they were rules (as opposed to abstract principles) but that they were construction-specific and language-specific. If they were not, there could be no Minimalist objection to them, and indeed in recent years they have promi-
nently returned, in a non-construction-specific shape [Epstein and Seely, 2002]. One might even be able to construct an argument that they should, everything else being equal, return, and it is useful to elaborate on this argument because it nicely illustrates a typical piece of Minimalist reasoning.

Say we have reduced the rule component of the grammar to one single transformational operation, ‘Affect $\alpha$’ [Lasnik and Saito, 1992]. This was conceived as a maximally unrestricted operation: it allowed doing anything to the syntactic category $\alpha$ in the course of a derivation at any time, including inserting, moving or deleting it, as long as no sub-theory or component (‘module’) of UG would forbid the ensuing result. The consequence of this move is a system that is maximally simplified in its operations and that compensates for this simplicity through the constraints incorporated in various modules of the grammar, among them Theta-theory (which sieves out structures violating constraints on the thematic structure of lexical items), Case-theory (requiring arguments to have cases), and Binding Theory (capturing anaphoric and pronominal dependencies), plus the constraints provided by some few principles, such as ‘Subjacency’, ‘cyclicity’ conditions (illustrated below), or the ‘Empty Category Principle’. Quite plausibly, the Lasnik and Saito model is the best that P&P had to offer, and its purest incarnation. But although we satisfy an aim for maximal simplicity by having a theory like that, one feels that a much better design would be if structures that are to be sieved out by certain constraints later on in the derivation wouldn’t be produced in the first place, since producing them would, as it were, be vacuous and a waste of computational resources. Put differently, a ‘crash-proof system’, in which rules can only build structures that meet the constraints, hence where structures failing certain constraints cannot exist, would seem to meet ‘best design’ principles much more convincingly. Thus, we are not surprised to find the feasibility of the ‘crash-proof’ program to be a question at the forefront of current Minimalist inquiry (see e.g., [Frampton and Gutmann, 2002]).

Although the field has not unanimously accepted this particular line of argument, it has made a move in a direction equally opposite to that of Lasnik and Saito, by adding avoidance of operational complexity as a design principle to the grammar, an example of what is called an economy principle in grammar. Operational complexity means that rules may only apply under ‘last resort’ conditions: they apply not freely, but only if forced. Chomsky’s formulation of this economy principle in [1995:200] reads:

\begin{quote}
(1) Last Resort:
A step in a derivation is legitimate only if it is necessary for convergence.
\end{quote}

That is, had the step not been taken, the derivation would have ‘crashed’. A derivation is said to converge, if it violates no constraints imposed on the output.

\footnote{See Putnam (ed.), forthcoming, for book-length discussion. What should also be mentioned here are arguments rejecting the ‘derivational’ approach just discussed, in favor of a ‘representationalist’ model of the grammar, where all constraints on derivations are re-coded as constraints on representations (see [Brody, 2003]).}
of the grammar by the outside systems with which the linguistic system interfaces. The most important of these constraints is ‘Full Interpretation’:

(2) **Full Interpretation:**
Representations at the interfaces must be fully legible.

Conditions on ‘legibility’ will be relative to what the external systems are. Minimally, however, workable linguistic representations will have to be of such a form that they can be converted by the Phonetic Component (PC) into a structure that can be used for externalization (be it vocal communication, as in spoken languages, or visual communication, as in sign languages), and by a Semantic Component (SC) into a structure that functions in thought, reference, planning, and intentional language use more generally. One might argue, in fact, that, optimally, these two ‘interfaces’, which are called SEM and PHON, respectively, not only minimally exist, but maximally as well: by a minimalist logic, ‘minimally two’ should actually mean, ‘exactly two’. In particular, there should be no other levels of representation in the grammar than interface representations: representations that have an immediate interpretation in terms of external systems.

Last Resort will explain, for example, why, in the structure (3),

(3) \( \Delta \) seems [John to be nice]

(where \( \Delta \) indicates a position in the structure that will eventually have to be filled), ‘John’ will have to move to make the structure well-formed:

(4) John seems [t to be nice]

Here ‘t’ (a ‘trace’) indicates a position that is phonetically null and is located in the launching site of this movement, i.e. the ‘base’ position of ‘John’ (the position of ‘lexical insertion’). Assume that ‘John’ in (3), like any other nominal, has a (nominative, NOM) Case feature to check against a relevant Case-assigner. If a verb assigns NOM to a nominal, and that nominal has a NOM-case feature, the two features are brought to a match and play no other role in the derivation: they are eliminated. This is good because Case is a feature that at least appears to serve a purely grammatical function without having an interpretation in either PC or SC at least in English: not in PC, because English has a severely impoverished system of phonetically overt Case-marking; not in SC, because structural Case (NOM and ACC) does not seem to make a difference to semantic interpretation.

---

8Note that this particular formulation of Last Resort makes reference to convergence at the interfaces — the endpoint of the derivation — and hence involves a computationally suboptimal form of ‘look-ahead’. This problem, however, is not endemic to the idea of Last Resort as such. Thus, e.g., if movement of a syntactic object \( \alpha \) to a syntactic object \( \beta \) is conditional on the latter object containing a feature that needs to be checked against one in the former, one might argue that movement is forced without reference to convergence (on feature-checking see right below). The phase-based framework mentioned later also contributes to restricting look-ahead.

9Thus e.g., it makes no difference whether we say *He believed him to be an idiot*, or *he believed that he is an idiot*, where *he* and *him* differ in (overtly marked) Case. In either case the person referred to is the subject of the embedded clause (that is, the idiot).
If convergence at the interfaces in the sense of (1) drives grammatical processes, Case must be checked before the derivation reaches the semantic interface — else Full Interpretation would be violated. Now, Case assignment doesn’t happen in the clause where ‘John’ is first generated (=lexically inserted), because this clause is non-finite (its Tense is not either present or past). The movement of the NP John saves the structure because the finite verb ‘seems’ has an (abstract) nominative Case feature against which that same feature of ‘John’ can be checked.

Under Last-Resort conditions, we will now also predict that a structure like (5), where the verb of the embedded clause is finite, hence where Case can be checked in the embedded clause, ‘John’ need not, hence, by (1), cannot, move. This prediction is born out by the ungrammaticality of (6):

(5) ∆ seems that [John is nice]

(6) *John seems that [t is nice]

‘Last Resort’ illustrates both the departure from the P&P model as driven to the limits by Lasnik and Saito [1992], and an idea very central to the Minimalist Program: the idea that requirements of the (semantic and phonetic) interfaces — so-called ‘bare output conditions’ — drive syntactic processes: the syntax is in this sense externally conditioned and explainable from the way it fits into the structural context of the rest of the mind. A related economy principle is the Minimal Link Condition (cf. [Chomsky 1995:311]):

(7) **Minimal Link Condition:**

K attracts α only if there is no β, β closer to K than α, such that K attracts β.

This principle is illustrated by ‘superraising’ examples such as (8) (see [Chomsky, 1995:295-7]):

(8) ∆ seems [that it was told John . . . ,

in which the node that ∆ occupies would correspond to K, it to β, and John to α. Two things could happen in this derivation to satisfy the requirements of the inflected matrix verb ‘seems’: ‘it’ could move so as to yield (9), or ‘John’, so as to yield (10):

(9) it seems [that t was told John . . . ]

(10) John seems [that it was told t . . . ]

In (9), ‘it’ satisfies the requirement of the matrix verb to have a nominal in its subject position, but not its requirement to check nominative Case: for, by our above reasoning, the relevant Case checking has taken place in the embedded clause already. So we might predict that ‘John’ will move instead, which is both nominal and has an unchecked Case feature (passive verbs do not assign accusative Case). But given the intermediate position of ‘it’, this move would be longer than allowed
by (7): although the Case feature of ‘it’ is lost, its move would be a legitimate one, hence must take place in preference to the movement of ‘John’. If a ‘shortest move’ requirement of the sort that (7) expresses is part of the grammar, we can thus explain why (8) does not and cannot converge by any continuation of the derivation.

A final idea of an essentially historical importance in Minimalism illustrating the same kind of economy principles is the Shortest Derivation Requirement (see [Collins, 2001, 52] for discussion):

(11) **Shortest Derivation Requirement:**

Of two derivations, both convergent, prefer the one that minimizes the number of operations necessary for convergence.

This is a principle of *derivational economy*, as opposed to the *representational economy* of the principle of Full Interpretation, (2), which bans interpretationally superfluous elements of representations. (11) would make little sense if it would apply unrestrictedly to all derivations, for then, if it was operative, the derivation in which nothing happens would always win. The question is thus which of a number of convergent derivations that are built from the same collection of lexical items (called a Numeration), economize on the number of steps needed until convergence. This principle is prima facie rather unappealing, as it implies a form of ‘look ahead’: numbers of steps are counted relative to the need of convergence at the last line of the derivation. The comparison of derivations it invokes leads to an explosion of computational complexity and has been heavily criticized for that reason (see e.g. [Johnson *et al.*, 2000a; 2000b; 2001]). This critique misfires, however, for one thing because computational explosion is ‘bad design’ only on the subsidiary assumption that the linguistic system is designed for use. But it need not be. Language as a system may be of vast computational complexity and be used, while those parts of it that cannot be used simply are not used (see [Uriagereka, 2002:ch.8] for further discussion of this and related issues). Put differently, what we make of language is *as such* no indication for what its structures truly are. The use of the system may be the result of the restrictions of the usability of the system that its interfacing cognitive systems impose. Setting aside this methodological problem of the critique, however, the objection is mute for another reason. The idea of global comparison of derivations has essentially been given up in mainstream current Minimalism, along with the associated idea of ‘look-ahead’. In line with much earlier considerations of the ‘cyclicity’ of derivations, to which I return, Chomsky [2001] introduced the idea that a derivation is ‘by phase’: it proceeds in relatively small computational cycles within which a portion of a Numeration (initial selection of lexical items) is first used up before a new cycle can begin (for considerations against look-ahead see also [Chomsky 2001a/2004]). For other more recent concerns with notions of ‘efficient computation’ in the language faculty, see [Collins, 1997; Epstein, 1999; Epstein *et al.*, 1998; Epstein and Seely, 2006; Frampton and Gutmann, 1999].
1.4 Single-cycle generation

Above we encountered a ‘best design’ consideration to the effect that the grammar, if it has to have interfaces, as it clearly does, necessarily has to have at least two of them, and thus ideally should have exactly two. In the light of this, let us consider the basic ‘Government & Binding’ (GB) model of grammar, itself the mature form of the ‘Y’ model stemming from the so-called ‘Extended Standard Theory’ (EST) (Fig. 1):

This model consists of 5 independent generative systems. The operations they consist of connect distinct ‘levels of representation’ in the linguistic system. These are single (unified) syntactic objects that are legitimate (only) if satisfying a list of constraints. For example, at D-structure, the thematic structure of the verbal head of a clause must be fully specified: the theta (for ‘thematic’) roles (AGENT, PATIENT, INSTRUMENT, GOAL, etc.) that a verb obligatorily assigns to its arguments must be discharged by appropriate arguments. D-structures are assembled by selecting a number of items from the mental Lexicon and putting them into configurations satisfying the laws of ‘X-bar theory’. The latter in particular stipulate that the configurations in question must be ‘hierarchical’ and ‘headed’: as for hierarchy, phrases have constituent structure, in the sense that some categories contain others or have them as parts; as for headedness, at least one of the syntactic objects that combine must ‘project’, i.e. label the resulting complex...
object (see [Speas, 1990]). A V-N combination, for example, will be either verbal — a Verb Phrase (VP) — or nominal — a Noun Phrase (NP), and couldn’t just become a Prepositional Phrase (PP), for example. The rules constructing a D-structure moreover operate cyclically, in the sense that there is an intrinsic order to how things happen in a derivation: what happens early in a derivation cannot depend on what happens later in it, say when the derivation has proceeded. For example, filling in the external argument of the verbal phrase after completing that phrase would be a counter-cyclic operation (assuming the verbal phrase to constitute one such cycle). We may more abstractly think of cyclicity as a ‘compositionality’ principle, in the sense that a complex combinatorial object is determined by its parts and their structural configuration. Determination in this sense logically requires that the parts do not in turn depend on the whole (or on what is larger than themselves). On the determination of what the relevant cycles are — called phases in recent Minimalism — see [Frampton and Gutmann, 1999; Chomsky, 2001a/2004; 2007; 2008a; 2008b].

All other four computational cycles are wholly or partially compositional in this sense. The second cycle in Fig. 1 is the mapping from D-structure to S-structure, the result of transformational operations displacing constituents to other positions in the phrase marker assembled at D-structure. Thus e.g., (13) would be the D-structure of (12), (14) its S-structure, the structure that in GB captures the level after displacements have occurred:

(12) noone seems to be in the cupboard

(13) ∆ seems ∆ to be noone in the cupboard

(14) noone, seems t_i to be t_i in the cupboard

In (13), noone is in its base position, the same position where it appears overtly in (15)

(15) There seems to be noone in the cupboard

Without there, however, there is nothing to fill the positions in (12) that the symbol ∆ indicates must be filled in order for the structure to be well-formed. The so-called EPP-requirement (‘Extended Projection Principle’) in particular stipulates that all clauses need a subject — the subordinated clause [to be noone in the cupboard] as much as the matrix clause [seems [to be noone in the cupboard]]. In the absence of there, in (14) noone must move to satisfy the EPP-requirement for each clause. It does so ‘successive-cyclically’, by first moving to the embedded subject position, and then to the matrix subject position (again, first satisfying the matrix position would be a ‘counter-cyclic’ operation).

(14) is an example where a displacement of a syntactic object is overt: even though interpreted as the agent of the event of being in the cupboard, in the phonetic output it is not heard there, but only where it is displaced, in the matrix-subject position. It became a standard GB-assumption, however, that not all
movements to satisfy certain requirements need to be overt in this sense. Thus, e.g., in Chinese, where quantified NPs never appear displaced overtly (they stay in situ, i.e. in their base position), certain expressions have properties of a kind which we would predict them to have precisely if the NPs in question would move in these expressions as well, albeit not overtly, or without a phonetic reflex. The suggestion thus was made that our theory of grammar would be simplified by assuming that movement, as a phenomenon effecting syntactic levels of representation, need not affect an expression’s phonetic form. This gives rise to a third compositional cycle, leading from S-structure to a new level ‘LF’ (for Logical Form), where the same computational operation ‘Move α’ applies, though now after the point in the derivation where this matters for the phonetic output [Huang, 1995]. But the S-structure has of course also to be mapped to the interface with the sound systems — the phonetic component, PC — which gives us a fourth cycle. And finally, there is the fifth cycle, the mapping of an LF-representation to a semantic representation SEM of the thought expressed by a sentence, as constructed in the semantic component, SC.

Now, a system with five cycles seems biologically possible, but the Minimalist question will immediately arise: ‘Why five?’ The problem aggravates in the light of the fact that there appear to be great redundancies in the mappings involved, and that essentially the same operation — the basic combinatorial operation, Merge — applies in all these independent generative systems. In short, best design considerations would suggest, in the absence of contrary evidence, a single-cycle generation [Chomsky, 2005; 2007; 2008a; 2008b]. The boldness of this vision seems somewhat tantalizing given the complexities of Fig.1, but it is now clear that its possibility cannot be dismissed, and its basic underlying idea will now be explained.

In a single cycle architecture there are no levels of representation at all that are internal to the grammar — that is, levels of representation at which grammatical conditions apply but which are not interface representations that all and only answer output conditions. S-structure and D-structure have been widely assumed to be grammar-internal levels of representation in precisely this sense, and thus they should not exist. Minimalism has risen to this challenge and re-considered the empirical arguments that were put forward in favour of these levels, in an effort to argue that the relevant data follow from independently motivated conditions that make weaker assumptions. Among the strongest data arguing for a D-structure level, in particular, is the distinction between raising and control constructions, but as Hornstein et al. [2005, chapter 2] carefully argue, this distinction follows quite easily on the basis of one single simple stipulation, namely the derivational requirement that arguments cannot move to the syntactic positions where they receive (or get assigned) their thematic roles. Other functions of D-structures can be taken over by an analysis of how the basic structure-building mechanism (Merge) works. As for S-structure, why should it be a level of representation? Conceptually necessary about it is only that at some point in the grammar, the derivation must branch, in order for phonetically interpretable material to be separated from
semantically interpretable material. Ideally, then, S-structure should just be a branching point, or perhaps several such points [Uriagereka, 1999], but not a level.

This dispenses with cycles 1 and 2. As for LF, as long as we distinguish it from SEM, the semantic interface itself, it is still grammar-internal. Better, then, would be if the derivation feeds into the SC directly, without the assumption that prior to the computation of a semantic representation a single syntactic representation of the LF-sort is assembled and then subjected to conditions of well-formedness pertinent to that level. If there is no empirical reason to assume that such a single syntactic object exists (as [Chomsky, 2001; 2001a; 2004] argues), the weaker option that LF in earlier models is no more than a descriptive artifact should be accepted. This deals with cycles 4 and 5. What remains? A single cycle, departing from the lexicon, LEX, that feeds directly into the interfaces, with recurring points of ‘Transfer’ where the structure constructed up to those respective points is shipped off to the PC (a sub-process of Transfer called ‘Spell-out’) and the SC (‘phases’ being the objects that are so transferred). Neither an LF nor a PF is ever assembled as such, their only analogs in the system being the single and never unified chunks of structure constructed within a phase before they are shipped to phonetic and semantic interpretation:

![Diagram of the model](image)

Figure 2.

In comparison to the earlier GB model depicted in figure 1, this model, whose basic structure and minimality couldn’t have been contemplated as a possibility even a few years back, clearly looks much less baroque. More positively, it is a paradigm of beauty and parsimony (even if false). Unless prevented to accept it on the basis of some extremely recalcitrant data, it would have a greater explanatory force. Not only would it be the basis for deducing the relevant data on entirely minimal theoretical and structural assumptions, it would also depict the language faculty itself as ‘optimally designed’, as meeting rational design expectations.

---

10Such a separation seems necessary since primitives and relations in phonetic form are strictly different from those in semantic form. Conceptual-intentional systems on the other side of the semantic interface wouldn’t know what to do with motor instructions to the speech articulators.
What exactly that idea and the methodology behind it means will be the subject of the next sub-section.

1.5 Methodological vs. substantive minimalism

Minimalism not only continues, as we have seen, the basic strategy of the P&P tradition seamlessly, but also recalls some of the earliest considerations regarding a ‘simplicity’ criterion for evaluating grammars (hypothetical rule systems) for a particular language [Chomsky, 1955]. Simplicity in that sense was not only the standard notion of simplicity that is broadly used as a desideratum for good theories in the core natural sciences. It also was intended as an evaluation measure for grammars that was internal to linguistic theory. When comparing two grammars, the ‘simpler’ one of them, in some defined linguistic sense, should win, as long as conditions of descriptive adequacy and other such external conditions were met. The idea was not an actual discovery procedure for how the child would, using simplicity as a criterion, arrive at the right grammar for a language in its environment. It was assumed that the child could arrive at grammars of languages in all sorts of — largely ill-understood — ways (intuition, guess-work, partial methodological hints, etc.). Rather, the goal was to ‘provide an objective, non-intuitive way to evaluate a grammar once presented, and to compare it with other proposed grammars’ [Chomsky, 1957, 55-6], where all grammars are assumed to be a priori generated within a range of admissible ones (namely, grammars conforming to the laws of UG). By this simplicity criterion, for example, a grammar that contained transformations was arguably ‘simpler’ than a grammar containing none.

The idea is continuous with Minimalism, except for a crucial difference, hinted at towards the end of the last sub-section, namely that over the decades simplicity has come to be applied not merely to theories of language, but the very object that these are about — the language faculty itself. Language itself, as our object of study, the new idea is, may satisfy principles of design optimality not even suspected in the early years. The reader might wonder whether the difference between simple theory and a simple object is not perhaps too thin. But note that a theory may be simple or convoluted crucially no matter what the nature of the object is that it is about: if the object violates all rational expectations of how it should be (as happens often in biology, for example in the design of the eye), we will still aim for the simplest theory of it. It could have been, e.g., that brain structure is so convoluted, inscrutable, and complex, that we would give up all hopes of developing a deeper understanding of it. Again, even in such a case, we would still aim for the optimal theory of it. Interestingly, however, there are indications that at least in this particular case this pessimistic conclusion need not be drawn: Chris Cherniak’s ‘best of all possible brains’ hypothesis points to

---

11 There are other critical points of agreement between Minimalism and some of the earliest work in generative grammar, for example the view that a syntactic transformation is just a copying operation: take a syntactic object and copy it somewhere else (see [Chomsky, 2004, 153] for comments).
exactly the opposite conclusion, where the optimality in question is really one of the object itself, not merely its theory.\textsuperscript{12} There is some basis then for saying that the optimality of the theory and the optimality of the object must be conceptually distinguished. The distinction is vital to Minimalism, as it has been in other fields in the natural sciences. In sum, with Minimalism, broadly methodological applications of the simplicity criterion have shifted to what is called \textit{substantive Minimalism}, the attempt to show that the language faculty itself is — surprisingly — an instance of design perfection. That is, the facts of language are ultimately the result of a very small set of computational operations that are ‘minimal’ in the sense of being the minimum needed for language to be used. A ‘minimal’ theory, in this sense, may, again very surprisingly, also be a theory that satisfies all other tests of adequacy. Substantive Minimalism is what we saw at work in the previous section.

We are now ready to state Minimalism’s basic explanatory strategy. Clearly, if a natural object satisfies expectations of optimal design — it is as we would rationally expect it to be, or as a rational designer would have made it who devised it from scratch — we also understand \textit{why} it is the way it is. We will say: the object simply has the minimal structure — the structure needed to be usable at all by outside systems — and no more structure besides. Call this the Strong(est) Minimalist Thesis:

\textbf{(16) Strong Minimalist Thesis (SMT):}

Language is an optimal solution to the task of satisfying conditions on legibility.

If this thesis made sense (were true to any significant extent), there would, in a sense, be no further questions to ask. Questions arise where an object does \textit{not} make rational sense, not if it does. We will, of course, have to ask questions about the language faculty’s \textit{development} in evolutionary history, a question not answered by synchronic considerations of its intrinsic design; about its implementation in the brain; and about the mechanisms that use it to various purposes. Yet, in a sense, the basic ‘why’-question — of why language is the way it is — will be answered. We will, in particular, not have to look at the role of contingent history and specific evolutionary pathways to explain why language does not adhere to some vision for how it should be. This is true, of course, only as long as our metric of ‘optimality’ (minimality, elegance, etc.) is itself not conditioned by considerations of history. In some cases, this may well be relevant. For instance, Carstairs-McCarthy [1999] argues that the evolutionarily earlier structure of the syllable was a crucial internal determinant for the evolution of clause structure. Relative to the historically contingent circumstance in which the syllable evolved,

\textsuperscript{12}Cherniak develops this hypothesis departing from the principle ‘save wire’ within given limits on connection resources in the brain [Cherniak, 1995; 2005]. It is a central function of the brain to connect, and given computer models of how such wiring might be perfected, the finding is that refinement we actually find in the brain is discernible down to a ‘best-in-a-billion’ level. Conway Morris [2003, ch.2], raises similar ‘best of all possible genomes’ considerations regarding the structure of DNA. There are many other examples.
the SMT can still be explored. That said, in the general case, considerations of contingent history are not the sort of considerations that feed our intuitions when ‘Galilean’ perfection in nature is what we seek to establish. The guiding intuition here is that principles guide the maturation of the language faculty which are no different in character from those guiding the growth of a crystal. A crystal grows by natural law subject to the environmental parameters that provide the context within which these laws apply. In this regard, the minimalist vision is like the Pre-Darwinian vision of organic forms as immutable natural forms or types that are no different in kind from inorganic ones and built into the general order of nature. This perspective does not conflict with a neo-Darwinian adaptationist outlook: the afunctional givens of physical law will necessarily interact with the conditions of historical circumstance. But we don’t expect the latter to greatly distort the canalizing influence of physical law. Nor is this ‘Pre-Darwinian’ perspective in any way foreign to contemporary biology, as for example a recent paper of Denton et al. [2003] suggests, who point out that

when deploying matter into complex structures in the subcellular realm
the cell must necessarily make extensive use of natural forms (…) which like atoms or crystals self-organize under the direction of natural law into what are essentially ‘pre-Darwinian’ afunctional abstract molecular architectures in which adaptations are trivial secondary modifications of what are evidently primary givens of physics.’ [Denton et al., 2003]; see also [Denton, 2001]

Note, on the other hand, that the idea that language is a case of natural design perfection is something noone expects. But Minimalism need not be vindicated for it to make sense as a research program. It rather raises the question how strong a ‘minimalist thesis’ can be entertained — that is, to what extent it might actually be true that there is something ‘minimal’ about human language design. Exactly insofar as this is not the case, we will have to invoke external conditions that influenced the evolution of language, be it some of the contingencies of genetic evolution (genetic ‘tinkering’), unknown facts about the architecture of the brain, or something else. But that is to say that a partial or even a total failure of the SMT will likely have taught us something about human language: why it is not perfect in the way we expected.

There is thus nothing mystical in Minimalism’s proclaimed aim to vindicate principles of ‘design perfection’ in the domain of language, and to go beyond the goal of ‘explanatory adequacy’ that P&P had set [Chomsky, 2004]. Explanatory adequacy had meant no more than that we understand how a language can be learned from the data available to the child alone. The answer was: by the child’s biological endowment with a set of principles and parameters whose values it can easily set in the light of the data in question. If our aim is the deepening of our understanding of the mechanisms fulfilling this explanatory task, the general idea of vindicating principles of design optimality appears to make good sense.
1.6 Structural and functional design perfection

When speaking of optimality many readers will at first have expected, or still now expect, that principles of functional optimality are intended, of the sort that adaptationist biologists will have to at least partially assume, when they claim particular traits to be designed by natural selection. It may even seem unclear what other notion of design optimality could actually be intended in Minimalism, if not some such notion of functional optimality — optimality with respect to some task. Moreover, in Minimalism there is a task with respect to which language design is heuristically hypothesized to be optimal, as noted: the task of optimally satisfying the external conditions imposed on the language system for it to be usable by systems it is interfacing with.

To see this idea more clearly, imagine a pre-linguistic hominid with something like the cognitive organization of a modern ape: a system of conceptual understanding, with whatever capacities and limitations it has, is in place, hence presumably a semantics, perhaps encoded in some kind of ‘language of thought’. Undoubtedly, then, a designer of the language faculty would design the latter so as to be usable by this conceptual system (this is what was termed ‘legibility’ above). The same holds for a partially pre-linguistic system of sound production. As noted, whatever other structures the language system will have to have, it will minimally have to satisfy usability constraints at its meaning and sound interfaces (SEM and PHON, respectively). Mapping structures to these interfaces means to link them, for language is a mapping between sound and meaning. In the limit, therefore, language would be also no more than a linking system: it is not independently ‘innovative’ for either sound or thought, but just links them, hence externalizes forms of thought given as such beforehand. This clearly is an ideal of functional optimality.

Yet, this is not quite the kind of functional optimality an adaptationist biologist will typically have in mind: the minimalist question is not in particular how useful language is relative to various purposes for which one might use it in an external environment, but how optimal it is as a system (‘organ’) interacting with a number of other systems already in place — systems external to the language-faculty, but internal to the mind. Let us call this an ‘internalist functionalism’. Importantly, a system optimal in this sense may still be sub-optimal if it comes to its use with regards to various purposes to which language is put, e.g. communication or pronunciation. In short, while optimal in the internalist-functional sense with respect to at least one of its interfaces, it can still be quite sub-optimal in an externalist-functionalist sense and with respect to its other interface. From a minimalist perspective, this is quite expected, for as Chomsky has long since argued, there is no particular empirical reason to regard language as being ‘optimized for use’ in the externalist’s sense. A quite imperfect instrument of communication,

---

13 I take it that a pre-linguistic ‘language of thought’ may have a quite different structural format than actual human language: I see no structural necessity, in particular, that there would be propositional kinds of meanings or thoughts expressed by such a ‘language’.

14 There is no general biological or other reason why languages made available by the language
indeed, is what the Fregean tradition in philosophy at large has taken language to be. In short, while it is true that we use language and that it is useful, and while it may also be true that the language faculty was partially selected for (after it existed) because of the communicative uses it has, these facts need not be intrinsic to it. As noted, language may, in the internalist sense, be a perfect system, even if only partially usable. We would simply use the parts that are usable. But those that are hard to use or unusable may then still consist of expressions that are optimal in the sense of satisfying interface conditions optimally, with a minimum of computational resources (for example, a syntactically and semantically immaculate expression may simply be too long or complicated to ever be used by a creature with a memory capacity as we have it).

That language is not optimized for its communicative use, which may well have postdated its emergence, might also be suggested by the fact that traces are phonetically null. They are causes of great difficulties in processing sentences, hence is a sub-optimal feature of language in this sense. Thus, e.g., (17) is bordering deviance:

(17) who do you wonder whether John said solved the problem?

(17) has a perfectly coherent semantic interpretation, however:

(18) who is such that you wonder whether John said that he solved the problem?

The deviance of (17) thus illustrates one of the typical constraints that abound in language and that we don’t know how to explain functionally or semantically. The difference between (17) and (18) is that the latter contains a resumptive pronoun (‘he’) in the position where who is interpreted (base generated), a position indicated by a trace t in (19):

(19) who do you wonder [whether John said [t solved the problem]]?

An answer to (19) could be (20):

(20) I wonder whether John said Harry solved the problem.

In other cases, traces are not recoverable at all. Consider (21), where the moved ‘how’ is meant to question the mode of the problem-solving of Harry, not of the wondering or the saying:

faculty should be fully accessible (...). The conclusion that languages are partially unusable, however, is not at all surprising. It has long been known that performance systems often “fail”, meaning that they provide an analysis that differs from that determined by the cognitive system [of language] (...). Many categories of expressions have been studied that pose structural problems for interpretation: multiple embedding, so-called “garden-path sentences,” and others. Even simple concepts may pose hard problems of interpretation: words that involve quantifiers or negation, for example. Such expressions as “I missed (not) seeing you last summer” (meaning I expected to see you but didn’t) cause endless confusion. Sometimes confusion is even codified, as in the idiom “near miss,” which means “nearly a hit,” not “nearly a miss” (analogous to “near accident”) [Chomsky, 2000, 124]. See also Chomsky [2008] for the ‘primacy’ of the semantic interface over the phonetic one.
(21) how do you wonder [whether John said [Harry solved the problem \(t\)]]

A semantically perfectly coherent answer to this question would be:

(22) I wonder whether John said Harry solved the problem [in three minutes]

But (21) is again deviant. Why then is it that traces, even if their phoneticization _would_ ease interpretation and hence communication, are not heard in overt speech? An answer could be: because language did not evolve for communicative use, hence is not optimal with respect to it. The language faculty economizes on representations and derivations that yield elegant sound meaning pairings, even if this makes processing harder. The entire phonetic side of language may be a later development in the evolution of the language faculty, and that prior to its externalization language was used for thought alone. That line of thought may also derive plausibility from the fact that we find all regularity and universal aspects in language (of a sort that reflect ‘rational design’ conditions) in the computational system of language which generates syntactic derivations on the path from LEX to SEM (‘narrow syntax’). By contrast, it is much less clear whether we find such features in the morphological and phonetic aspects of languages, aspects that are moreover widely held accountable for the surface differences among languages. If the computation of PHONs is governed by different kinds of design constraints than that of SEMs, we might conjecture that the peculiarities of the phonetic channel is something that language design had to accommodate to as well as it could, while existing before that as a system used for thought alone.

Having emphasized the distinction between design ‘for a use’ and the minimalist quest for design ‘to be usable (at all)’, we should nonetheless ponder the fact that minimalist optimality is still a functionalist one, which moreover has an _externalist_ character in the specific sense that we centrally take language to answer conditions imposed by systems external to it (the ‘bare output conditions’ mentioned above). We assume, that is, that the system is nothing in itself, as it were: it is not ‘creative’, does not generate new structures of its own, but rather reflects the structures of what is given to it. This very contention is built into yet another putative ‘best design’ condition, namely that the grammar satisfies a condition of _Inclusiveness_ [Chomsky, 1995, 228]:

(23) **Inclusiveness:**

No new objects are added in the course of the computation.\(^{15}\)

If (23) holds, a derivation begins with a set of lexical items (these being sets of phonetic, semantic, and syntactic features), and the derivation does nothing other than _re-arranging_ these items and their features. If so, the syntax is banned from doing anything that is not already contained in lexical items.

\(^{15}\)As Michiel van Lambalgen notes, using Hilbert’s program in meta-mathematics as an analogy, this restriction on the system may be somewhat unmotivated and problematic: mathematics does not work like this. Hilbert attempted to reconstruct the system of arithmetic from a finite initial base. But proofs in this system will make reference to the infinite, for example, and Hilbert’s program failed.
I shall critically return to precisely this aspect of Minimalism in the next section. But it is worth pointing out already here that there is a sense of structural perfection that is different from and more attractive than any functionalist notion of design optimality, including the internalist sense of the mainstream Minimalist program that I have just described: a good example for perfection in this sense is again Cherniak’s research on neural component placement mentioned above. This is perfection in structural organization, Cherniak argues, which, despite being formidably functional, has no functional rationale. It comes “for free, directly from physics”, i.e., [is] generated via simply exploiting basic physical processes, without intervention of genes’ [Cherniak, 1995]. Functionality disappears as a rationalizing factor for design that we find: it plays no explanatory role, which indeed we do not expect it to play in physics or mathematics. If we found factors from within these fields operative in the organization of language, the way we would understand it would have no functional dimension at all (even though, again, such design could be eminently functional and lead to appropriate selective episodes in the evolutionary history of the organ in question).

Structural perfection in this sense we normally expect in the physico-mathematical domain alone, but it is not confined to it. A well-known example is sunflower phyllotaxis, where proportions of left- and right-turning spirals consist of neighboring numbers of the famous Fibonacci series (generated by the principle that any Fibonacci-number is the sum of the previous two: 1, 1, 2, 3, 5, 8, 13, 21, . . . ), whose ratio converges to the golden angle [Uriagereka, 1998]. Again, this design is eminently functional and adaptive, as it prevents leaves from ever shielding others from the influence of the sun. But this doesn’t explain the constraint in question — it doesn’t explain how this pattern comes to exist. Mitchison [1977] provides such an explanation, which derives the pattern as a mathematical consequence of principles of stem growth and lead placement (see [Amundson, 1994] for illuminating discussion). Mitchison’s explanation of the origin of form in this case is a paradigmatic example of an internalist one that contrasts with both the standard externalism of adaptationist biology — organismic morphology is rationalized as a consequence of competitive organism-environment interactions — and the functionalism that is implicit in the minimalist idea of ‘motivating language from interface conditions’. I return to these different senses of design perfection in section 4.

2 METATHEORETICAL AND PHILOSOPHICAL ASPECTS

2.1 The ‘Galilean style’ in linguistics

Ultimately, the Minimalist project is nothing other than the project of a ‘natural philosophy’ begun in the early 17th century by Descartes, and it is ‘rationalist’ in much the same sense (see [Hinzen, 2006]), with the considerations of design perfection above possibly representing a peak in rationalist theorizing about the mind. This project’s rationalism affects its methodology of scientific inquiry as much as
its epistemology. Epistemologically (i.e., in its underlying theory of knowledge), it departs from a ‘naturalism’ that views basic forms of higher cognition as forms of nature, that is human nature. It can therefore make no more sense of the ‘justifiedness’ of basic forms of knowledge than Plato or Descartes could: knowledge (of mathematics, grammar, music, etc.) is essentially there by nature, hence no more subject to a normative assessment or ‘justification’ than any other natural object could be. Being a creature of a certain kind, certain things will seem obvious or intelligible to us, with no further justification possible than to say that this is how our minds works.

The anti-psychologistic tradition in philosophy influenced by Frege would consider much of this unintelligible, as the mental aspects of human beings, naturalistically conceived, are a mere aspect of ‘psychology’ for this tradition: the structure of thought is logic, and if made the subject matter of empirical inquiry, all we can see is logic as transpiring through human limitations. In short, the mental as studied empirically is necessarily a distortion of the true nature of thought, which as such is something abstract and mind-independent. The mind is then no natural object but a failure-prone representational medium or machine. But the objection of psychologism may be missing a point. The basic structures of logic, mathematics, music or language characterize the specifically human mind if anything does: our mind is, of its nature, a device that can grasp these things, apparently in contrast to most other minds that arose in nature. That, as Frege claimed, the study of logic should not be conceived as the study of how we happen to factually think is consistent with that. The study of generative grammar is also not the study of how we happen to talk (this is the competence-performance distinction). It nevertheless is the study of a naturally occurring generative system that, together with a myriad of other cognitive systems, enters into the way we talk and think.

The verdict of ‘psychologism’ issued against the sort of enterprise introduced in this chapter has made philosophy move away from the naturalistic study of the mind, leading instead to a theoretical focus on language as a medium of communication, a system of norms, a game, an instrument serving for purposes of the representation of what is true in the world, or for expressing language-external or ‘public’ constructs (e.g., mind-external propositions). Much of philosophical logic and the philosophy of language is based on such conceptions to this day, and the generative enterprise, by virtue of its non-functionalist perspective on the mind, remains an isolated rationalist endeavour in a philosophical half-century that has virtually exclusively favoured empiricist, pragmatist, or hermeneutic trends. An internalist perspective on the intrinsic structures of the mind (and the question of ‘mind design’) seems essentially absent in current reflections, or is regarded as extraneous to philosophy. This is a sharp departure from the naturalistic outlook of the early modern philosophers, where logic, language, morality, aesthetics, etc. were all regarded as intrinsic features of the human mind.

As for the rationalism in its philosophy and methodology of science, Minimalism exhibits a quasi-Galilean reliance on abstract formal models of nature as something
that alone provides us with a grip on how it works. The Cartesian rationalist does not reason from effects to causes by inferring the latter from the former (as on an ‘inductive’ model of scientific inquiry), but deductively, from causes to effects. Effects are data that we expect to fall into place if we depart from the right abstract ideas or mathematical models, which show their value or truth in the course of this process — but it doesn’t really matter where these ideas or models come from. For all the rationalist cares, they come from the caverns of our minds, from nature itself, or from God. In the Minimalist Program, too, we reason, from a model of how language should be, to effects or data that this model would predict. If the data are then different, we try to fit them into the model; only if they are too recalcitrant do we give up on the rational standard that our model sets, and go for something that makes a priori less good sense. But data, generally speaking, are not what a theory is about. It is about what they follow from — and understanding this in the case of some data can be better than a maximum of data coverage. At the outset of a theory, Chomsky keeps reminding us, we may often want to discard a great deal of data that fly in the face of our theory, much as Galileo couldn’t explain why, on this theory, we don’t fall off the earth, despite the fact that this would seem a rather important piece of data. In Minimalism too, at first, ‘all the phenomena of language appear to refute it’ [Chomsky, 2002, 124]. But science does not work in the way that theories are trashed in the light of ‘facts’ alone. Facts are of our own making, they are generated by theories with concomitant biases. If facts appear to refute a theory, it is standard scientific practice at least in the physical sciences to also evaluate the facts in the light of the theory, rather than merely the theory in the light of the facts: the facts may not have been described correctly, there may be another way of looking at them that preserves and perhaps enhances explanatory power, and what we are calling ‘facts’ may be the overall result of too many factors whose confounding effects we can’t dissociate. In this sense, science is not the representation of reality, but a form of experimentation with structures — explanatory constructs, models — that we find occurring in our minds, by a process that Peirce called ‘abduction’. The reason that science consists in doing experiments is that representing facts is as such of no particular interest: facts are the result of many confounding factors, and experiments have the function of isolating those that one’s theory says one should be interested in. As Ian Hacking has put this general point strikingly:

‘One chief role of experiment is the creation of phenomena. Experimenters bring into being phenomena that do not naturally exist in a pure state. These phenomena are the touchstones of physics, the keys to nature, and the source of much modern technology. (...) Most of the phenomena, effects, and events created by the experimenter are like plutonium: they do not exist in nature except possibly on vanishingly rare occasions.’ [Hacking, 1982, 71-2]16

16For elaborations on rationalist method, see [Hinzen, 2006, especially section 2.2].
Some of the resistance to the Minimalist Program in linguistics and philosophy may also derive from the conviction that linguistics is more like psychology than mathematical physics, or even better, a form of engineering [Dennett, 1995, ch.13]. In both cases, the ‘Galilean’ style will be viewed as a doomed and misdirected effort, and another scientific aesthetic will reign. For Galileo, who inaugurated the aesthetics underlying Minimalism, nature only produces ‘perfect’ objects, it being our task to figure out in what way they are perfect. Nature, ‘generally employs only the least elaborate, the simplest and easiest of means’ (quoted in [Chomsky, 2002, 57]). In just this way, the Strongest Minimalist Thesis (SMT) entertains the idea that language design is economical and minimal, guided by ‘least effort’ principles.

This aesthetic — that nature, deep down, is fundamentally perfect and simple, with all complication and convolutedness being a hint for a lack of understanding on our part — is a basic constant in the history of modern (particle) physics, and perhaps its identifying rationale: beauty is a guide to the truth, just as ugliness in a theory points to error and human misperception. Hence theories can also be selected for elegance, and not merely because they are true. Whatever the origin of this aesthetics, and whatever the reason it works, this aspect of Minimalism should not as such be surprising; surprising should be its application in a domain where we would hardly have expected it to bear any potential scientific fruits: language, a system that so many are tempted to simply look at as a conventional and imperfect object that can exhibit essentially arbitrary features, subject to constraints on communication and convention alone. Even if language is viewed as a natural (as opposed to merely conventional) object, there is, perhaps indeed some hybris in the bold attempt to create a ‘Galilean’ science of the language. After all, this is a central component of the most complex organ in the most complex creature that we know of, a recent and isolated product of evolution that is at the core of our humanity and the basis for our culture and history, and that we barely begin to be understood in its relation to molecular and physiological levels of description.

2.2 Philosophy of mind

As explained above, Minimalism is an attempt to characterize the mind and its organizing principles on a model of scientific understanding that takes its inspiration from physics. It is thus worth asking in what sense exactly we are here talking about the ‘mind’, and to what extent we do so differently in Minimalism that in earlier versions of the generative project. In what sense are we endorsing a ‘physicalism’ or ‘materialism’ here? The short answer is that we aren’t, and that

17Similarly Descartes [1641/1988], for whom it is obvious imperfections in God’s creation — in human cognition, for example, where our conceptual capacity and faculty of judgement could for example be much more perfect, given how easily we fall into errors — that cry out for a special explanation: why does error and falsehood exist in the first place? (Cf. the third of the Meditations.)
this is because post-Newtonian science itself adopts no such view. But let us proceed more slowly. Chomsky’s practice and recommendation for many decades has been to speak about the mind in an essentially informal and ontologically neutral sense. That is, when talking about ‘mental’ aspects of organisms we are doing no more than singling out a domain of inquiry, just as we would when talking about the organism’s chemical, optic, or visual aspects. In the intended informal sense, there uncontroversially are such mental aspects of organisms: understanding a language, seeing colours, missing someone, or regretting one’s body weight, are some of these. It seems that engaging in such activities will involve internal mechanisms of the organism as much as external ones, the latter relating to the organism’s embedding in a wider physical and social world. We simply expect that a part of what accounts for language use — our overall explanandum — are internal structures in the organism that are due to this particular organism’s nature. To entirely dispense with those — to treat the organism as essentially a ‘black box’ whose internal structure does not matter since it is entirely malleable by external forces — is the contrary strategy pursued by Skinner [1957]. In this regard, Chomsky’s anti-Skinnerian ‘internalist’ stance [Chomsky, 1959; 2000] is a view that would seem no more controversial than a rejection of a radically externalist Skinnerian account would be when applied to bodily organs: obviously, no biologist would seriously entertain the view that organic development is wholly due to organism-external factors. What organs develop in ontogeny is not primarily a function of what physical forces act on us, but also of what creature we are. Nobody, also, would defend a ‘nativism about plants’. Such a nativism is basically assumed in biology, and yet it is widely held that it is in need of a thorough defence in psychology and the study of human nature. This reflects a methodological dualism in the study of human and non-human nature, or mental and non-mental aspects of nature. In this way, we can say that Chomsky’s internalism amounts to little more than the rejection of a methodological dualism: the view that radically different principles must apply to the study of ‘physical’ and ‘bodily’ aspects of an organism, on the one hand, and its ‘mental’ ones, on the other. Assuming a broad methodological monism (naturalism), rather, we hope our account of ‘mental’ organs to pattern along with that of ‘bodily’ ones. In fact, we would deny — empirical differences between the respective scientific domains notwithstanding — that a principled ontological distinction between these domains should even be made.

In short, we start out from a metaphysically neutral position — neutral in particular on (and hence consistent with) the issue of metaphysical (Cartesian) psycho-physical dualism. That order of things, as opposed to the order that places metaphysical contentions prior to scientific conclusions — is in itself a post-Kantian imperative (see [Friedman, 1993]), and perhaps characterizes the scientific revolution and modern philosophy more generally. From Descartes onwards, the struggle for a ‘natural philosophy’ precisely was a struggle for a philosophy unpremised by empirically unsupported metaphysical-scholastic speculations. Descartes’ ‘method’ [Descartes, 1637] would forbid appeal to any other explana-
tory tools than reason alone and experiment: his mechanical conception of matter and motion was not a metaphysical view but manifests a rationalist standard of explanation and of intelligibility that the new scientific method had set. Cartesian dualism was a consequence of a form of scientific inquiry committed to these very tools. In contrast to this, the materialist (or, as it has been preferentially called: ‘physicalist’\(^{18}\)) consensus in the second half of 20\(^{th}\) century philosophy of mind has not been a consequence of scientific facts. There was no specific discovery in the brain sciences that led U. T. Place to proclaim the mind/brain identity theory in 1956, or that triggered subsequent developments in philosophy leading up to functionalism, anomalous monism, instrumentalism, or eliminative materialism. It rather seems fair to say that the truth of some sort of materialism and the undesirability of Cartesian dualism — to this day regarded as a paradigm of an ‘anti-naturalistic’ and ‘anti-scientific’ stance (see \cite{Meixner, 2004}) — was mostly a starting assumption of the contemporary philosophy of mind, the question not being whether materialism was right, but how it could be \cite{Place, 1956}. Note that the trivial observation that mental phenomena crucially depend on the brain was of course both as such accessible to Descartes, and consistent with his specific dualism.\(^{19}\)

Chomsky’s basic strategy, then, which combines an internalism with a metaphysical neutralism and methodological naturalism is very different from the one that has characterized much of 20\(^{th}\) century philosophy of mind. It is, moreover, intellectually interesting to see that metaphysical naturalism (physicalism) often went along with a methodological dualism: specifically, a veto against the applicability of naturalistic inquiry to the mind. Physicalism is the ill-defined view that physical facts are all the facts there are — in particular, mental/psychological facts must in some way really be physical facts — where the notion of the ‘physical’ is indexed in essence by physics as it is (rather than by a potential future physics, that, through some conceptual revolution, comes to account for the mental by new laws). Given this basic physicalist assumption, a methodological dualism can for example take the form of the contention that the mental does not exist, hence cannot be studied \cite{Churchland, 1981}, or if it exists, must be functionalizable or consist in certain input-output relations that a computer could perform too; that it falls outside of natural science because there are no natural laws to be found in this domain at all \cite{Davidson, 1980}; that it is essentially normative and social, and in these respects not natural or physical \cite{Bennett and Hacker, 2003; Kripke,}

\(^{18}\)The latter term has been preferred to the former because of the unclarity of the notion of ‘matter’. A ‘modern’ materialism, it is thought, must simply take its departure from whatever it is that physics says exists. It is unclear for reasons outlined below whether this move succeeds in making materialism more coherent (in essence, because post-Newton the notion of ‘the physical’ is an all-encompassing notion).

\(^{19}\)There is still a persistent presumption that an ontological independence of the mental is actually inconsistent with the known laws of nature. This is surprising in light of the fact that most of those who defend versions of that independence today are actually physicists rather than philosophers, whose usual doctrine of ‘physicalism’ usually forbids any kind of dualism \cite{Stapp, 2006; Thompson, 1990; Barrett, 2006; Penrose, 1994}.
1980;Voltolini, 2003]; or that it requires an autonomous science independent of physics with its own distinctive kinds of laws, to which biology and physics may even be somewhat irrelevant [Block, 1995]. Note that all of these views must be classified as anti-naturalism from the viewpoint of those who wish to regard mental phenomena as aspects of nature among others and apply the tools of naturalistic inquiry to them (this anti-naturalism of much contemporary philosophy is what [Chomsky, 2000], centrally explores). It is worthwhile reflecting on how much the early modern project of ‘natural philosophy’, for which a methodological naturalism is essential, is really still a minority endorsement. Indeed, philosophical misgivings about the general methodology of generative grammar as a science of the mental (see e.g., [Devitt, 2003] continue with Minimalism, sometimes with a renewed invigoration [Rey, 2003; Searle, 2003].

Defending the legitimacy of a methodological naturalism as applied to the study of language, however, is one thing. Quite another is to say something more positive about the metaphysics of generative linguistics, hence to move beyond metaphysical neutrality. Can Minimalism support any metaphysical conclusions at this point? If physicalism is asserted today, this is taken to mean that the putative mental facts that there clearly are — speaking a language, enjoying a melody, thinking about one’s holiday — either cannot exist, or, if they exist, must, in philosophical parlance, ‘supervene’ on physical facts. Supervenience is a strong form of metaphysical dependence, suggesting that two seemingly different layers of reality — here, the mental and the physical — are in some sense actually a single layer (idealists would let the physical supervene the mental, physicalists the mental on the physical). If supervenience is endorsed, then, if one were to make an inventory of things that exist, mental ‘things’ do in principle not have to be mentioned. For example, in such an inventory, a bathtub probably won’t have to be mentioned, as it is fully given if all the molecules it is made up of are given, and if they are arranged in one specific fashion. There simply seems to be no way for the molecules and their arrangement to be given, and yet for the bathtub not to exist: the bathtube, in a strong sense, reduces to the molecules and their arrangement. In a similar way, a picture on a TV-screen consists of a particular arrangement of pixels. There is no aspect of this picture that can change independently of what happens at the level of pixels and their arrangement. Everything, then, one might propose, ultimately supervenes on the physical. Given that the doctrine seems intelligible enough, what is it that Chomsky [2000] finds unintelligible in it? Indeed, he has gone on record for decrying Cartesian dualism as incoherent and unintelligible as well.

Plainly, and as a point of pure logic, the coherence of Cartesian mind-body dualism as much as the supervenience doctrine just described depends on some determinate notion of what ‘body’ (or ‘matter’, ‘the physical’) is. Indeed, in Descartes’ case that was a very specific notion. Here is what Descartes tells us about what he means by a body:

‘whatever has a determinable shape and a definable location and can occupy a space in such a way as to exclude any other body; it can be
perceived by touch, sight, hearing, taste or smell, and can be moved in various ways, not by itself but by whatever else comes into contact with it.’ [Descartes, 1641/1988:81]

The notion is clear enough. If you then think about how, by contrast, we think of our thoughts, say the thought that there are seven sources of true beauty, it seems we simply don’t conceptualize these as being three inches long, say, or as weighing five ounces, as smelling badly, as visible, as making a noise, or as being caused mechanically as opposed to arising freely and creatively. In short, it seems that thought has properties that no physical object ever studied by a physicist had, such as that it has intentionality and can be assessed as true or false. From this point of view, the assumption that a science of the mind is outside the scope of a physics based on contact mechanics seems plausible indeed. However, this entire notion of body of course was trashed by Newton shortly after its inception. At the end of the very century that had given birth to the ‘mechanical philosophy’, post-Newtonians came to accept that bodies exert a gravitational force on other bodies without being in physical contact with them. How their matter gives rise to such effects was left unanswered by Newton, and little has changed since then either with respect to our capacity to address such ‘how’-questions or our willingness to endorse theories that leave them open. Gravity remains the least explainable of all presently understood physical forces, and quantum mechanics has added further frustrations in our quest to make good conceptual sense of what nature is like. Physics has played havoc with our common sense intuitions and we have become accustomed to the lower standard of intelligibility that it has set, in comparison with the vision of mechanistic explanation that so centrally characterized the earlier Galilean and Cartesian frameworks.

Strikingly Descartes, sticking to an intuitive notion of physical matter, had defined extension in space as the one essential properties of physical bodies. But the post-Newtonian Jesuit Boscovich already, in one of the first systematic systems of physics after Newton, declared matter to be actually unextended (see [Yolton, 1983]). Hume would later echo the spirit of this fundamental sceptical crisis in regard to our conceptual understanding of the world, when arguing that both extension and solidity belonged to matter’s ‘secondary’ properties. Locke added that ‘matter can think’, thereby establishing that there was no coherent notion of matter on which to base materialism: at the very least, any such notion would have to exclude mental properties. As a consequence, materialism stopped being a metaphysics that could be said to underlie natural science, and no notion of ‘the physical’ as required for a contemporary ‘physicalism’ can repair this fundamental problem in turning either dualism or materialism into a coherent doctrine again. It is unclear on what scientific basis our notion of a physical ‘supervenience base’ could rest (cf. [Lewis, 1994]).

What are the consequences of this for our practice in developing a science of the mind? Shunning metaphysical biases on what ‘matter can be’, we will simply proceed to describe the ‘mental’ aspects of the universe as good as we can, attempting to collect a ‘body of theory’ for them in the same way as Newton had
developed for what he called ‘those properties of gravity’, leaving the unification problem aside. It is that attitude that Joseph Black, in a situation in the history of chemistry in many ways analogous to that of today’s linguistics, defended when writing:

‘let us receive chemical affinity (…) as a first principle, which we cannot explain any more than Newton could explain gravitation, and let us defer accounting for the laws of affinity, till we have established such a body of doctrine as [Newton] has established concerning the laws of gravitation’. (cited in [Chomsky, 2002, 54])

Generative linguistics as understood by Chomsky has followed precisely this course: developing a body of theory, rather than to engage in metaphysical speculations (‘reduction’, ‘supervenience’) at the outset of inquiry. However, there is an important difference between chemical and mental aspects of the universe: in particular, there is not even the beginning of a ‘body of doctrine’ comparable to Newton’s on what the Cartesians perceived as the essential creativity of language use (the problem of why humans say or think what they do, when they do). The more that observation holds, musings over the ‘reduction’ or ‘supervenience’ of mind may seem pointless and anachronistic.

There are two ways in which the Minimalist phase in generative grammar speaks to these conclusions. First, Minimalism has raised an entirely novel issue and research program for the philosophical reflection on mind and human nature: the question of human mind design [Hinzen, 2006] and the more specific question of its ‘perfection’, which raises novel concerns in the study of human evolution. Secondly, reflections on the metaphysical implications of generative grammar should not stop at the overall Chomskyan conclusion that we arrived at in the last paragraph. These considerations stop with stating a ‘unification problem’, adding the suggestion that we should no more translate ‘lack of unification’ into ‘dualism’ than chemists should have done so in the 19\textsuperscript{th} century. But Hinzen and Uriagereka [2006] argue that while it seems entirely appropriate to pursue a physicalist inquiry into the nature of the gravitational force or chemical bonding, to pursue such inquiry for thoughts might be as nonsensical as pursuing it for numbers or complex topologies. To start with, there seems to be nothing incoherent in this stronger conclusion, as Chomsky’s stance implies. But, more positively, even a unification problem for linguistics as construed on analogy with the chemical one in the 19\textsuperscript{th} century points to a ‘duality’ that we have at present no reason to expect to go away. Even if a unification were to happen, we would if anything expect current linguistics with its ‘body of theory’ to provide constraints for how it does, rather than to be eliminated in the course of this unification (put differently, the new unified theory would have to be true of or to encompass the mental aspects of nature that linguistics theorizes about). Finally, the ‘stronger conclusion’ just mentioned may actually be supported by empirical considerations arising in the

---

\textsuperscript{20}This is not a form of ‘Platonism’ as elaborated and defended by Katz [2000]. Language for Katz is a mind-external object.
course of Minimalist inquiries. Consider ‘jumps in representational order’ in the arithmetical system. Number systems exhibit such jumps in the following sense: the natural numbers 1, 2, 3, . . . , form an algebra that is closed under the operation of addition, for example, definable in terms of the successor function. But as we go about enumerating this sequence of objects, we find that we can also perform the inverse of the basic algebraic operation we started with, namely subtraction, a decision that takes us into a new ontological domain, the integers, which asymmetrically entail the ontological domain we started with. In a similar way, inverting the operations of multiplication and exponentiation by performing division and roots opens up new mathematical spaces with inhabitants such as rational, irrational, and complex numbers, again with concomitant relations of asymmetric entailment between the arising ontological layers. This hierarchy doesn’t end with complex numbers, beyond which we have quaternions and octonions. So the process is evidently productive — let us call it ontologically productive: it yields new kinds of objects that we never hit upon when applying linear operations within a given vector space, say when generating the natural numbers.

The standard machinery for deriving a syntactically complex linguistic expression however is *not* taken to be ontologically productive in this sense. In fact, the process (see section 3 for details) seems to be essentially viewed as one of adding ever more brackets, as in (24), which give rise to ever deeper embeddings, without any jumps in representational order of the kind just illustrated:

(24) [. . . [. . . [. . . [. . . ]]. . . ]]. . .]

But why should there be no ontological productivity in the linguistic system, analogous to the arithmetical one? Hinzen and Uriagereka (2006) suggest a much closer analogy between arithmetic and language than Chomsky’s frequent evocations of this parallelism allow (e.g., [Chomsky, 2008a]). The basis for this suggestion is that languages, in their ‘parts of speech’ systems,

i. exhibit universal hierarchies that do not seem to follow from anything other than the syntax (hence, in particular, not from the semantics associated to them), and

ii. these hierarchies exhibit asymmetric entailments that are of a necessary character.

It would then follow that if

iii. nothing other than the type of hierarchy we find in the numbers provides an analysis for the linguistic hierarchies in question (and their necessity),

iv. the issue of ‘unification’ of these linguistic hierarchies with physical matter would indeed be as nonsensical as the unification of numbers or mathematical topologies would be.

As for an example of one such linguistic hierarchy, consider (25), where ‘<’ stands for ‘less formally complex than’:
(25) The nominal hierarchy:

abstract < mass < objectual/count < animate

The noun beauty would illustrate an abstract nominal space (i.e. a noun with an abstract denotation), beer a mass noun, mug a count noun, man an animate count noun. In a deliberately intuitive sense, the formal complexity of the mental space respectively denoted by these kinds of nouns increases: a mass is formally or topologically more complex than an abstract space, as for the former we need a substance that extends in time and space and has (mass-) quantifiable parts; and to be countable is to involve more than a mass, namely some sort of boundary. Moreover, we see an asymmetric entailment among these layers, in that a thing, if it is animate, also shows restrictions for concreteness, mass, and abstractness, while the opposite is untrue. The question is what such entailments can follow from. They do not plausibly follow from any independently given semantic ontology, or the intrinsic structure of reality: reality in its post-Newtonian guise precisely does essentially not seem to ground our conceptual intuitions and their intrinsic structure. From a world in which most of matter is ‘massless mass’ (in John Wheeler’s phrase), or solid objects are mostly empty space, the conceptual constraints we see operative in the human conceptual system are not operative (another language than natural language, stemming from another corner of our minds, is used in physical science). Moreover, the hierarchy of semantic denotations actually correlates with the syntactic complexity of the expressions that have them as their meanings. Hence syntax (or whatever formal operations it reflects)

21 Thus observe:

i. (a) We gave the man/*institution our pictures.
   (b) I saw most men/*beer.
   (c) It’s a ‘man eat man’ world.
   (d) He’s more man than you’ll ever be.

In (ia) we see the expression man in one of its canonical uses: as the obligatorily animate beneficiary of an event in ‘Dative Shift’ guise; the equally plausible, albeit inanimate, institution doesn’t work. In (ib) we see man in a normal quantificational use, its animacy now being irrelevant (observe that non-count expressions like beer do not work). In marked contexts, man can also appear in a purely mass usage (in (ic) the expression is true if men never actually eat entire men, so long as they eat some man) and even in a purely abstract guise (in (id) man means manly, denoting the prototypical attributes of being a man). It is trivial to show that a canonically abstract expression, say beauty, is impossible in all but the most generic contexts of the form in (i) — and when it is coerced into #we gave beauty our pictures, then to the extent this works it is invoking some personified reading, where Beauty, for instance, is taken to denote a goddess.

22 Thus beauty (in its abstract sense) does not take articles such as the, in the way that beer does, is not measurable, and doesn’t take a plural (*we saw different beauties in the museum). In turn, beer is only classifiable or quantifiable by the rough estimates much or little, whereas, if it comes to mugs, we see languages applying more classificatory (e.g., number and gender markers, in languages having overt repertoires for this purpose) and quantificational resources (e.g., four mugs), with further such resources showing up in the case of animate nouns like man (e.g., ‘personal’ markers in many languages). In short, grammatical complexity tracks lexical-semantic complexity.
seems to be causally involved in the genesis of these denotations. This suggests
the conclusion that as the derivation unfolds (structural complexity builds up), se-
monic complexity builds up; but syntactic complexity engenders the asymmetric
entailments in question only if the derivational process has jumps in representa-
tional complexity in the sense above, hence is ontologically productive; hence
maybe this is precisely how the syntax is organized.

This surprising conclusion would point us to a system in which constraints of
a mathematical nature are operative that are no more interpretable in physical
terms than constraints operative in the number system itself. It may well form
a basis for a metaphysics of linguistics that is as different from contemporary
philosophical physicalism as one could get.

3 THE COMPUTATIONAL SYSTEM OF HUMAN LANGUAGE

3.1 Hierarchy, Recursion, and Merge

Having talked about issues of basic architecture above, let us now turn to the Min-
imalist conception of the basic operations of the computational system of human
language $CS_{HL}$, which makes the picture outlined possible. A basic assump-
tion of modern linguistics from its inception has been that the most elementary
property of language — apparently unique in the biological world — is that it
is a system of discrete infinity, consisting of objects organized in a hierarchical
rather than merely linear way. These are two different though related constraints.
Discrete infinity stands for the old Humboldtian observation that language makes
infinite use of finite means: just as there is no largest natural number, there is no
longest sentence. For, in both cases, if there was one, one could use the very oper-
ations that built this object to construct a larger one. As for discreteness, just as
each natural number is a discrete unit, each of the unbounded number of elements
of a particular language is a discrete unit, which does not blend into any other.
Discrete infinity relates to hierarchy in that the result of combining two units of
the system in question creates higher-order units which are not per se contained
in any of its constituents, and in particular mean something else than any of their
constituents. Constructions — structural configurations of lexical items — plainly
have meaning too, which differs from that of lexical items. E.g., a verb phrase such
as *kill Bill* is not contained in either *kill* or *Bill*: it is different both syntactically
and semantically from either of them. Syntactically, because the phrase contains
its constituents, and semi-antically, because a verb phrase depicts an event with an
intrinsic participant in it, where this event moreover is telic (bounded) in the sense
that it intrinsically ends with Bill’s death, this being inherently a sub-event of the
larger event in question. If we seek a principled explanation of language, then it
is discrete infinity, linearity, and categorial hierarchy that we have to explain.

Linearity is plausibly a phenomenon most directly relating to speech (or exter-
nalization): it is a paradigmatic interface constraint that arises as a feature of
language because we have to send the productions of the language system through
a linear phonetic channel, in which constituents are ordered according to the relations ‘before’ and ‘after’. What however can we say to explain the hierarchical order that we find in these objects on principled grounds? Throughout the first decades of the generative enterprise, the idea that expressions in their underlying syntactic structure have constituents (hence, a part-whole structure) was spelled out through phrase structure theory, which exhibits the familiar tree diagrams whose nodes bear categorial labels such as V’ (‘V-bar’), NP, PP, and so on. Intermediate (X’) and maximal (XP) labels are regarded as projections of the lexical items bearing the relevant label X. Phrase structure (PS), in short, is ‘lexical-entry driven’, or ‘projected from the lexicon’ (see [Borer, 2005] for discussion). As this was viewed in the earlier P&P tradition, PS provides a structural format within which the semantic information contained in a lexical item is to be coded if it is to become available to wider thought and cognition. This format, it was held, wasn’t arbitrary: that is, there is a systematic connection between the configurational syntactic positions in which lexical items appear in a phrase structure tree, on the one hand, and the individual lexical requirements that a lexical items makes on the syntactic objects it co-occurs with. We may call this the interface aspect of early PS-theory, the interface in question being that between the lexicon and the syntax, or lexical knowledge and structural knowledge. Put differently, the way that lexical knowledge becomes available to wider thought and cognition is mediated by the rules of PS, hence is linguistically specific — an assumption diametrically opposed to the assumption in generative semantics and much philosophy of language that thoughts are independent of what syntactic format codes them.

Let us distinguish this interface aspect of PS from the purely syntactic aspects of PS — the specific structural constraints on a well-formed PS-tree (see [Chametzky, 2003, 194-5] for this distinction). The latter aspect was first technically implemented through PS-rules, an idea that however quickly gave way to X-bar theory. The latter was based on the idea that the individual lexical requirements of a lexical item need not redundantly be joined by PS-rules that partially recapitulate the same information. The task arose to decide what to keep: the lexicon as an independent module, or PS-rules. The choice fell on the former, which meant that PS-rules disappeared. So individual lexical items came to be seen as essentially ‘driving’ the syntactic process, with X-bar theory providing as the only main restrictions that all phrases must be hierarchically structured and be ‘headed’, as noted above. With X-bar theory reduced in this fashion, it is no surprise that we find early papers in Minimalism aiming to entirely ‘derive’ whatever is left of X-bar theory, or even to ‘eliminate’ it (see e.g. [Chomsky, 1995, 378]). This meant that the properties of X-bar theory could be ‘derived on principled grounds’. Basically the idea is that the one operation that we minimally need to get a discrete infinite system will also account for phrase-structural hierarchy (or whatever needs to be preserved from it). It is widely argued that this operation, now called Merge, is as such n-ary and that binarity follows from other considerations, e.g. interface considerations. Binary Merge will minimally take two lexical items (LIs) α and

---

23 Most famous among these is Kayne’s Linear Correspondence Axiom (LCA) [Kayne, 1994;
\( \beta \) in order to put them into a (crucially unordered) set:

\[
(26) \text{Merge } (\alpha, \beta) = \{\alpha, \beta\}
\]

If \( \alpha \) is a verb and \( \beta \) is a noun, for example, then, by Minimalist logic, this minimal set should be all there is to a complex phrase. This entails: there are no projections, no bar-levels, and no XPs, all of which would be violations of Inclusiveness, which (recall 23) was the stipulation that syntactic objects arriving at the interfaces are no more than rearrangements of the features of LIs (nothing is added in the derivation). But there is still hierarchy on this picture, since a relation ‘contain’ automatically arises from the Merge process as so viewed, containment being set-membership. If Merge is recursive, not only hierarchy but discrete infinity follows. Recursivity means that after creating a set, Merge can apply to that set and another set or LI again, to create a larger set. This other syntactic object (a set or LI), \( \gamma \), can either be external to both \( \alpha \) and \( \beta \), or can be contained in (be internal to) either \( \alpha \) or \( \beta \). In either case, the resulting syntactic object is the new set

\[
(27) \text{Merge } (\gamma, \{\alpha, \beta\}) = \{\gamma, \{\alpha, \beta\}\}.
\]

Expectedly, we speak of ‘external Merge’, in the first case, and of ‘internal Merge’ in the second. If we derive (27) in the second way by internal Merge, we talk of the \( \gamma \) that is merged to \( \{\alpha, \beta\} \) and the \( \gamma \) that is initially contained in (say) \( \beta \) as two ‘copies’ of one another. This is the ‘copy theory of movement’, plausibly a minimal account of what movement is. In this way, internal Merge becomes identical to ‘displacement’ or ‘movement’:

\[
(28) \text{The copy-theory of movement:}
\]

Internal Merge of \( \beta \) to \( \alpha \), with \( \beta \) contained in \( \alpha \), yields two copies of \( \beta \), one contained in \( \alpha \), one external to it.

What we further see in (27) is that Merge leaves the syntactic objects to which it applies unchanged, the so-called ‘no-tampering’ condition:

\[
(29) \text{The no-tampering condition:}
\]

Do not tamper with syntactic objects already constructed.

This is another natural condition to impose on computationally efficient design: Merge cannot break up syntactic objects once constructed, or add anything to them [Chomsky, 2008a; 2008b]. Given (28), it now need not be stipulated any more, as was necessary in earlier PS-theories, that Merge (no matter whether external or internal) is invariably to the root of a syntactic ‘tree’. This now follows from principles of computational efficiency. Also, there is no more now to our PS-notion of a ‘complement’ than to our notion of what is ‘first merged’ to some
syntactic object $\alpha$, and there is no more to our PS-notion of a ‘specifier’ than there is to our notion of some $\gamma$ being ‘second-merged’ to $\alpha$ [Chomsky, 2008]. It is in this way that PS is ‘derived’.

None of this, on the other hand, explains why there should be movement or displacement in the first place, or why it takes place if it does (another problem, on which see section 3.2, is why, apart from Merge, another operation, namely Adjunction, should exist). All that internal Merge yields is the free availability of this option. So what Minimalist syntax has added to this basic and austere apparatus is a theory of what triggers applications of internal Merge/movement: namely, the existence of morphological features within lexical items that need to be ‘checked’, to use the terminology introduced above. Morphological feature checking is subject to a constraint of Agreement, both as regards Case-features and $\varphi$-features (person, number, and gender): such features, when checked against one another on two syntactic objects that have them, need to match. Yet, even an account of what triggers Movement is not yet an account of why it is triggered if it is, or why the triggering features exist. Here the answer has again largely been that ‘interface conditions’ can shoulder this explanatory load, even though there is disagreement on which interface is responsible for this. Chomsky [2008a; 2008b] argues that if the syntax makes operations such as external Merge and internal Merge freely available, it is natural to assume that these will also be recruited to express different kinds of semantic conditions that the SC imposes on the outputs of the grammar at the semantic interface. It has in particular long been argued that there are two fundamentally different such conditions: the ‘duality of semantics’, which consists in the duality of argument structure (thematic relations, ‘who does what to whom’ structure), on the one hand, and ‘discourse properties’ (quantification, scope, reference, force, focus, finiteness, etc.), on the other. Assuming this duality to exist on the ‘other’ (non-linguistic) side of the semantic interface, Chomsky argues that both internal and external Merge receive a ‘principled explanation’, in the following way: external Merge is motivated by the need to express argument structure, internal Merge by the need to express discourse properties. Note however that this reasoning must assume the duality in question to be independently given, i.e. not to depend on the evolution of a syntactic language, in the absence, as far as I can see, of evidence from comparative cognition to support this conclusion. Clearly, if these conditions depended on a syntactic language evolutionarily, they could not be invoked to motivate the structure of that language.

This is the point to return to our earlier note in section 1.5 above, that at least Chomsky’s mainstream version of the Minimalist Program retains a functionalist flavor. The line of reasoning on the rationale of movement that I just sketched clearly is a functionalist one: it explains movements from the need to express certain properties, or to ‘satisfy’ certain functions (see [Moro, 2004] for the same point). This kind of explanation is problematic if there is no independent evidence for the duality of semantics mentioned. In the absence of such independent evidence, the causal arrow could as well point in the opposite direction: that there is a
duality of semantic interpretation because there is a duality of structural resources that the language faculty makes available, engendering new kinds of thoughts not accessible otherwise. These two options could not be more different: in the former, Chomskyan option, syntax answers semantics (or expressive needs); in the latter, it is the other way around. Note that even if independent evidence was available for the duality in question, the suggested functionalist reasoning remains conceptually problematic: on a Darwinian (as opposed to Lamarckian) conception of evolution, organismic structures don’t evolve to satisfy certain needs. They evolve for independent reasons, and then the environment in which this happens finds some uses for them. Chomsky’s [2008a; 2008b] story as depicted above is consistent with this.24 Yet, it does assume that the environmental ‘problem’ that internal Merge ‘solves’ — the problem of expressing discourse properties — predates its solution. There is evidence suggesting that although our nearest relatives are capable of symbolic understanding, hierarchical organization, and even a limited form of a systematic combinatorics, neither their thoughts nor communicative acts exhibit intentional reference or propositionality [Terrace, 2005; McPhail, 1998; Fitch and Hauser, 2004]. That is, although a good case can be made that the ability to handle relations and to reason transitively — a hallmark of rational inference — is fundamentally shared with monkeys [McGonigle and Chalmers, 2006], reference and propositionality may have humanly specific features that depend on the evolution of a syntactic language.

Is an attempt to ‘motivate’ movement from interface conditions more promising if we turn to the phonetic interface? Moro [2000] is a sustained attempt to argue just such a case, an attempt born out of a deep suspicion against the functionalist flavor of the story just discussed, and in fact the whole Minimalist strategy of using morphology as a trigger for movement. The relevant trigger for Moro is the need for a linear compression of a hierarchical phrase marker when sent through the phonetic channel. If a phrase marker exhibits certain ‘points of symmetry’ — as e.g. when two maximal projections XP and YP are merged, as opposed to a head and a maximal projection — linearization difficulties arise if Kayne’s [1994] proposal for linearizing hierarchical phrase markers is assumed, leading the grammar to ‘save’ these pockets of symmetry by displacement of relevant constituents for the sake of obtaining an anti-symmetric arrangement that is linearizable again. Moro concludes that the reason that we have movements (and everything that results from them in the language faculty) is purely phonetic. This does not preclude that as a result of the re-arrangement of the phrase markers in question that need to be linearized, new semantic effects will also arise. Moro in fact claims that they do. Nonetheless, movement is not driven by ‘expressive’, i.e. semantic, needs or effects.

Hinzen [2006] opts for the alternative that the reason that movements exist is entirely internal to the syntactic system, and is induced by its intrinsic architecture: they have no externalist motivation or rationale at all. In particular, chains

---

24I find this much less clear for his earlier papers. See [Hinzen, 2006: section 5.5, esp. pp. 216-7] for a more extended discussion.
sets of contexts of copies of lexical items — are as such intrinsically interpreted, and are assigned kinds of meanings that would not exist in the absence of these syntactic objects or the processes giving rise to them. In other words, semantics intrinsically depends on syntax — the evolution of argument syntax and transformational syntax gave our minds entirely new thoughts to think. The rationale of movement in this sense is not phonetic. This is good news insofar as the more we make semantics independent of what happens in the syntax, the less we can use syntax to explain it. Semantics and syntax seem to be tightly related, and if the independent evidence for motivating the latter from the former is lacking, and comparative cognition studies do not as yet support it either, as noted, we might try it the other way around, to see semantics at last partially as a consequence of a particular kind of syntax. It may independently be a very natural assumption that semantic complexity builds up as syntactic complexity does. As different forms of syntactic combinatorial devices evolve — from adjuncts to argument structures to predications to chains — the specific kinds of meanings arise that correlate with these modes of combination (on the evolution of various combinatorial devices with different computational power and their semantic consequences, see also [Uriagereka, 2008]).

Returning to Moro’s different suggestion, does it solve the conceptual problem pointed out for Chomsky’s ‘functionalist’ view? Well, although it expressly identifies and distances itself from the functionalist flavor of Chomsky’s alternative, it preserves that flavor in a certain sense. Suppose that phrase markers that would not be linearizable would all crash at the phonetic interface. Then why is this a reason for movements to exist and to take place that ‘save’ these structures? If phrase structures cannot satisfy certain needs ‘imposed’ on them, one might point out, all the worse for them! Maybe they simply wouldn’t be phoneticized. Maybe language wouldn’t have evolved as a communication system to start with, remaining a system for expressing thought in the individual. Worse accidents have happened in evolution. While Moro’s alternative may thus explain why un-linearizable structures are not used to a phonetic purpose, it does not explain why a mechanism exists in the first place that transforms them to be so usable. It may explain why such a mechanism is selected for after it exists. Even then, these structures are being selected for because of a phonetic reason; and that reason seems quite unrelated to — and hence does not explain — the fact that because we are equipped with a transformational syntax, we are enabled to grasp entirely new kinds of thoughts that we couldn’t grasp otherwise.25

We began this section with some of the ‘big facts’ of language — that it exhibits discrete infinity as well as hierarchy, and we surveyed the current minimalist account of this hierarchy, which replaces or ‘derives’ earlier accounts based on PS-rules and X-bar theory. This surely looks like a stunning simplification of the grammar, but we should also note its downside. Might we have minimalized our phenomenon so much that we have lost it? Calling, as Chomsky does, the ‘elimination’ of PS a ‘derivation’ suggests that PS is actually preserved in Minimalism

25Naturally, one might deny that this is a fact.
rather than eliminated in the strict sense of the word. Certainly, a Merge-based system is structurally much simpler than a phrase structure grammar of the original kind. If Merge is construed as in (26), in particular, minimalist ‘phrase structure’ comes to look rather flat, looking essentially as in (30), which is the unofficial tree notation for the official syntactic object \{the, book\}:

(30)

\[
\begin{array}{c}
\text{the} \\
\end{array}
\begin{array}{c}
\text{book}
\end{array}
\]

No categories form here that are anything over and above lexical items (see [Collins, 2002]). In this sense, a categorial form of hierarchy does not get off the ground. Chomsky’s classical [1995] system has a similar consequence, even though it seems superficially dissimilar, and more complex. Thus, it regards the result of Merge not as merely a set, but as a labeled one, as in (31), where \(\gamma\) is the label:

(31) \(\text{Merge } (\alpha, \beta) = \{\gamma, \{\alpha, \beta\}\}\)

\(\gamma\) is not a syntactic object itself, in the sense of a constituent to which Merge applies; rather, it labels such productions (specifies their type). However, Chomsky [1995, ch. 4] also argues that the label is necessarily identical to one of \(\alpha\) or \(\beta\), hence a lexical item. But then, on this slightly more complex conception of phrase structure we again never meet anything other than lexical items as we go up a syntactic tree:

(31)

\[
\begin{array}{c}
\text{the} \\
\end{array}
\begin{array}{c}
\text{the}
\end{array}
\begin{array}{c}
\text{book}
\end{array}
\]

If labels are retained (see e.g. [Boeckx, 2006]), their rationale may be argued to be one of computational efficiency: on Chomsky’s view of labels, in particular, the label is the only thing that the computational system ever ‘sees’ when accessing an object such as (31), this label being what by hypothesis carries all the information about the constructed syntactic object that is relevant to further computation. Note that in this case the syntax never sees phrases (or projections) when accessing complex syntactic objects. It may seem rather strange in this regard that the only place where complex syntactic objects (phrases) are ever ‘seen’ are the interfaces: obviously, in particular, ‘the book’ is an object semantically interpreted very
differently than either ‘the’ or ‘book’. This is strange because what ‘sees’ these objects are, by definition, non- or extra-linguistic systems, in particular the system of ‘thought’ (the so-called ‘conceptual-intentional’ system). But why would these, of all systems, be able to see and interpret phrases? After all, phrases depend on linguistic structural resources which we do not assume in those pre-linguistic systems of thought.

These are not the only open issues facing the current minimalist take on phrase structure. Chametzky [2000; 2003] argues that Merge as defined in the above ways is not the most minimal option, actually, which is rather that the basic combinatorial operation should be simply a form of Concatenation (see also [Hornstein, 2005]). Set-formation is intuitively a more complex option than merely concatenating the relevant LIs, in which case we would define Merge simply as follows:

\[(33) \text{Merge}(\alpha, \beta) = \alpha \cup \beta.\]

Whether concatenation is simpler in formal terms, however, can be doubted, as it is intrinsically ordered, which an unordered set is not (see the next section on adjunction for more on this issue). Hornstein [2005] argues that the system starts out as a concatenative one and becomes hierarchical only when labels are added. This is a problematic result not only because it leaves out of account why syntax should be hierarchical, but also because we have seen that in current label-free accounts of Merge, hierarchy is still possible. But what kind of hierarchy are we in fact still talking about, in a projection-free system?

Chomsky [2008a, fn. 12] proposes an extremely weak account of what hierarchy amounts to, and claims:

‘Hierarchy is automatic for recursive operations, conventionally suppressed for those that merely enumerate a sequence of objects’.

In particular, enumerating the sequence of natural numbers yields hierarchy. As Chomsky reconstructs this process, there is an initial ‘lexical item’, which we may for convenience think of as the empty set, \(\emptyset\). Then Merge as the recursive operation ‘set-of’ applies to it, and forms the singleton set of \(\emptyset\), which is the set \(\{\emptyset\}\). Being recursive, the operation iterates indefinitely, and we can think of this sequence as the sequence of natural numbers, with which it is isomorphic:

\[(34) \emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \text{etc.}\]

The basic idea here is that \(c_n\) is defined by induction on \(n\) as follows: \(c_1 = \emptyset\) and \(c_{n+1} = \{c_n\}\) for each \(n\).\(^{26}\) But there doesn’t seem to be a particular reason to view Merge in this sense as being essentially a one-place operation. So let us contemplate it as an \(n\)-place operation, of which \(n=1\) and \(n=2\) are restrictions. Merge with \(n=1\) yields arithmetic, which we may regard as a ‘minimal language’

\(^{26}\)The more standard kind of hierarchy that underlies the definition of the successor-function of arithmetic and subserves the notion of an ordinal number is the following:

\[(i) \emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}, \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \text{etc.}\]

In this series, each member is constructed by collecting all and only its predecessors into a set.
with a lexicon of cardinality one. Merge with \( n=2 \) yields language, when a larger lexicon is added.

But what does (34) tell us about natural language? Merge as recursive-set-formation in the above sense depicts an essentially linear and mono-dimensional system. It is precisely not ‘ontologically productive’ in the sense introduced above. It may be for that reason that the system tells us nothing about the specific categorial hierarchy that emerges as a matter of course whenever we combine a head with an argument in human language: if we build up a clause compositionally, from the bottom up, we first construct a Verb Phrase that canonically embeds a Noun Phrase, then an Inflectional Phrase which embeds the Verb Phrase, and finally a Complementizer Phrase which embeds the Inflectional Phrase. If Merge is construed on the lines above, it won’t tell us anything about this presumably universal structural skeleton of the clause, whose origins must now derive from an independent mechanism. If that mechanism is labeling in the sense above, and labels are lexical items, it won’t help. It may thus well be that categorial hierarchy is, contrary to what mainstream Minimalism assumes, either build into the combinatorial operation Merge, or follows from operations of an essentially non-linear form which interact with Merge. One of these options should be pursued if the categorial hierarchies in question and their intrinsic interrelations cannot be blamed on anything extra-linguistic, like the supposed ‘conceptual-intentional systems’. Yet, both represent significant departures from how the computational apparatus of language is standardly viewed in mainstream Minimalism. Moreover, it compromises Minimalism’s basic explanatory strategy, to motivate syntax from interface conditions externally imposed on it. As I have argued here (and see again [Hinzen and Uriagereka, 2006]), whatever operation bootstraps categorial hierarchies seems to be internal to the syntax, as opposed to pre-syntactically given. Minimizing the computational system to trivial set-formation may endow the non-linguistic systems of thought with a richness they lack.

3.2 Adjunction

The standard definition of Merge assumes that the sets that are constructed by Merge are unordered: there is no asymmetry between the ‘Merge-partners’ A and B. Asymmetry follows from a mechanism independent of Merge, labeling. Merge as such is purely symmetrical: it is not the case that in the set \{A, B\}, A is merged to B, say, as opposed to B being merged to A. This is different for the case that B is adjoined to A, an operation that, on Chomsky’s canonical 2001a/2004 proposal, is crucially asymmetrical and yields the structure of an ordered pair:

\[
\text{(36) Pair-Merge(A, B)=<A,B>}
\]

Using the standard Kuratowski definition of ordered pair, ordered pairs are sets of (unordered) sets:

\[
\text{(37) <a,b> = \{a\}, \{a,b\}\}}^{27}
\]

\[^{27}\text{Chomsky’s [2004] definition of ordered pairs omits brackets around ‘a’, a non-trivial move}\]
This is somewhat surprising since adjunction is in many ways a simpler operation than argument-taking. This is the case semantically, in that an adjunction structure like run quickly is interpreted conjunctively on the lines of ‘there was a running and it was quick’ (with the two propositions flanking ‘and’ non-ordered), whereas an argument-structure like John runs is of course not interpreted as ‘there was a running and it was John’. In short, semantically an adjunction structure corresponds to the operation of predicate composition, whereas standard Merge accounts for argument structure and discourse (or ‘edge’) properties. But the same is also the case syntactically, for adjuncts don’t take part in the paradigmatic kinds of things that arguments partake in: they don’t move for Case-reasons, are not part of the Agreement system, adjunction of B to A does not change the categorial status of A, adjuncts do not receive theta-roles, and adjunction can be re-iterated indefinitely, while the argument system is highly restricted. If anything, then, one would guess that prior to the evolution of a full-blown modern language, whatever proto-language existed would have consisted of the adjunct system only, viewed as a more archaic sub-part of modern language. This is in diametrical opposition to the account of adjuncts in Chomsky [2001a/2004], where adjuncts come out as a complexification of a Merge-based symmetric system: merely looking at (36)–(37) makes clear that we are dealing with a formally more complex system.

The asymmetry of adjunction is argued for by Chomsky [2001a/2004, 117] on the grounds that in a structure like

(38) old man

the whole expression functions as if ‘old’ weren’t there, for further computation, apart from semantic interpretation. Thus, the adjunct ‘old’ has no theta-role in (38), although the structure does, namely the same as ‘man’; ‘old’ is unselected for by ‘man’, and the selectional properties of ‘man’ are retained after adjunction. So, the adjunct acts fundamentally different from the head. But then, we should ask whether these asymmetries can be grasped by a system that doesn’t have (or cannot deal with) these notions (like ‘head’). So, it might be only from the point of view of the complexified system that includes argument structure that the adjunctive system which is a sub-part of it creates ‘asymmetries’. For a system that doesn’t know about ‘heads’ and ‘selection’ in there isn’t an asymmetry and ‘old’ and ‘man’ are entirely on a par.

What is more striking is that Merge as technically defined in Chomsky’s recent work actually fits the adjunct system in crucial respects quite well, since, as we saw, Merge so conceived crucially does not project any new categories either. Adjunction yields the hierarchies exhibited by linear systems like the natural numbers, but not the categorial ones. Also the infinite iterability of Merge fits the essential unrestrictedness of basic adjunctions. Since arguments are correlated with restrictions on the application of Merge, one wonders again whether the argument structure system does not build on the adjunctive one as opposed to vice versa. In the limit, this line of thought suggests, somewhat provocatively, that Merge does whose rationale is unclear to me.
not yield core syntax, but rather yields adjuncts that fall short of much of syntax, in fact have little syntax to them. One wouldn’t, then, begin one’s reconstruction of language with unordered sets, and define adjuncts in terms of them. Rather one would begin with adjuncts, and view the argument system as something that builds on adjunction while adding a crucially new element to it: thematic roles, which are basic to the argument structure system.

This conclusion ties in with the conclusion reached at the end of the last subsection. There we suggested that syntax could only be so poor as standard Minimalism suggests if we assume that the systems on the non-linguistic side of the semantic interface are implausibly rich. Now we can submit the further thought that it might well be that adjunct syntax, but only it, can be ‘motivated by conditions imposed by the semantic interface’: adjunction structures have a very simple semantics, and the very operation of adjunction might well be explained on essentially semantic grounds (see [Ernst, 2002]). But the more we then see that the argument system (and the transformational system) is crucially different, introducing novel elements of a structural nature and kinds of restrictions that nothing in the adjunction system predicts, the whole project of ‘motivating syntax from interface conditions’ will look misconceived.28 We shouldn’t begin with Merge, and then ask why adjuncts also exist. The existence of adjuncts is the easy part. That of arguments is much harder.

What all this certainly allows us to conclude that there is still a problem with adjuncts in the theory of syntax. Minimalism has the virtue of bringing this very problem into a much sharper theoretical focus: for in Minimalism everything in syntax is highly contrained, following ideally from ‘virtual conceptual necessities’ and ‘last resort’ conditions. Adjunction — an operation that by no conceptual necessity needs to exist, and which is outside the core of syntax that is guided by the necessities and last resort operations in question — simply seems to find no natural place in a system viewed in this fashion.

4 CONCLUSIONS

I have tried to give an indication of an extremely ambitious scientific program in the understanding of human nature that, if successful, opens up entirely new perspectives upon ourselves and the nature of our minds. It pushes explanatory scope to its limits. The truth be told, there is still any possibility at this moment that the project will prove misguided, as it makes wrong presumptions on human nature. The strange thing however is the extent to which the project has been

28Interestingly, Chametzky [2003, 206-7], in his discussion of Chomsky’s construal of adjuncts, remarks that this construal gives rise to numerous technical problems, apart from being conceptually unsound as well. He concludes that since Minimalism fails to account for adjuncts, this is an instance of the larger fact that it fails to account for phrase structure altogether. My line of argument has been different here: that Minimalism with the basic operation Merge perhaps accounts unproblematically for adjunct structures, and the way it succeeds in this as such suggests a reason why it may fail to account for syntactic hierarchy.
found intriguing on such a global scale and led to fruitful works opening up totally new avenues in epistemology and language evolution as well. It also promises — though it has not been received in this way outside of generative grammar — a unification between current disparate theories of grammar, since minimalization as such is a task that may help to uncover a core of conceptual necessities that point to a theoretical structure any theory of grammar must have. Different grammar theories may compete for minimality, and thus have a common theoretical agenda. Moreover, Minimalism forces us to sharpen our focus on large theoretical issues, such as the syntax-semantic relationship, which I have emphasized throughout this chapter (and see [Hinzen, 2006; 2007]). Clearly, the basic picture now in generative grammar is that syntax and semantics are tightly interwoven to the extent even of being unified with one another, or at least of standing in a transparent relation. This is not the stance of, for example, Jackendoff [2002], where the lack of transparency in the syntax-semantics mapping is a prime motivating factor for adopting a ‘parallel’ architecture. This is to resign from more optimistic visions of architecture and to content oneself with a hypothesis that is clearly not the null hypothesis. Before non-transparency is explored or endorsed, one wants to see a more restrictive theory assuming transparency fail. This is what Jackendoff argues by appealing to various facts about syntactic categories and argument structure, among other things. But, as always, and as Minimalism particularly emphasizes, facts are in part of our own making, and fall out differently as the theoretical perspective changes.

ACKNOWLEDGEMENTS

I am very grateful for numerous discussions in which the various views of this paper emerged: especially with Noam Chomsky, Angel Gallego, Michiel van Lambalgen, Andrea Moro, Massimo Piantelli-Palmarini, Martin Stokhof, and Juan Uriagereka.

BIBLIOGRAPHY


COMPUTATIONAL LINGUISTICS

Gerald Penn

1 DEFINING COMPUTATIONAL LINGUISTICS

Until late 1959, we accepted the label “MT”, but two months ago we petitioned for a change. Our new titles are linguistic research and automatic language-data processing. These phrases cover MT, but they allow scope for other applications and for basic research.

Machine translation is no doubt the easiest form of automatic language-data processing, but it is probably one of the least important. We are taking the first steps toward a revolutionary change in methods of handling every kind of natural-language material. The several branches of applied linguistics have so much in common that their mutual self-isolation would be disastrous. The name of our journal, the name of our society if one is established, the scope of our invitation lists when we meet, and all other definitions of our field should be broadened — never narrowed. In 10 years we will find that MT is too routine to be interesting to ourselves or to others. Applied linguistic research is endless.

David G. Hays [1961]

Clearly, Dr. Hays was a gentleman who appreciated the value of a fitting name. His preface to the English translation of Axmanova et al. [1961] [Axmanova et al., 1963] contains one of the first written uses of the term, computational linguistics, a term to which he had resorted a year earlier in naming the Association for Machine Translation and Computational Linguistics (later renamed the Association for Computational Linguistics, or ACL), of which he was the first Vice President, and indeed a term that he is widely credited with having invented [Kay, 2000].

Computational linguistics, we are to believe, is not merely the study of mechanical translation or machine translation (MT), the automated translation of one language to another, nor is it mathematical linguistics, a much older term that was in widespread use in North America before 1962, and which remained ideologically more acceptable during the Cold War among academics that answered to Cominform members. The difference between computational linguistics and mathematical linguistics is less clear, but it is likely that the former was regarded as a more grant-worthy appellation for certain topics within the purview
of the latter. This was to become a particularly acute concern in 1965 when the
circulation of an influential, negative assessment of machine translation by a Na-
tional Academy of Science panel, published the following year [ALPAC, 1966],
threatened an imminent collapse of support for MT research in the United States.

This same NAS report went out of its way to distinguish computational linguistics
as the field that consists of all of the other computational research, including
“basic,” less application-driven research, that was germane to the new science of
linguistics than the tired, old topics of machine translation and information re-
trieval, on which the sun was then setting. Perhaps this was a ruthlessly practical
way to circumscribe the boundaries of an academic discipline, but for its time it
was relatively enlightened. Up to this time, mathematical linguistics, like com-
puter science itself, had been defined at least as much by the post-war political
economy of funding agencies such as the United States Air Force and Office of
Naval Research as by some paradigm shift in our view of the world. With this
new characterization, computational linguistics was being encouraged to venture
out and acquire a body of basic scientific knowledge that would support a later,
more mature investigation of applications such as machine translation.

Out of concern for the probable fate of computational linguistics funding at the
hands of an indictment of machine translation, the Academy’s Committee on Sci-
ence and Public Policy asked [Brooks, 1966] the authors of this report to include a
statement on the need for computational linguistics funding. This statement iden-
tified two important tracks of research that inter alia, computational linguistics
should engage in (p. 31):

1. “basic developmental research in computer methods for handling language,
as tools to help the linguistic scientist discover and state his generalizations,
and as tools to help check proposed generalizations against data;” and

2. “developmental research in methods to allow linguistic scientists to use com-
puters to state in detail the complex kinds of theories (for example, grammars
and theories of meaning) they produce, so that the theories can be checked
in detail.”

In very broad strokes, these describe an undercurrent that remained very influ-
ential in computational linguistics research until the mid-1990s, and that placed
the field of computational linguistics necessarily in very close proximity to the
contemporary research programmes of theoretical linguistics.

2 NARRATIVES OF PROGRESS

Then something happened. The post-1995 ACL community have woven a num-
ber of somewhat fictional narratives about what that something was, but the re-
sult has very clearly been a redefinition of computational linguistics that is much
more answerable to its potential applications (including, very prominently, ma-
chine translation systems) and much less beholden to other branches of linguistics for its direction.

2.1 The Advent of Statistical Methods

Perhaps the growing rift with linguistics can be felt most prominently today through the preference for statistical techniques in computational linguistics, a trend that generative linguists have been very slow to embrace. In CL, “statistical” is used to refer to both statistical sampling methods as well as probabilistic models. It will be used here to refer to the larger-dimensional statistical methods that are pervasive in engineering and statistical pattern recognition, as opposed to the statistics that are widely used in the life sciences and elsewhere for descriptive hypothesis testing. Indeed, quantitative approaches to experimental design and significance testing are not taught in computational linguistics curricula, have only begun to appear in CL publications within the last ten years, and remain poorly understood by most CL researchers.

It would be inaccurate to suggest, on the other hand, that statistical methods in the pattern recognition sense were only recently introduced to this area. In fact, statistical methods were not unknown to CL even before the break with machine translation research. Frumkina [1963] declared: “We can show that a large number of linguistic situations exist that can be described both fully and briefly only by means of statistical rules,” referring mostly to concepts from corpus linguistics such as Zipf’s law, relative frequencies and sample sizes necessary to achieve certain error bounds. Paducheva [1963] contributed a paper to the same collection on machine translation, with a précis on statistical methods that included Shannon’s noisy channel model, conditional probabilities and entropy, as well as Markov processes. Paducheva’s précis was incomplete for understanding a modern statistical MT system; notably, Bayes’s rule is not mentioned (see section 2.3 below). She also reached the conclusion that there were limits to the information-theoretic analysis of language because of (citing Chomsky and Miller [1958]) determinate rules governing grammaticality, those governing meaning, and a perceived need to distinguish the two within the formal structure of an information-theoretic model. The list of concepts from probability and information theory that she brought to bear on the problem is nevertheless eerily prescient to any 21st century computational linguist. Jurafsky [1992] observes that lexical entries for verbs were being annotated with probabilities of what we would now call their subcategorization frames as early as by Ulvestad (1962; cited by Jurafsky [1992] as Ulvestad [1960]).

Many researchers today in statistical machine translation cite a memorandum by Weaver [1949] as among the earliest suggestions that their current approach to MT was worth exploring, based on an analogy of translation to decipherment. This, too, must be placed into an appropriate historical context, as this proposal was only the third of four that Weaver [1949] made [Hutchins, 2000]. The others were: (1) to disambiguate word tokens by examining the adjacent words in their context, (2) to approach translation in a purely logical, almost proof-theoretic
manner, because “written language is an expression of logical character,” and (4) to approach translation by way of the common logical structures inherent to the grammars of all languages. The earliest investigation of the first seems to have been a set of human-subject experiments by Kaplan [1955] that was widely regarded by his contemporaries as a demonstration of the great potential inherent to using adjacent words of context [Mel’chuk, 1963], even a very limited amount of context, to disambiguate word tokens. Even now, Kaplan [1955] is still regarded as one of the earliest studies in word-sense disambiguation, as this topic is now called (again, see below). At this particular period in time, however, it was widely assumed that most of the ambiguity that would be faced in natural language understanding or machine translation was lexical and moreover limited to certain substantive categories of words. This first proposal would have been considered not as a research programme in a separate subject called “word-sense disambiguation,” therefore, but as one in machine translation itself. The second and fourth are more akin to the paradigms of machine translation that were dominant before the advent of the IBM statistical machine translation models in the late 1980s. The fourth is suggestive of what eventually became known as an interlingua approach to machine translation, the appropriateness and structure of which quickly became a central topic of discussion in the earliest machine translation conferences [Reifler, 1954; Yngve, 1961]. So, while Weaver [1949] was indeed among the first to suggest the possibility of translation by analogy to wartime cryptographic methods, in another sense he deserves no credit at all, as he did not, almost certainly could not, and would probably have been unwilling to provide any guidance as to which of his four proposals would prove most advantageous. Weaver [1949], unlike Hays twelve years later, looked out upon machine translation at the very outset and saw an area replete with fascinating possibilities, all of which in his mind doubtlessly deserved further investigation.

Perhaps the most lasting legacy of the early, postwar statistical approaches was the enabling metaphor of language as a code (see especially Miller [1951] on defining this term). The basic message units in this code were sentences, not documents, and these units were not viewed as inherently endowed with

---

1 This use of limited contexts is vaguely similar to what is now called a language model (see below), a term that appears not to have been used in print until Jelinek [1976], but an idea that had only just started to be applied to English text by Shannon [1951]. Shannon [1951] referred to these contextual units as n-grams, a term that is now often used synonymously with language model, even though Shannon’s [1951] n-grams were character-level, i.e., adjacent letters of context that can be used, for example, to restore obscured letters on an optically scanned page of printed text.

2 Carnap’s work was known and cited, particularly his work with Bar-Hillel on “semantic information theory” [Bar-Hillel and Carnap, 1953–1954], but there was still no general appreciation in computational linguistics at this stage of the difference between a sentence and an utterance.

3 There was some interest in speech-to-speech machine translation even then, and while the continuous nature of the speech signal as opposed to the discrete nature of written text was very much appreciated, the difference between the registers of continuous, interactive, colloquial speech on the one hand and read speech, i.e., written text read aloud, on the other had not received much attention yet in this community. Even today, however, “document” is often used to refer to a computer file containing audio data, as well as to one containing textual data.
meaning, but rather as conveyances of it. Some of them could potentially convey multiple meanings; these sentences were ambiguous, which in the then-current use of this term, appears to have included the potential of being underspecific if the possible extensions could be differentiated in the target language in which the translation output was written. It was the goal of the translator to determine which of these meanings had been intended by the sender so that the correct meaning would appear among the possible meanings intended by the new code in the target language. “Meaning” itself was not regarded as something that required disambiguation; it was the text that had to be disambiguated.

It was understood that an observed message may at times be ill-formed, and there was a very early awareness that some facility for recovering from these errors was necessary. It was also understood that errors in meaning were possible, e.g., a disagreement in time between a temporal adverb and the tense of the verb it modified, but these errors were not regarded as recoverable, in the sense that the intended meaning was deemed to have been lost in this case and so no translation was possible. Neither class of error had anything to do with truth value, which was regarded to be outside the scope of the translation enterprise altogether — even outside the scope of disambiguation.

These shared beliefs about language and meaning, which were tacitly presupposed in the publications of this period, were common to both statistical and non-statistical work. Even after statistical and probabilistic methods had fallen into disrepute, the presuppositions remained for many years. What is more interesting is that, upon the return of statistical methods in the 1990s, the view of text as a container from which an objective meaning must be extracted — this time, for translation and many other purposes — returned with a vengeance as well. The significance of statistical techniques in connection with these presuppositions may be their utility as a means of disambiguation, which is necessary for this extraction to be successful, together with the attraction of having a simpler, self-contained classification problem, to which these methods can fruitfully be applied.

Language was no ordinary code, moreover. It was one that was uniquely suited to conveying meaning. Although later approaches that called themselves interlingua-based would use very abstract representations of meaning and syntactic structure, early interlingua representations in MT, in keeping with the use of that term in the planned language movements of the period, used human language itself, oscillating between a pivot language, an initial target language that served as the source for subsequent translation into several other languages, and a model or regularized target language, the syntax of which was controlled so as to make translation from one or more particular source languages easier [Reifler, 1954].

2.2 The Futility of Knowledge-Rich Approaches

Nevertheless, abstract representations, of meaning, syntax and other linguistic knowledge, did soon become the norm. When statistical methods were supplanted, it was not by numberless clones of the same models and algorithms, but by those
based largely on symbolic inference and computation: new and promising areas in their own right.

Indeed, once statistical methods began to regain prominence in the 1990s, non-statistical approaches were often characterized as knowledge-rich, as opposed to their knowledge-lean statistical cousins. The nature and value of knowledge that these terms assume can be very difficult for someone outside the field of computational linguistics to intuit. “Knowledge-rich” is a bad thing here, because incorporating a variety of knowledge sources into a natural language processing system in advance is labour-intensive to create, prone to error, and a nuisance to maintain. It is much better to be knowledge-lean, at least at the outset, and then learn what is required for a task automatically. Conditional random fields are a modified version of hidden Markov models that exhibit superior performance on many labelling tasks in computational linguistics such as labelling (or tagging) each word with its part of speech, but a CRF-based part-of-speech tagger is not knowledge-richer than an HMM-based part-of-speech tagger. If anything, it is knowledge-leaner, because it can acquire more knowledge from the same amount of annotated data; so knowledge of statistical pattern recognition does not count. Statistical methods that rely on manually annotated corpora of texts are knowledge-richer than statistical methods that use naturally occurring samples of text, but still knowledge-leaner (which is a good thing) than a non-statistical approach that manually incorporates the exact same or less knowledge than what appears in the annotation. Upon careful consideration, it appears that what is being commoditized here is not knowledge at all, but time, especially my time, the CL researcher’s, at the expense of the annotator’s and the statistician’s, and regardless of the consumer’s (who may actually waste more time, if the resulting appliance does not work as well). “Knowledge” is a brittle and time-wasting encumbrance, which may or may not be true and is in fact characterized by a lack of epistemological justification, because the system did not infer it itself.

To a great extent, early CL researchers never really had a choice, because the same opportunities for advancement with corpora did not present themselves. The first corpus-based study of American English was not published until 1967 Kučera et al. [1967], and the ASCII encoding standard for electronic text did not even appear until 1963. Computer memory was also prohibitively expensive. Corpora are essential for computational linguistics, especially for statistical methods, even if formal grammars that express the same abstract annotations are available. Corpora ground the instances of those abstract concepts in a naturally occurring context, they force the annotator to consider real data, and, once annotated, the instances of those abstract concepts can be counted, from which initial probabilities can then be estimated. Surely, it can be no accident that the enthusiastic re-uptake of statistical methods should have coincided so closely in time with the appearance of the World Wide Web, and the birth of the very influential Linguistic Data Consortium, which serves as a distribution channel for corpus data that were only within reach to the likes of IBM and Bell Labs until the late 1980s.
In early CL, knowledge-richness was also regarded in a very different light, probably because the novelty of being able to formally express propositions and rules of inference at all had still not worn off. Knowledge of language was no different. Then, as now, non-statistical approaches were thought of as deductive, but the rules of deduction about linguistic structure were not the usual rules of logic that had been the subject of philosophical investigation. These new rules of linguistic deduction were instead the rules of grammar, which could either derive the grammaticality of a sentence, or transform one grammatical sentence into another. The empirical basis for grammaticality judgements received no explicit discussion, and there is no evidence from the early machine translation or computational linguistics literature, moreover, that grammaticality was even regarded as an empirical issue. It appears to have rather been the final, missing, but necessary link to complete the chain of analogical reasoning that justified the appropriation of the new technology of symbolic inference to the task of working with the complex structure of language. In this analogy, grammaticality was what served in the role of truth.

The complete and methodical explication of these new rules of grammar for the purpose of supporting MT was itself a novel pursuit to the engineers, physicists and mathematicians that were engaged in these projects. It is likely that the abstract representations of meaning offered by deductive approaches — both first-order logic and a variety of relational representations derived from associative models of memory in the early, very influential work of McCulloch and Pitts [1943], and later of Newell and Simon [1956] and other pioneers of artificial intelligence research — were an attractive force on early computational linguists of equal or greater magnitude to the abnegation of statistical methods in the early work of Chomsky, beginning with Chomsky [1956].

2.3 The Linguists Made Us Do It (or Stop Doing It)

Among computational linguists, there is a long and venerated tradition of casting blame upon non-computational linguists. This extends at least as far back as ALPAC [1966], the concluding paragraphs of which jibe that the only fruitful result to emerge from the otherwise disappointing attempts that applied circa-1950 linguistic theory to computational models was “shaking at least some inquisitive linguists out of their contentment.” Within ACL circles today, there is a special measure of contempt reserved for theoretical linguists, particularly Chomsky, for having delayed the more thorough investigation of statistical methods that are now regarded to have been both inevitable and a matter of commonsense.

Chomsky’s earliest objection to the use of statistical modelling for natural language was essentially that statistical models could not distinguish between a low/high assigned probability on account of grammaticality or other formal aspects of a candidate sentence’s syntax and a low/high assigned probability on account of some contingent fact about the world as described by the candidate. At face value, this claim is now known to be false [Pereira, 2000], although the means by which the two can be distinguished does seem to require a more abstruse model
for this specific purpose than computational linguists in the late 1950s and early 1960s would have had any other reason to resort to. It is indeed heartbreaking to watch research programmes such as Paducheva’s [1963] stop dead in their tracks because of an assertion like the one she cites from Chomsky and Miller [1958], but it must be acknowledged that early citations such as this are exceedingly rare. In the main, the influences that are ascribed to Chomsky through citations from early (pre-1966) computational linguistics research are affirmative ones, although often speculative, and situated within descriptions of larger research programmes that are entirely non-statistical. Victor Yngve remarks [Yngve, 1961] that one of the more promising current threads of research taking place at his lab at MIT is a study by Edward Klima on translating imperatives, -ing forms, relative clauses and pronouns, following “the theoretical work of Noam Chomsky.” None of the other research he describes used statistical methods, nor apparently followed Chomsky’s theoretical work. Amidst a very broad portfolio of research projects on natural language at the RAND corporation, Hays [1961] acknowledges the influence of Chomsky’s work on “grammatic transformations;” but the only statistical influence he acknowledges is the view of distributional semantics advocated by Chomsky’s advisor, Zelig Harris.

A common alternative explanation is that linguists were generally prone to discounting the value of statistical methods because they have historically phrased their theories with a bias towards the generation of strings of words from underlying logical forms, rather than towards an analysis of input strings into logic. This bias was perceived even among some early deductive MT enthusiasts; Oettinger and Sherry [1961] boast that Chomsky’s theory is concerned only with “sentence synthesis” (what is now called surface realization, the final step in generating natural language text) whereas they have a theory of sentence analysis (parsing, the dual of surface realization; see Syntactic Structure in CL below) that is nevertheless consistent with contemporary views of syntactic phrase structure. Unlike the Chomskyan explanation, this one interprets the numbers that a statistical model produces not as scores of the degree of grammaticality of a sentence, nor as scores of how typical a sentence is in its contingent use of words, but as scores of how likely it is that a particular sentence should be analyzed in a particular way, where the possible analyses are distinct alternatives formulated in some abstract representation language. As a result of parsing, the highest ranking one of these is then selected. If a single correct representative were known in advance, as in generation, then there would be no need to rank a set of alternatives.

Relative to the historical evidence from early computational linguistics, this alternative view seems almost beside the point, because early CL was so focussed on machine translation. MT systems either did employ an abstract meaning representation, in which case they also incorporated a sentence analysis component, over which the linguists would presumably have held less sway, or they did not employ one, in which case a putative input representation’s determinacy would have been irrelevant. In either case, it is also relatively unusual that the input to a surface realization algorithm, then or now, would be completely specified in every
respects that could have an impact on the order and choice of the words generated. The choice of which preposition to use in a translation, for example, requires a great deal more work involving lexical collocations and syntactic analysis than one would expect of the higher-level planning component of a natural language generator. If specific prepositions appear in a semantic representation, however, this is exactly what must happen, in addition to the representation becoming less portable across language pairs. Statistical approaches to surface realization and other problems in natural language generation came relatively later than in parsing, but their value is now also clear. In this later work, we see probabilities being used both to guide further specification of an input semantic representation, as well as to rank candidate surface realizations by their degree of acceptability or grammaticality.

This alternative explanation also promotes a false dichotomy according to which grammars either parse or generate. In fact, much of generative linguistics, going back as far as the venerated Indian grammarian, Pāṇini (c. 450 BCE) [Joshi and Kiparsky, 2006], designed grammar not by blindly generating strings from a given abstract syntactic or semantic representation, but rather in a mode that might be called verification, in which both an abstract representation and an (in)correct string are known in advance, and the grammar must correctly (not) license a derivation of the latter from the former. In this mode, in which both ends are fixed, there is arguably less need for disambiguation or selection by a statistical component. What happens when one end is freed and the grammar is then actually used for true parsing or generation? In this situation, theoretical linguists (notably, again, Pāṇini) have been known to resort to one or more default mechanisms or some measure of economy in a clear attempt to restrict the potential over-generation of strings. In Chomskyan linguistics, this happened very late (early 1990s) — so late that the computational linguists who studied parsing had already broken away, either to look for psycholinguistically plausible restrictions on the parsing algorithms themselves, or to investigate the statistical disambiguation of parses produced by context-free grammars (CFGs), a syntactic formalism that again dates from the mid-1950s, and one that is prone to massive ambiguity in its syntactic analyses. The present use of statistical methods in parsing by computational linguists is then, in at least one respect, quintessentially early-Chomskyan in that it has obediently received a view of grammar that has been very carefully circumscribed to exclude any built-in symbolic apparatus for disambiguation, as well as most of the world knowledge or reasoning that might be useful for doing so. The mere attempt to disambiguate under these circumstances is one that can be credited to twentieth century linguistics.

---

4 In computer science circles, Pāṇini is sometimes credited with the invention of CFGs, because of a cosmetic similarity between certain conventions for specifying grammar used by both Pāṇini and an early means for defining CFGs called Backus-Naur Form, but Pāṇini’s grammar is nothing like a CFG.
2.4 Computational Linguistics as Artificial Intelligence

By the early 1970s, the refrain that had become familiar was that statistics have no place in computational linguistics because statistics are for disambiguation, disambiguation requires world knowledge, and computational linguistics is not about world knowledge [Kay, 2011]. This, too, rings hollow because knowledge representation theory was very much about world knowledge, and it, too, had eschewed statistical methods. Both CL and knowledge representation are sub-disciplines of artificial intelligence, and, as such, both were profoundly influenced by the same symbolic systems research that took place during the late 1950s and 1960s. Indeed, the term “symbolic” has often been used over the last twenty years in CL to describe non-statistical approaches, even though discrete probability distributions are widely used within statistical approaches. This use of “symbolic” almost certainly began as a reference to this genre, not to the use of discrete symbols itself.

The AI genre was historically preoccupied with deductive systems that classically were not numerically parametrized, and its influence reached much further than computational linguistics. It included, for example, machine learning, a term that, within the computational linguistics community, has now become synonymous with what engineers would more precisely call statistical pattern recognition, but which historically referred to a variety of learning and inference algorithms, many of them non-statistical, that were inspired as much by research in psychology and cognitive science as by probability and information theory. This was fitting, because the goals of early AI research centred around the development of thinking machines, and our ability to attribute thought to a computer was clearly determined by more than its performance on some collection of mundane classification tasks. The computer had to perform in a manner that corresponded to human cognition. While there was a great deal of debate as to how close and at what level that correspondence had to be, it is fairly clear from of the position papers by McCarthy, Minsky, Newell and Simon that touched on the subject that using statistical methods was, if not cheating, then of no consequence. This was certainly due at least in part to the Zeitgeist of symbolic inference during this period. It may have also been a reaction to the Cybernetics research programme of Norbert Wiener, which needed numerically rich models in order to simulate analog processes and was regarded by early AI as too low-level to be at all illuminating of higher aspects of human cognition. Nilsson [2009], in explaining the choice of the term “artificial intelligence,” cites a quote by McCarthy which implies that there was some personal rivalry between Wiener and him that kept the two fields at a distance.

It was ultimately the social connection to artificial intelligence, particularly to knowledge representation, that brought CL out of its ideological shell with respect to presuppositions about language and meaning. The eventual aspiration not just to decode secret messages into static meaning representations as goals in their own right, but to understand language as a human agent would, led to an ac-
knowledgement that language must be subjectively interpreted, and to a series of very influential works, beginning perhaps with Winograd [1972] and Schank and Colby [1973] and culminating no earlier than with Sowa [1984], that attempted to reconcile advances in knowledge representation with natural language processing (during this period, often calling natural language understanding). This quest typified CL research during the late 1970s and early 1980s. The inference that was conducted in these systems used a wide range of knowledge sources, including factual knowledge about the world, and the ability of these systems to compute truth values was essential to disambiguating natural language expressions. Meaning and interpretation were often not clearly distinguished, on the other hand, and the authors of these interpreted texts were often regarded as having a perfect competence of language and being immune to error [Hirst, 2007].

Advances in speech processing together with Grice’s [1969] influential contributions to the study of intention resulted in a parallel effort to discover not only an objective interpretation of meaning in text, but the speaker’s or author’s intent. Discourse analysis refers to the study of this in written text, especially as it pertains to determining the overall structure of a text’s logical argument. Dialogue systems research attempts to recognize plans and intentions in speech transcripts and to respond to them constructively and naturally.

3 SEMANTICS IN CL

Much of the recent CL research that uses inference is dedicated to question answering, an idea that originated in the information retrieval community, in which queries to search engines that have access to very large databases of documents return not a ranked list of complete documents related to the query, but a ranked list of actual answers, possibly supported by evidence from the original sources. Thus, “Who won the 1989 Nobel Peace Prize?” would not return web pages or articles related to the Nobel Prizes, but the answer The Dalai Lama of Tibet or another short phrase that refers to the winner.

If a fixed set of questions can be determined in advance, question answering becomes very similar to information extraction, which can be as simple as identifying the people, places, corporations and other named entities in a text, involve extracting two-place relations that connect those entities as they are described in the text, or require a more detailed template to be filled in for each text. An information-extraction approach is common, for example, when the documents are known to describe the same kind of event, e.g., news reportage of a terrorist attack (where and when did it take place, how many were killed, who took credit for it). As the name implies, the document in this task is perceived not so much a conveyance of meaning as a somewhat haphazardly organized collection of information, much of it extraneous. Our job is to recover the needles from the haystack.

Another notable exception has been the Pascal Recognising Textual Entailment (RTE) Challenge [Dagan et al., 2005; Bar-Haim et al., 2006]. In this competition,
every system is provided with several pairs of sentences, and the systems must answer in one of four ways for each pair: the first sentence entails the second, the second entails the first, the two sentences entail each other (equivalence), or no entailment relationship exists. The participants are provided with a common corpus of sentence pairs, annotated with the correct entailment answers, to develop their systems with. Many of the participants in this challenge do not agree with the answers provided by the annotators, mostly because entailment is often only established in the presence of background knowledge, and the amount and kind of permissible background knowledge is difficult to gauge.

Summarization, whether of text or speech, can be thought of as the search for the document that comes closest to conveying the same meaning as the original while not exceeding $N\%$ of its size. $N$ usually falls somewhere between 5 and 30. With a compression rate this small, some part of the original meaning must go unstated, of course, but the closest documents are to be found by retaining the most salient parts. What counts as salient is not always easy to determine. Sometimes a cue is provided in the form of an initial query (query-based summarization); this is essentially an acknowledgement that salience is determined in part by the reader’s own value system. On other occasions, another, typically older document serves as a proxy for background information that is no longer salient (update-based summarization). This acknowledges that salience can alternatively be determined in the context of mutually shared belief, either by a failure to infer the former from the latter, or by the observation that the former contradicts some consequence that is usually inferred from the latter. The summary often consists only of sentences selected from the original document (extractive summarization). Otherwise, the summary contains text that has been generated automatically by the system for inclusion (abstractive summarization). In either case, a summarizer must address many of the same difficulties that arise in generation, such as ensuring that referring pronouns are used naturally and with their correct referent, and coherently ordering the summary sentences [Mani, 2001].

Paraphrase is a recent alternative in which there is no stated attempt to compress the original text, but only to rephrase it. Paraphrase algorithms typically operate at the sentence level, not the document level, and have drawn much of their inspiration from statistical machine translation research, by considering the defective case in which the source and target languages are identical.

A number of other very intriguing classification problems have recently emerged that involve aspects of traditional reasoning, but somehow manage to avoid inference altogether. There are classifiers, for example, that attempt to identify the scope of downward entailing contexts within sentences [Danescu-Niculescu-Mizil et al., 2009], and classifiers that attempt to classify documents as having either positive or negative sentiment about a topic. Sentiment analysis [Pang and Lee, 2008] looks not at propositional content, but at the manner in which it is expressed to determine an attitude or disposition towards it. It is generally assumed in this work that a determinate sentiment is deposited by the author in the text where it can be discovered, without consideration of whether readers will agree that there
is a consistently expressed sentiment, what it is, or whether word choice and the
other cues used to express it will vary through time.

4 BAYES’S RULE

Perhaps the most important equation in modern CL is Bayes’s Rule. Suppose we
have two random variables, $I$ and $O$, for the input to and output, respectively,
from a process that performs a task that we wish to imitate, such as language
translation, and we wish to model their joint probability. It is a commonsense
assumption about how conditional probabilities relate to joint probabilities that
the joint probability $P(I, O)$ can be computed by determining the probability of
the value of one of the random variables by itself, and then multiplying that by
the probability of the value of other random variable, given the first:

$$P(I, O) = P(I|O) \cdot P(O) = P(O|I) \cdot P(I).$$

It does not matter which random variable we consider first, so this gives us two
equations. But if this is true, then, assuming there are no zeros, we also have:

$$P(O|I) = \frac{P(I|O) \cdot P(O)}{P(I)}.$$

This is a simple, algebraic derivation, but the late Rev. Bayes assigned to it a very
special significance: he said it encapsulates a view of scientific reasoning or belief,
if we allow ourselves to interpret probability in this way. In this view, $P(O)$ is a
prior that represents our reasoning or belief about the distribution of unobserved
samples from a random process, $O$. Then we make some observations of another
random process, $I$, which we know to correspond to samples of $O$ according to
the conditional distribution $P(I|O)$. Based upon the evidence supplied by ob-
servations of $I$, we may then change our minds about $O$, and this is reflected in
the posterior, $P(O|I)$, which represents our reasoning or belief about $O$ in light
of those observations. This must be renormalized by the marginal $P(I)$ to be a
legitimate probability with a range of between zero and one.

In CL, the underlying view about probability is not terribly important; Bayes’s
rule is often used side-by-side with treatments of probabilities that are in every
other respect frequentist (in which probabilities are assumed to be the limit of
relative frequencies taken over ever larger sample sizes). For CL, what it provides
is a theoretical justification for combining two very important sources of knowledge.
The first ($P(O)$) forces us to produce outputs that look legitimate. In a German-
to-English machine translation system, for example, there is no point to producing
output that looks nothing like English. The second ($P(I|O)$) forces us to be
faithful to the relationship between the input and the output. To consider machine
translation again as an example, it is very easy to produce perfectly natural-looking
English sentences if they are not required to have the same meaning as the German
input. We must do both to be considered successful at translating German to
English.
4.1 Language Models

In CL, priors often go by the name of language models. We usually work with sequence data (strings of words, for example), and a language model tells us whether a particular sequence of words looks like one we might typically find in the text of some language. The most common kind of language model is called an $n$-gram model, which assumes that the previous $n - 1$ words of context are sufficient to predict the next word, no matter how long the sequence of text ($N \geq n$) that we want to know the probability of:

$$P_n(w_1 \ldots w_N) = \prod_{i=1}^{N} \frac{C(w_i, w_{i+1}, \ldots, w_{i+n-1})}{C(w_i, w_{i+1}, \ldots, w_{i-n+1})}$$

$$P_n(w_N|w_1 \ldots w_{N-1}) = \frac{P_n(w_1 \ldots w_N)}{P_n(w_1 \ldots w_{N-1})} = \frac{C(w_N, w_{N+1}, \ldots, w_{N+n-1})}{C(w_N, w_{N+1}, \ldots, w_{N-n+1})}$$

In the definition of $P_n(w_1 \ldots w_N)$, we take every negatively indexed $w_{-j}$ to be some special symbol that is not part of the lexicon. This formulation assumes that we have “primed” the generative process with $n - 1$ instances of this symbol before starting to generate text. $C$ is a count that provides the number of times that its argument has occurred in a training corpus, which is assumed to have been sampled from the same distribution.

It is sometimes said that $n$-grams assume that the occurrence of a word is independent of all of the previous words except the most recent $n - 1$. There is a degree of independence there, but this is not entirely true; an $n$-gram model assumes that all of a word’s dependencies on distant, much earlier words can be completely characterized by factoring them into how the most recent $n - 1$ previous words depend on them. This assumption, together with the assumption that each $n$-word tile is sampled, one after the other, from a time-invariant binomial distribution (like a coin toss or a die roll, only this die has as many sides as the size of the language’s lexicon taken to the $n^{th}$ power) makes this a Markov process.

There are language models other than $n$-gram models. Typically, they incorporate some amount of overall syntactic structure or semantic restrictions onto word choice. These alternatives generally cannot beat $n$-grams on the classical language modelling problem of predicting the next word in a sequence, but they are used to model other processes in which their extra structure is more explicitly required.

Given an unlimited amount of data and computing resources, the quality of the language model by the standards of any application context would improve as $n$ increases. In practice, this is not the case because there are many $n$-gram sequences, even for not so large $n$, that occur very rarely, which makes their probabilities very difficult to estimate. Even today, when we have access to a massive amount of English text on the World Wide Web (approximately 84% of the World Wide Web’s text is in English), we still do not have enough text to reliably model $n$-grams beyond $n = 5$, and in all but the most resource-intensive of applications, $n$ is still usually set to about three (called a trigram language
The sparseness of word distributions is the bane of CL’s existence. The distribution of word frequencies is known to be hyperbolic in human languages — the $i^{th}$ most frequent word type has a relative frequency that is proportional to $1/i$. At the level of 1-grams, this property goes by the name of the Zipf-Mandelbrot equation or Zipf’s first law. Much of the research that CL undertakes is devoted to working around this empirical fact.

### 4.2 Channel Models

The likelihood $P(I|O)$ is generically called a channel model, in honour of Claude Shannon’s early work on using probabilistic models to reconstruct acoustic signals that were transmitted over noisy channels. Here the “input” is a code that has been received by us over a noisy channel, and so our input is the channel’s output, and our output — what we wish to reconstruct — is the channel’s input. Thus $P(I|O)$ represents the possible distortion of the channel’s output given its input. Reconstructing the channel’s input is often referred to as decoding, and it is important in these algorithms to understand that roles of input and output have been reversed.

For different CL applications, the channel model is typically given a more fitting name, such as one of those in Table 1. In some of these applications, using a channel that operates in the reverse direction to the task being solved has the added benefit of making the model easier to estimate. A speech recognizer, for example, receives a sequence of acoustic data and produces a sequence of words as text. To directly model $P(O|I)$ would require us to somehow obtain multiple, identical or nearly identical instances of the same acoustic data in order to estimate the probability of individual sounds or words being spoken at those times. $P(I|O)$, on the other hand, merely requires multiple acoustic observations of the same sounds or words being spoken, which can be collected by recording one or more speakers repeating the same words.

### 4.3 Word-Sense Disambiguation

Returning to the lesson of Kaplan [1955], we may attempt to disambiguate word sense using the adjacent words of its context, a probabilistic model and Bayes’s
rule. In this example, we will make a few simplifying assumptions, all of which are commonplace in word-sense disambiguation research.

The first is that homograph disambiguation is word-sense disambiguation, and all of word-sense disambiguation can, at any rate, be thought of as a discrete classification task, as would be appropriate for homograph disambiguation. In this task, several well-defined alternative meanings are known to exist for a word, \( w \), and we must choose one from among them or rank them in order from best to worst for each instance of \( w \) that we observe in a corpus. Thus we find ourselves taking the very un-Bayesian step of stripping away the probability distributions in search of one or more highly ranked alternatives, \( \bar{s} \), given a context of words, \( c \):

\[
\bar{s} = \arg\max_s P(s|c)
\]

Here, \( s \) ranges over the alternative senses of \( w \).

The word, \( bank \), can refer to a financial institution or to the shore of a river. While this word is ambiguous, these meanings are generally not regarded as two senses of the same word, but as meanings of two different words that happen to be spelled (and pronounced) the same way. \( Bank \) is a homograph. A proper word-sense distinction, on the other hand, tends to be a finer difference between two related meanings. A corporation that is in the business of providing financial services, and a building occupied by such a corporation to transact financial services are two different senses of one of the \( bank \) homographs above, for example. Verbs are known to be more often polysemous (having many senses) than homographs (written the same as other words). \( Swim \), for example, is identified by the WordNet lexical database [Fellbaum, 1998] as being ambiguous between the act of self-propulsion through water from one location to another and the act of treading water in place so as to stay afloat. Many of these fine-grained sense distinctions are remarkably difficult to disambiguate; some would argue that many instances of such words could not be assigned a single sense absolutely, but only compared to their use in other contexts and to other words’ meanings on a more relative or fluid scale. Much of the recent research on word-sense disambiguation has reconnected WSD to its roots in machine translation, treating it as a lexical choice task in which the lexicon of the target language is effectively used to enumerate the word senses or homographs of ambiguous words in the source language.

The second assumption is that one word sense is used per discourse. This assumption allows us to use as much context as we think is useful, regardless of where other occurrences of the same word, \( w \), in the same discourse may be, and to make our classification based on the aggregation of the evidence provided, without worrying about which context word conditions which instance. The third is that the order of the context words does not matter. The result is sometimes referred to as a bag-of-words model, in which context words are merely counted. This is particularly useful for reducing problems in computational linguistics to well-studied classification and regression algorithms in computer science that work on large-dimensional vector spaces, because each dimension of the vector can store the count of one of the context words. On occasion, we only care whether a context
word appears within a certain sized window of \( w \) or not, in which case the bag-of-words model uses only ones or zeros as the values of its vectors.

The final assumption is known as the \textit{naïve Bayes assumption}. This says that the occurrences of these context words are all conditionally independent of each other. This is not the same as Bayes’s rule, although it becomes particularly relevant in the presence of Bayes’s rule. Because \( P(s|c) \) is hard to measure directly, we use Bayes’s rule:

\[
P(s|c) = \frac{P(c|s)P(s)}{P(c)}
\]

whereupon the naïve Bayes Assumption simplifies one of these factors:

\[
P(c|s) = \prod_{v_j \in c} P(v_j|s)
\]

\( v_j \) is a single word type that appears in context. The resulting model is then:

\[
\tilde{s} = \arg\max_s \frac{P(s) \prod_{v_j \in c} P(v_j|s)}{P(c)}
\]

Notice that our decision to focus on the maximum value of \( s \) removes the need to estimate \( P(c) \) as it does not vary as a function of \( s \). Both kinds of probabilities in the resulting product can be estimated by using relative frequencies from a corpus in which the instances of \( w \) have been annotated with their correct sense:

\[
P(s_i) = \frac{C(s_i)}{C(w)} \quad P(v_j|s_i) = \frac{C(v_j,s_i)}{C(s_i)}
\]

Here, \( C \) refers to counts from the corpus of instances of \( w \) (\( C(w) \)), instances of \( w \) that have been annotated with a particular sense (\( C(s_i) \)), or sense-annotated instances of \( w \) that occur together with a particular word of context (\( C(v_j,s_i) \)).

### 4.4 Part-of-Speech Tagging

A part-of-speech tagger attempts to annotate every word in a text with its part of speech, such as \textit{noun}, \textit{verb} or \textit{adjective}. Assigning these labels is important because it serves as the first step in a number of other tasks, including word-sense disambiguation (some homographic pairs have different parts of speech) and parsing.

The set of tags that should be used in the annotation is still a matter of some debate, mostly because of a tension that exists between tags that human annotators can quickly assign with a high level of agreement, and tags that embody several cross-cutting categorical classifications in a taxonomy that accommodates a large number of genres and languages. Ideally, we would like both, but this is difficult to achieve. Some tagsets for English are internally ambiguous, for example, in that the same tag serves to label both gerunds and participles because both
end in -ing. Many English tagsets are inconsistent with each other. Some choose to tag subordinating conjunctions with the same label as coordinating conjunctions, for example, because they are both conjunctions, while others tag them as prepositions, because many words that can be used as subordinating conjunctions can also be used as prepositions.

Many words have more than one possible part of speech, especially in English, in which there is a collection of regular semantic processes that allow most nouns to be used as verbs. It is assumed that every instance of a word in context has a unique part of speech, however. This qualitative characterization of the distribution of parts of speech led to the naïve assumption early on that syntagmatic information about a word’s part of speech, gained from the statistical analysis of sequences of part-of-speech tags in a language-model-like fashion, would be more useful than paradigmatic information about a word’s part of speech, gained from relative frequencies computed over all of that word’s various instances. An example of the first is the observation that DT-VBD-NN (determiner followed by a past-tense verb followed by a common noun) is much less frequent than DT-NN-VBD. An example of the second is that can is more often used as a verb than as a noun, whereas duck is more often used as a noun than as a verb. In fact, paradigmatic information alone is a remarkably good baseline, and far better than syntagmatic information alone, although combining both yields better performance than either by itself.

For a simple method for combining the two, we can again resort to Bayes’s Rule:
(4) that word emissions only depend on the corresponding tag, not on adjacent tags. Independence assumptions such as these are common in CL not because we agree with them, but as a practical consideration because they result in probabilistic parameters that are easier to estimate. We also assume that tag sequences are generated by a Markov process, which simplifies to the form in (5), where we take \( P(t_1 | t_0) \) to be defined as the probability, \( P(t_1) \).

This model in (5) is called a hidden Markov model or HMM. HMMs are important not just because of their widespread use in CL, but because they establish an important connection to an area of computer science called automata theory. As shown in Figure 1,

![Figure 1. An automaton-based depiction of an HMM. The probabilistic interpretation of the automaton is reflected in the range of the numerical weights (between 0 and 1), and the source-normalization of the weights, i.e., the sum of all of the emission probabilities at a single state and the sum of all of the transition probabilities from a single state are both 1.](image)

HMMs can be thought of as probabilistic finite-state transducers, in which probabilities of the form, \( P(t_i | t_{i-1}) \), correspond to probabilities that the automaton transits from state \( t_{i-1} \) to state \( t_i \), and probabilities of the form, \( P(w_i | t_i) \), can be thought of as emission probabilities that control the output of symbol \( w_i \) at state
By convention, the indices used on $w$ and $t$ do not enumerate all of the possible words and tags in succession, but tell us which word was emitted at time $i$ and which tag state the automaton passed through at time $i$ in the course of generating some output sequence. As a result, some words and tags will be associated to more than one index, the indices will change if the output sequence changes, and it is possible for the automaton to pass through the state called $t_i$ at a different point in time than $i$, at which it generates some other word than the word called $w_i$, and arrives from a different state than the one called $t_{i-1}$.

Hidden Markov models are “hidden” in that we cannot see which state the automaton is actually in at any given point in time. Instead we are forced to guess the most probable sequence of states, given a sequence of output symbols (which we can see) and the transition and emission probabilities of the model. The fact that the model emits words from states that correspond to part-of-speech tags is a consequence of using a noisy channel model. Part-of-speech tagging is then accomplished by “decoding” a sequence of word emissions back to their most probable input state (tag) sequence. The terminology is very faithful to the philosophical underpinnings of this area: the tags are really there in the text, but we cannot see them.

### 4.5 Machine Translation

The noisy channel model can even be applied to machine translation by regarding German, for example, as a distortion of English through a channel with a very peculiar kind of noise. This is not the same as saying that German has no grammar of its own, but rather that the grammar of German can be completely characterized by a factorization once again, this time into regular, observable correspondences between German and English, together with the grammar of English. To find the most probable English translation of our German input, we apply Bayes’s Rule:

$$
\hat{E} = \arg\max_{E} P(E|G) = \arg\max_{E} P(G|E) \cdot P(E)
$$

which yields the product of a translation model $P(G|E)$ with a language model of English, $P(E)$. The language model serves as a proxy for a grammar of English — it predicts what English strings look like. There has been a great deal of research on using linguistically better informed models than $n$-grams for $P(E)$ in the context of this task.

The translation model can be constructed by way of hypothesizing alignments between English and German words, based on regular correspondences between German word occurrences in a corpus of German sentences and English word occurrences in a manually created English translation of the German corpus [Berger et al., 1996]. Figure 2 illustrates an example alignment for one pair of sentences. All of our alignments allow a single English word to map to many German words, but every German word must correspond to a unique English word. There are
Yesterday I went home with my sister

Gestern bin ich mit meiner Schwester nach Hause gegangen

Figure 2. An example alignment between a German sentence and its English translation.

different word orders, both in English and in German, that can be used to express the same proposition, but alternative word orders and alternative phrasings with different words are only evident in this model to the extent that the same German sentence fragment appears more than once in the corpus, each time with a different English translation.

Ideally, $P(G|E)$ would be found by summing over all possible alignments $A$:

$$P(G|E) = \sum_A p(G, A|E)$$

but for reasons of efficiency, we generally approximate this by making one very good guess at an alignment, and assuming it to be true. Thus:

$$\bar{E} \approx \arg\max_E p(G, \hat{A}|E) \cdot p(E)$$

We find the alignment along with the translation, by building $P(G, \hat{A}|E)$ from three different parameters:

$$P(G, \hat{A}|E) = \prod_{i=1}^{\vert E \vert} P(n(e_i)|e_i) \cdot \prod_{j=1}^{\vert G \vert} P(g_j|e_{a_j}) \cdot d(\hat{A}|E, G)$$

$n(e_i)$ is a random variable called the fertility of the English word $e_i$. It tells us how many German words are aligned with $e_i$ in $\hat{A}$. The sum of these fertilities tells us the length of the German string $\vert G \vert$. Having determined that length, for each German word to be generated in order of their English correlates, $P(g_j|e_{a_j})$ is the lexical transfer probability that tells us which German word to use, given its English source. Then the distortion model $d(\hat{A}|E, G)$ reorders the German words to a more typically German word order in light of the German words chosen and the English input.

The fertility and lexical transfer models are estimated by training on a bilingual corpus of English-German sentence pairs. The distortion model is often trained on alignments that have been preprocessed by very language-pair-specific word-order transformation rules in order to force the empirically observed alignments into a simpler class of functions. For example, in German-to-English translation, a large number of German clauses render the main verb and all of its phrasal arguments except the subject in exactly the opposite order that their translations appear
in English (the verb appears last, temporal adverbial phrases precede locatival
adverbial phrases, etc.). Performing this reversal step on the German input often
simplifies the subsequent alignment.

5 SYNTACTIC STRUCTURE IN CL

Classically, parsing referred to the analysis of words into their meaningful com-
ponent parts, called morphemes, but it is now universally used in CL to refer to
the analysis of sentences into component phrases. This shift in usage corresponds
neatly to the difference between the syntax of highly inflected languages such as
Ancient Greek, Latin and Sanskrit and the syntax of English, in which word order
plays a far more prominent role. The analysis of words into their component mor-
phemes is now referred to as morphological analysis. This is still a very necessary
first step in parsing qua sentence analysis, as well as in text-to-speech synthe-
sis, because of the crucial role that morphological structure plays in the proper
pronunciation of words.

What constitutes a phrase is unmistakably clear in CL, although usually not
for the right reasons. Apart from the usual empirical problems with defining
constituency in generative linguistics, most syntactically annotated corpora in CL
admit no direct way of indicating a partial or total lack of commitment to the
phrase structure of a sentence in the corpus, even when contemporary linguistic
theory provides no insight or wildly conflicting views on the matter. What we find
instead often looks very arbitrary and is led by the chosen syntactic representation,
e.g., headedness assigned to the leftmost word in a phrase that has no head simply
because the formalism (or, possibly, the algorithm) requires one to be assigned to
every phrase, flat tree structures where there is no agreement on where a subphrase
should attach to the rest of the tree, head-modifier relationships that are posited
only because the branches of a dependency tree are not allowed to cross, etc. For
this reason, and because of the substantial amount of linguistic expertise that
is required to formulate or even correct an annotation, the vast majority of CL
research on statistical parsing and generation accepts parses as objets trouvés in
the annotated corpora that they are trained on. Statistical algorithms in this area
are therefore mostly abstract classification tasks with far less input from linguistics
than one might suppose.

This is in stark contrast to the approaches to parsing and generation that were
prevailing before 1990. In this approach, grammars were not learned from corpora,
but written by hand — often the same hands that wrote the systems, because
every algorithm seemed to come with a warning label that it would not terminate
on grammars in which a particular form of syntactic rule was used. There was
a dizzying assortment of search control strategies that interleaved the prediction
of structure with the traversal of sentence input so as to capture every possible
grammar within some particular class. The goal of the parser plus grammar was
to find all and only the correct analyses of its input sentences, and vice versa for
surface realization algorithms. While a set of sentences may have been carefully
Figure 3. Two syntactic analyses of an English sentence with a prepositional-phrase attachment ambiguity.

delineated in an unannotated corpus — these often served to indicate the minimum amount of coverage that the system had to have — the issue of what it meant to be a correct analysis was a complex and often introspective process in which computational linguists personally engaged. There was a constant tradeoff between building restrictions on syntax into the system itself and building them into the grammar, as well as a tradeoff between the informativeness of the resulting syntactic structure and the coverage of the grammar. Many attested sentences often received no analysis.

Such is not the case with statistical parsers, at least, most of which assign at least some analysis to every input, regardless of how well-formed it is. The grammars that underlie most statistical parsers and generators are implicitly understood to massively overgenerate analyses, but most of those analyses receive very low numerical scores, and so will be outranked by correct analyses.

Some degree of polymorphism if not overgeneration is in fact required of grammars because human languages, unlike, for example, computer programming languages are inherently ambiguous, and not just at the level of part-of-speech tag labelling. Pairs of phrase structure analyses such as those in Figure 3, for example, are said to indicate a syntactic ambiguity in the analyzed sentence. The location of the subtree for the prepositional phrase with a telescope in (a) reflects a reading of the sentence in which telescope is a possession of the man seen, whereas the attachment in (b) reflects a reading in which the telescope is an implement used by the speaker to see the man. This is known as an attachment ambiguity. Consider, however, the following pair of sentences:

- Three boys carried a piano.
- Three boys carried a light bulb.

Both of these sentences are ambiguous because of a collective-distributive ambiguity as to whether the three boys collectively carried a single object, or each one carried his own object. In the case of both ambiguities, we can prefer a particular reading because of facts at our disposal about telescopes, pianos and light bulbs...
such as how they are typically used or their weights. Nevertheless, the ambiguity in Figure 3 is regarded as a syntactic ambiguity, and collective-distributive ambiguities are regarded as semantic ambiguities because conventional phrase structure has no structural means of distinguishing between collective and distributive readings. The distinction between a syntactic ambiguity and a semantic ambiguity is thus not completely clear because it relies on an explicit choice of which ambiguities to treat within the formal representation defined by the grammar.

5.1 Phrase Structure Grammars

In CL, those representations generally take one of two forms. The first is given by a phrase structure grammar that generates analyses such as those shown in Figures 3 and 4. Again, two potential analyses of the sentence Fed raises interest rates are shown as trees. This time, the difference can be traced back to an ambiguity in the assignment of part-of-speech tags to the words raises and interest, and, as a result, is perhaps a more clear-cut case of syntactic ambiguity. In phrase structure analyses, the part-of-speech tags are the labels of the tree nodes that directly connect to the nodes that are labelled by the words of the sentence. This difference, too, leads to different semantic interpretations arising from the choice of a different word as the main verb.

Generally, the tree branches cannot cross, which means that every such annotation can in principle be generated by a context-free grammar (CFG), such as the one shown in Figure 5. Here, context-free refers to the convention that allows any occurrence of a leaf at the bottom of a partial phrase structure tree to be replaced with one of the local trees that has the same label at its root as the leaf, regardless of any other properties of the tree. Because all of the local trees in a context-free grammar are labelled by syntactic categories, a tree in which all of the leaves are labelled by words cannot be expanded any further.

A grammar is said to be more or less lexicalized if the assignment of structure is more or less tightly constrained by the words that appear in the input. CFGs
are generally not very lexicalized because many of their local trees have no word-labelled nodes, as in Figure 5. There are grammar transformations that establish that property without changing the language generated simply by changing the local trees, but it is more common to enhance the syntactic category labels so that each one includes the head of the subtree rooted at that label (Figure 6). The head is a syntactically distinguished word from which most of the syntactic properties of a subtree’s entire phrase are inherited. It is also common to combine the local trees of the grammar into larger local trees that are more than one tree level deep, as shown in Figure 7, where again the scope of headedness determines the extent of the tree fragments. Lexicalized grammars are important for computational linguistics because words are the last remaining vestige of world knowledge in modern syntactic representations. Statistical parsers, for example, can condition the attachment of the prepositional phrase, with a telescope in Figure 3 on the words saw, man, with and telescope to prefer the higher attachment that results in reading telescope as the speaker’s instrument of seeing.

Context-free grammars are not the only class of grammars that can generate these trees; tree-adjoining grammars [Abeillé and Rambow, 2001] are a well-known alternative. A branch of computer science research called formal language theory that also has its origins in the work of Chomsky [1959], is devoted to determining which sets of strings of words can be generated by different classes of grammars. Tree-adjoining grammars can generate some languages that context-free grammars cannot. There is a proof that at least some human languages cannot be generated by context-free grammars [Shieber, 1985], but most of the classical results of formal language theory, including this non-context-freeness proof, took place without considering statistical methods or numerically parametrized grammars. In the latter setting, it is possible for a probabilistic context-free grammar not only to assign
too many analyses to individual strings, but to assign analyses to a much larger set of strings than the sentences of the human language under analysis, giving very low probabilities to those that are not grammatical sentences. Already in 1963, Rabin [1963] had proved that if the language recognized by a probabilistic automaton is defined as the set of all strings that can be generated by the underlying discrete automaton with a probability of greater than some real number $0 \leq \lambda < 1$, then there are some values of $\lambda$ for which the resulting language cannot be generated by any discrete automaton, i.e., the use of probabilities actually adds something to the expressive power of this class of grammars. Furthermore, any $\lambda$ that adds this extra expressive power must be irrational, which means that no computer can realize this extra potential because of the limits on the numerical precision of its calculations. Fowler [in press] shows that probabilities also extend the expressive power of context-free grammars, but in a direction that is incomparable to tree-adjoining grammars, i.e., probabilistic CFGs cannot be used to simulate tree-adjoining grammars.5

Algorithms for statistical parsing and surface realization generally do not make use of Bayes’s Rule, although they often do incorporate language models. A good introduction to statistical parsing with phrase-structure-annotated corpora can be found in Jurafsky and Martin [2008]. Surface realization algorithms are often neglected by general introductions to computational linguistics, but Langkilde-Geary’s [2003] is a fine example of one that uses a statistical model.

5.2 Dependency Grammars

The second form of syntactic representation is generated by a dependency grammar, which produces analyses such as the two shown in Figure 8. There is a

---

5This is not the same as considering probabilistic languages, in which each string is paired with a probability merely by virtue of its membership in the set. Ellis [1969] proved that there are probabilistic languages that cannot be generated by probabilistic automata even though their support (the set of strings that have non-zero probabilities) can be generated by a discrete automaton. Kornai [in press] proved that this result holds even if the probability values are all rational.
straightforward mapping from every phrase structure analysis to a dependency analysis, provided that heads are uniquely identifiable in the phrase structure analysis. The reverse mapping is not as easy. Dependency structures are very popular in CL in part because there are far fewer of them; it is normally an invariant of dependency grammars that there are as many nodes in the dependency tree for a string as there are words in the string. This makes them easier for annotators to agree upon, and less arbitrary with respect to the higher phrasal structure that they specify.

Dependency structures are able to capture argument structure, the part of a sentence's structure that specifies “who did what to whom,” succinctly and accurately. There is plenty that they do not specify accurately or do not specify at all, however, much of which is necessary in order to bridge between syntactic structure and the logical representations that semanticists typically use:

- **Order of argument application**: dependency trees in which a head takes multiple arguments cannot distinguish an order of application or priority of combination with those arguments, as all of the arguments appear as equally ranked subtrees of a node corresponding to the head. Some dependency analyses label the edges of the tree with grammatical functions, thematic roles, numbers or other information that could in principle replace an order of application.

- **Discontiguous dependencies**: while arguments for constituency are equally problematic for phrase structure and dependency analyses, there is at least some semantic basis to constituency in both approaches. In sentences such as *A woman arrived who was wearing a hat*, the relative clause *who was wearing a hat* is usually taken to modify the noun *woman*, but this cannot be depicted in a dependency tree unless either (1) the leaves of the tree do not spell out the original sentence when read from left to right, or (2) the branches of the tree are allowed to cross. Most dependency analyses reject either possibility; these are called projective dependency analyses. A guarantee of projectivity makes dependency parsing much faster, but at the cost of failing to provide semantically transparent analyses of sentences such as this one. Phrase structure analyses also have problems with non-projectivity (there called the no crossing branches constraint), but their richer internal structure makes it easier to formulate repair operations (often called movement) that make it possible to observe semantic dependencies without crossing branches.

- **Complex word-word dependencies**: Analyses of head coordination are notoriously difficult in dependency analyses. In sentences such as *Ginger and Fred danced and sang*, for example, *and* is often taken as the head of coordinate phrases such as *danced and sang* in dependency analyses, but to assign it the argument *Ginger and Fred*, which also has *and* as a head, provides very little insight into the arguments to these two predicates semantically. Another difficult example is predicate adjective constructions such as *I watched*
the film alone, where, again motivated by a desire for simplicity of semantic interpretation, alone clearly cannot modify the same word that in colour does in the dependency analysis of I watched the film in colour.

- **Semantic role assignment:** The verb opened in the sentence The door opened a crack is clearly assigning semantic roles in a pattern different from John opened the door, but opened is merely a node with two arguments in the dependency analysis of both. In phrase structure, there is a wider variety of internal structures that the subtree for opened and its arguments can avail itself of. Semantic role labelling has become a research topic in its own right [Gildea and Jurafsky, 2002], motivated largely by a dissatisfaction with how transparently semantic roles can be expressed by purely tree-configurational representations, whether dependency-based or phrase-structure-based.

- **Scopal elements:** Words that define a scope, including quantifiers, interrogatives, and verbal morphology or adverbs that are involved with temporal interpretation, are also in want of extra tree positions in a dependency analysis that could be used to indicate their scope, which phrase structure provides to a greater extent, along with more repair operations to ensure the proper readings.

Until very recently, dependency grammar lacked an equivalent stratification into formal language classes that allowed us to speak of their generative capacity, the ability of a grammar formalism to define certain sets of strings or certain abstract syntactic structures. A very fruitful, recent trend [Kuhlmann, 2007] has been to view dependency structures not as the outputs of unwieldy dependency grammars, but as derivative extracts of phrase structure grammars, or to view them from the opposite direction, as annotated generalizations of strings. These intermediate structures can further stratify the classes of phrase structure grammar that are traditionally held to be equivalent only by virtue of the sets of (unannotated) strings that they generate. For most practical applications, syntactic analysis is a vehicle to understanding, so this small amount of extra structure about “who did what to whom” makes generative subclasses based on the dependency structures that a grammar formalism can generate more real and relevant.

A good introduction to algorithms for statistical dependency parsing is Kübler et al. [2009].

**BIBLIOGRAPHY**


THE METAPHYSICS OF
NATURAL LANGUAGE(S)

Emmon Bach and Wynn Chao

1 INTRODUCTION

Natural language metaphysics is the study of what kinds of things, distinctions, and so on are necessary for an adequate account of the semantics of natural language and natural languages [Bach, 1981; 1986b].

What is there? is the fundamental question of metaphysics.
What do people talk as if there is? is the fundamental question of linguistic semantics.

There is an old tradition, still very much alive, that says that the answers to such questions will vary from language to language. So what we believe or presuppose about the nature of reality will differ according to whether we are speakers of Hopi (say) or English.

The language Hopi (Uto-Aztec) is mentioned advisedly. Benjamin Lee Whorf [1956b] claimed that the basic view of reality — in a word, the metaphysics — embodied in the Hopi language was radically different from the basic world view of the Standard Average European languages (which would include Modern English). The latter was supposed to be fundamentally Newtonian: concepts of time and space as absolute containers, separate and rigid, as opposed to the much more Einsteinian or relativistic view underlying the Hopi world view.

It is unclear whether claims like Whorf’s are about a culture or a language, and some would question whether such a division is a help or a hindrance. In any case, we take the point of view here that it is possible to make such a division and ask about metaphysical assumptions that are built into the semantics of a language.

We follow the route of model-theoretic semantics. To spell out the semantics of a language is to associate denotations and other kinds of meanings with the expressions of the language. To give a general account of a framework for doing this for natural languages – Modern English, Swahili, Hopi — choices must be made for the objects that are to be candidates for these denotations: things, functions, numbers, mountains, people, wars, and the like. These choices imply what the universe of meaning is like, and in this sense provide a metaphysics for the language [Quine, 1948; Cresswell, 1973]. An immediate question is then:
What is common to different languages in their universes of meaning and where do languages differ? This is the import of the parenthesized plural in our title.

Linguistic semantics follows the general plan of attack familiar from other sub-disciplines such as phonology or syntax: it must provide a general theory which captures what is common to all languages and provides the means for registering possible differences among them. An essential part of the general theory is a model structure that contains the ingredients for specifying the denotations of expressions of the language(s) under consideration. At the least we need these components:

(i) truth values: \{TRUE (1), FALSE (\emptyset), \cup (undefined)\}

(ii) a set of individuals

(iii) a set of worlds: ways things could be

(iv) all functions that can be built out of the preceding

Note that there are no constraints on what the individuals are. In PTQ [Montague, 1973] the basic set of expressions in the set indexed by the category of names include John, Bill, Mary, and ninety, presumably people and a number.

We suppose (standardly) that expressions in a natural language are interpreted in context so we need to embed our semantics in a pragmatic theory (in one sense of “pragmatic”): this makes it possible to fill in denotations for indexicals (context-dependent elements) like tenses, and words like I, you, here.

Once we have established what is referred to by such indexicals in a given context, we can ask what the semantic value or denotation of the expression in question in a certain world or situation is.

Where and how can differences and generalizations about meaning be registered in semantic theories? We assume the following possibilities, each of which constitutes a major debate in semantics (see [von Heusinger et al., forthcoming] for details and references):

(i) denotations (“semantic values”)

Examples: John Smith denotes John Smith; snow denotes snow; it is snowing here denotes the value true if and only if it is snowing at the place of assertion at the time of the assertion, . . .

In the present context the important contrast is with concepts as meanings. The cat is on the mat makes reference to a real cat and a real mat (perhaps in this world, or another possible world.

(ii) indexicals and other context dependent items, variables

Besides the obvious examples already mentioned, sometimes we need to supply a parameter from “outside”. For example: the white dog might need a local context to support the uniqueness presupposition of the.

---

1 We assume a third truth value, for example, to take care of situations where we want to use sorted domains to enforce collocations and we want to say that a sentence may be neither true nor false. PTQ allows sentences like Ninety seeks a unicorn (more below).
(iii) entailments

*John went to the store* entails that John got to the store.
*John was going to the store* does not.

(iv) presuppositions

*Why did you let the dog out?* presupposes that you let the dog out.

(v) meaning postulates

A meaning postulate is a semantic tool to put constraints on denotations. For example, Montague used meaning postulates to ensure that certain verbs, like *see* guaranteed the existence of a thing seen, as opposed to a verb like *seek* which has (alas!) no such guarantee.

Sometimes this is the best way to understand the next item:

(vi) semantic parameters / features

Features are familiar from other parts of linguistic theory, but may play a role in semantics as well as signalling particular choices within a general domain of interpretations [Chierchia, 1998].

(vii) conventional implicatures

Elements of meaning that are not directly asserted (cf. [Karttunen and Peters, 1979; Potts, 2005]):

*He managed to get the job done* implicates that it was difficult to get the job done.

The term was originally coined by Grice to cover individual expressions whose meaning apparently does not contribute to part of the propositional content expressed in an utterance, but noneless is conventionally associated with a given word [Grice, 1975]

(viii) conversational implicatures

The concept of *conversational implicature* on the other hand, also a term coined by Grice, was to label those aspects of the interpretation of an utterance which are not intrinsically associated with what is expressed by the uttered sentence, but involve some step of inference made by combining what is directly expressed by that utterance with additional assumptions [Grice, 1975]. Upon recurrent repetitions over time, these may give rise to so-called generalized conversational implicatures, and constitute a quasi-idiomatic usage of language, as in utterances of (1) conveying ‘Please pass the salt (if you can reach it)’:

(1) Can you reach the salt?

These concepts, straddling conventional and nonconventional usage of language, have generated a great deal of discussion: see the debates as presented in [Horn and Ward, 2005].
(ii) connotations, associations etc.

These are not usually dealt with in formal semantic theory, but nonetheless constitute important parts of meaning in the broad sense: for example, so-called epithetic words convey attitudes of the speaker toward the referent, the bastard, the sweetheart, etc., and probably contribute to the common intuition that “you can’t really translate from one language to another”.

It is not always easy to know just where certain aspects of meaning are to be accounted for. We will see below (Section 4.4) one case where it is claimed that what is an entailment in one language is a cancelable implicature in others.

In the rest of this essay, we want to survey some of the main kinds of elements that have been put forward as possible further ingredients of the interpretations of natural languages. The means by which these additions or elaborations have been made are basically two: additions have been made to the basic model scheme outlined above, or structure has been attributed to already present or newly proposed domains. An example of the first type is the addition of Properties as an independent domain; an example of the second type is sorting the domain of individuals into Kinds, Individuals, and Stages. We discuss examples of both kinds below.

2 HISTORY

Apart from details about time and worlds, the foregoing is the basic structure of the first introduction of model-theoretic ideas into the analysis of meaning in natural languages from the work of Montague [1973] (others: [Cresswell, 1973; Lewis, 1970; Keenan, 1972; Partee, 1973; 1975; see Partee, 1996]).

One idea that was expressed at that time was that it was possible to separate questions about the meaning of particular words and idioms — lexical semantics — from the recursive specification of the meaning of constructions in syntax:

... we should not expect a semantic theory to furnish an account of how any two expressions belonging to the same syntactic category differ in meaning (Thomason, Introduction to [Montague, 1974, p. 48])

The passage is accompanied by the following footnote:

The sentence is italicized because I (i.e. Thomason) believe that failure to appreciate this point, and to distinguish lexicography from semantic theory, is a persistent and harmful source of misunderstanding in matters of semantic methodology. The distinction is perhaps the most significant for linguists of the strategies of logical semantics.

Some “logical” words were the exceptions to this division: quantifiers, be, conjunctions, and so on. As more details about English and other languages were incorporated into explicit semantic accounts it became apparent that it was necessary to “go inside” the meanings of lexical items.
This principled division foundered on the analysis of such constructions as the English progressive, where it turned out that the truth conditions depended precisely on the meanings of various verbs [Dowty, 1977; Vlach, 1981; see below Section 4.4]. Max Cresswell [1973] was one writer who took it upon himself to spell out the “metaphysics” of his various constructions: propositional languages and lambda-categorial languages used as a bridge (like Montague’s Intensional Logic) to interpret English.

Montague’s fragment PTQ [Montague, 1973] included the absolute minimum needed to illustrate his general proposals: three tenses (present, present perfect, future with will), singular nouns, names, and third-person pronouns, three determiners including the, the sentence adverb necessarily. These were used to illustrate Montague’s treatment of tense, modality, and intensionality. Naturally, in the early years of the adoption of his work into linguistic semantics there were proposals for extending the coverage of English and other languages: English words like proposition, plurals, and other tenses and aspects. We will start with more details about Montague’s work and these earliest extensions in the separate sections below. As we will see, many of the extensions pointed toward a more complicated story about structural and lexical semantics.

The earliest extensions of Montague Grammar in linguistics were aimed at covering a larger part of English, by including plurals, more options for tenses and aspects, propositions, and so on. For the most part these extensions used the apparatus that was already implicit in the basic model structure of PTQ, for example, Michael Bennett incorporated plurals into his treatments, and used sets to model them [Bennett, 1974].

The idea that there is a fundamental difference between the meanings associated with lexical items and with grammatical distinctions is an old one, obviously related to the difference just mentioned between structural and lexical semantics [Jakobson, 1959; Bach, 2005].

Model-theoretic semantics has been based most usually on set-theory and higher-order logic. That is, it countenances sets or classes of objects, functions on those objects, functions on those functions and so on. So insofar as it makes use of these tools, there is a sense in which it incorporates some metaphysical assumptions about such abstract objects. An alternative or supplemental basis is mereology — the theory of part-whole relations [Stein, 1981; Link, 1983; Champollion, 2010].

Another important stream in linguistic semantics comes from the works of Donald Davidson [1967; 1980], who insisted on the importance of events as ingredients in the interpretation of language. We will take up this theme in later sections (especially Section 4.4).
3 BASIC MODEL STRUCTURE

We consider some general questions about the basic model structure in this section. In the next we take up some more special questions about various domains.

3.1 Time and Tense

We take up first a question about the basic model structure outlined above. Montague’s model structure for interpreting English in PTQ includes a set of possible worlds and a set of times. The times are ordered by a relation $\leq$, which is total, transitive, and antireflexive: for all $i, j, k$ (points in time, say):

(i) $i \leq j$ or $j \leq i$
(ii) if $i \leq j$ and $j \leq k$, then $i \leq k$
(iii) if $i \leq j$ and $j \leq i$ then $i = j$.

How can we support or defeat these assumptions about the structure of time in terms of natural language (or English) metaphysics? It obviously won’t do just to ask speakers of natural languages. St Augustine said (Confessions, XI):

“What then is time? If no one asks of me, I know; if I wish to explain to him who asks, I know not.”

And indeed there have been other native speakers of Whorf’s “Standard Average European” (SAE) languages who have argued for diametrically opposed notions of time: Aristotle vs Plato; Leibniz vs Newton. On the one side is the idea that time does not have any independent existence, but is rather constructed from basic relations among events; on the other is something like Whorf’s idea about the SAE conception. But we can no more ask directly of native speakers that they do our analytic job for us in this domain than in other parts of linguistics. In the semantic domain we need to look at judgments of acceptability, inferencing, incompatibility, and so on, which constitute the data relative to which competing semantic theories are judged.

For example, we can test assumptions about time, by asking for judgments about specific expressions in a language:

(2) If Mary had left on her spaceflight, she would now be eating breakfast.

Interpreting this sentence requires reference to two possible worlds: the world in which Mary did leave and the other — the real world of “here and now” — in which she did not leave on her spaceflight. Moreover, it requires identifying times across these two worlds. It is doubtful that the possibility to make this identification should be a semantic assumption, given that it is not possible to identify times in different space-time regions in this world.
In the papers just cited, Bach argued for a rather different way of dealing with temporal relations, based on ideas of Wiener [1917], Whitehead [1920], and Russell [1929], and adapted by Hans Kamp [1980] (see now also [Fernando, 2009]).

The basic idea is that temporal notions can be construed as relations of precedence and overlap among events. A set of events connected by these relations is a (local) history. It is not assumed that all events in a possible world are so connected. With such an underspecified model it is possible to deal with situations or worlds in which series of events can be considered which are locally ordered but where it may not be possible to link across two such series for all events. Kamp gives two examples where such models would be useful: (a) a narrative where two or more story lines connect up at various points but where intermediate steps in the separate story-lines need not be temporally pinned down; (b) several changes where the exact point of the changes is left vague. Such a temporal model would also be well suited to accommodate versions of relativity theory. The main point is to allow a flexible semantic framework within which to consider the interpretation of examples like (2).

There are further questions about the structure and nature of time in a setup like that of PTQ:

(i) Are the elements of the time line points or intervals?

(ii) Does time conform to the structure of the real line, or the rational numbers, or the integers?

and so on.

We believe that time-talk in natural languages is pretty indeterminate in respect to these abstract qualities and that our model structures should reflect this indeterminacy. Some sentences seem to require that we can know what it means for one thing to happen “right after” another, while others seem to admit that between two happenings we can always find something that happened “between them” (compare [Bach, 1981], and the discussion of “achievements” below, Section 4.4).

3.2 Worlds and Situations

In standard possible world semantics, a world is just a way everything can be, everything. A number of writers have proposed the use of situations, understood as something “smaller” than worlds but of the same logical type. This idea has been spelled out in various ways by Cresswell [1973], Kratzer [1989], Barwise and Perry [1983], and others. (See [Kratzer On line, 2007/2009], for a good survey.) One way is to consider situations as partial worlds, with classical worlds being maximal situations. (We return to these ideas in discussing kinds, individuals, stages in Section 4.2.)

If we keep the basic model structure outlined above but reinterpret worlds as situations in this broader sense, we will gain some new options: the value or
extension of an individual (concept) will vary according to the situation it takes as an argument.

Kamp’s Discourse Representation Theory needs a separate rubric here. Although the units of his theory — Discourse Representation Structures (DRS’s) — bear some resemblance to situations in the above sense, they are really something intermediate between a linguistic representation and a model [Kamp, 1981; Kamp and Reyle, 1993; Asher 1993]. They share the quality of partiality with the situations we have just mentioned. Presumably metaphysical questions arise in the embeddings of the DRS’s into classical models.

3.3 Properties

In addition to worlds, times and events, it has been argued that our ontology needs to be expanded to give recognition to properties as an independent semantic type. Suppose people are the only rational animals. Then substituting rational animals for people in any sentence should preserve truth values. Now though this might hold for our world, it is easy to imagine another world in which this equivalence does not hold. The intensionality of possible world semantics, and the notion of sense which this concept is defined as reflecting, allows us to deal with the problem presented by sentences like these:

(3) People have two legs.
(4) Rational animals have two legs.

We can say that in our world (3) and (4) might be equivalent, they might very well not be equivalent in other possible worlds. This is a simple illustration of how intensions can help solve semantic problems. At the level of individuals, intensional meanings for names and some other nominal expressions allow us to solve the problem of sentences about the Morning Star and the Evening Star:

(5) The Morning Star is the Morning Star.
(6) The Morning Star is the Evening Star.

(5) is necessarily true, unlike (6).

Montague used intensional interpretations (senses) to solve such problems. An intensional interpretation in Montague is a function from a world (actually a world-time pair, an index) to some extensional denotation, the value of the function at that index. So even if in our world the terms Morning Star and Evening Star have the same extension, that is, the planet Venus, in some other possible world this might not be true. Similarly for problems arising with examples like (3) and (4). Gennaro Chierchia showed how adding properties as special meanings allows us to deal with problems like these. To take a more everyday example, in every world the set of things that are sold is the same as the set of things that are bought. So necessarily, if we build up a complex predicate compositionally by adding a
reference to an agent, we get the equivalence of sold by Mary and bought by Mary. Adding properties as an independent type of denotation allows us to circumvent this and similar problems [Chierchia, 1984; Chierchia and Turner, 1988].

4 ONTOLOGICAL CHOICES

We now consider some ideas about the nature and structure of special domains within the broad model structures posited for natural language semantics, and the ontology to which natural language distinctions appear to commit us. As Godehard Link put it: “our guide in ontological matters has to be language itself” [Link, 1983].

4.1 Things and Happenings

One of the peculiarities of the predicate calculus, from the point of view of natural languages, is that there is only one kind of predicate, so that the correspondents of dog and run have the same logical type and form. So also in PTQ: common nouns and intransitive verbs are syntactically distinct but map into the same type in the interpretation, sets of individuals or their intensional counterparts — individual concepts. Similarly for relational nouns, transitive verbs, and so on.

From the point of view of natural language, it seems that there is a pervasive difference between things and happenings, independently of how this is expressed in the syntactic and morphological categorizations of the languages in question. In English, for example, there is an affinity between certain predicates and happenings:

(7) The war lasted seven years.

(8) Jonathan lasted seven years.

(9) The parade took place in the center of town.

(10) ?Harry took place in the 20th century.

(7) and (9) are perfectly ordinary sentences. With (8) we have to supply some understood amplification: ‘as chairman’ or the like. (10) is anomalous unless we understand Harry as the name of an event or come up with a metaphorical understanding.

There has been a long-standing discussion about the universality of “parts of speech,” including much discussion about whether all languages distinguish nouns and verbs (and sometimes adjectives): [Sapir, 1921; Bach, 1968; Kinkade, 1983; Jacobsen, 1974; Jelinek and Demers, 1994; Jelinek, 1995; Demirdache and Matthewson, 1995; Baker, 2003; Evans and Osada, 2005]. A number of different questions are involved in these debates, among them the following:
(i) Does every language distinguish between nominal and verbal syntactic constructions?

(ii) Do all languages make a distinction between the lexical categories of *noun* and *verb*?

There is no doubt about the answer to the first question. Every language must be able to make truth-bearing constructions (assertions). And every language has a means to refer to an unlimited number of different entities. Quine’s minimalist language [Quine, 1948] makes do with a single lexical class but must still include variables as a second category (as well as quantifiers and other syncategorematic signs).

Compare Sapir [1921: 119]:

Yet we must not be too destructive. It is well to remember that speech consists of a series of propositions. There must be something to talk about and something must be said about this subject of discourse once it is selected. This distinction is of such fundamental importance that the vast majority of languages have emphasized it by creating some kind of formal barrier between the two terms of the proposition. The subject of discourse is a noun. As the most common subject of discourse is either a person or a thing, the noun clusters about concrete concepts of that order. As the thing predicated of a subject is generally an activity in the widest sense of the word, a passage from one moment of existence to another, the form which has been set aside for the business of predicating, in other words, the verb, clusters about concepts of activity. No language wholly fails to distinguish noun and verb, though in particular cases the nature of the distinction may be an elusive one. It is different with the other parts of speech. Not one of them is imperatively required for the life of language.

(This passage occurs immediately after a paragraph disparaging the idea that parts of speech might be universal. Sapir opts for a language-particular view of such distinctions.)

The second question is altogether different. It is still debatable whether lexical classes like *noun* and *verb* and *adjective* must be distinguished in every language. It is not in question whether these categories are part of a universal stock of available categories nor that *nouns* and *verbs* are specially connected to the syntactic categories associated with naming and referring, and predicating.

There is no doubt that we can refer to the same event using either nominal or verbal expressions. Further there must be logical connections between these references as shown by Terry Parsons’ examples [Parsons, 1990:18]:

(11) In every burning, oxygen is consumed.

(12) Agatha burned the wood.
(13) Oxygen was consumed.

What we have considered in this section is related to a question that will come up below when we talk more directly about events and other species of eventualities (Section 4.4).

The pertinent question in the present context is whether there is a significant semantic contrast between common nouns and verbs. A positive answer to this was given some time ago by Gupta who pointed out that nouns carried with them a principal of reidentification, so that it makes sense to apply the adjective same to a noun, whereas it is difficult to get that idea across in a purely verbal construction:

(14) Is that the same squirrel / man / explosion / burning that you saw?

(15) ?? Did Superman and Clark Kent run identically / the same / samely?

(16) Did Harry run the same run / race / route as Sally?

(Compare [Gupta, 1980], also [Carlson, 1977; Baker, 2003].) Note that this notion of sameness (and difference) can be applied to kinds and ordinary individuals (see next section):

(17) Is that the same dog? Yes, it’s a Basenji. / Yes, it’s Fido.

4.2 Kinds, Individuals, Stages

An early refinement of the simple model given above was introduced by Greg Carlson [1977] in his investigation of generics and related matters. The discussion continued [Carlson and Pelletier, 1995] and has been one of the liveliest topic areas in linguistic and philosophical semantics.

Carlson proposed that the domain of entities should be sorted into three subdomains: kinds, individuals, stages.

Two realization relations were posited to link kinds and individuals, and individuals and stages. We note here a parallelism between the relations of kinds and the individuals that realize them and individuals and stages — the latter something like local manifestation of individuals (compare here Cresswell’s notion of manifestations of individuals, [Cresswell, 1973]).

English has a variety of ways of referring to these various sorts:

(18) Horses disappeared from the New World between 10,000 and 8,000 years ago.

(19) Horses were running down the road.

(20) Equus evolved mostly in the Americas.

(21) Horses are mammals.

(22) A horse is a mammal.
(23) The horse is a noble beast.

(24) The horse is standing ready for you.

(25) Man is not grand.

(26) I hate rabbits, because they are destroying my cabbage patch.

The significance of the last example is that it shows anaphoric reference from one kind of interpretation (\textit{kind}) to another (\textit{stage}). Sentences like that one were taken by Carlson as evidence for locating difference between individuals and stages in the linguistic context rather than in the nominal itself. In fact there is widespread systematic understanding of different senses of lexical items (a big literature on this topic, see for example [Pustejovsky, 1998]).

Are Carlson's \textit{sorts}, or something like them to be found in every language, that is, should we think of them as a necessary part of \textit{natural language metaphysics}? It is certainly not the case that the means of expressing them are the same across languages. A good selection of the variety can be seen in our English examples: bare plurals, definite and indefinite noun phrases, bare Latin names in scientific parlance, and even a bare singular as in (24), a normal pattern in some languages. Let us also note here Daniel Everett’s claim [Everett, 2005] that there is no expression of genericity (or quantification) in Pirahä.

4.3 Mass, Plurals, Counting, Numbers

A similar challenge occurs in the subtyping of kind-denoting expressions. PTQ has eight common nouns: \textit{man}, \textit{woman}, \textit{park}, \textit{fish}, \textit{pen}, \textit{unicorn}, \textit{price}, \textit{temperature}. They are all count nouns, that is they have plural forms (\textit{fish} has two: \textit{fish} and \textit{fishes}), although as we noted the PTQ fragment has no plural forms. The full array of nouns in English comprises several other kinds, among them one — mass nouns — that has gained considerable attention.

An important (and old!) set of problems was brought to the fore especially in the work of Godehard Link on the interpretation of mass, count, and plural terms and predicates appropriate to them [Link, 1983]. In English, mass terms have no plurals, resist collocations with number words:

(27) #There were five muds on the floor.

(28) There were five blotches of mud on the floor

(29) These muds are quite distinctive.

Here, we have to understand the plural in (29) as referring to kinds of mud, and (27) can perhaps be coerced into this kind of understanding.

Plurals and conjunctions of names show several interpretations:

(30) The boys carried a canoe down to the lake.
(31) Sally and Mally lifted the suitcase.

These sentences can be made more precise by *each* or *together*.

(32) We saw dolphins on our excursion.

Is this sentence true if we only saw one dolphin? Group readings for plurals are easy to obtain and real in real life. The same people can meet as a finance committee, adjourn, and meet again as a policy committee [Landman, 2000; Schwarzschild, 1991]. Behind such familiar concepts, there are related questions about the universality of counting and numbers, starting from Everett’s claims about the Amazonian language Pirahã [Everett, 2005; Nevins et al., 2007]. Such discussions raise important points about the nature of universals as overt or covert categories in language, as potential or realized. We wish to draw attention to an important paper by Ken Hale [1975], which underlines the potential character of universals and also the role of cultural needs for one or another of such potential resources. Here it seems that all the necessary ingredients are potentially there for constructing the meanings of numbers, in the very notion of a set, presumably part of the metaphysical furniture of every language.

4.4 Verbal Aspect (Eventology)

Verbal aspect, *Aktionsart* (“eventology”) is the classification of eventualities and/or expressions about them due to Aristotle, Kenny, Vendler, Verkuyl, Dowty, and many others. A major part of the discussion has centered around the possibility and interpretation of (English) sentences using the progressive aspect:

Sentences like those set out below and many others have been used to establish classifications of eventualities into (at least) three types: *states*, *processes*, *accomplishments* (originally also *achievements*). Nowadays, a more favored terminology seems to be *telic* versus *atelic* events for accomplishments and processes, respectively. Terry Parsons, in the most detailed study to date of the role of events in the semantics of a language (English, [Parsons, 1990]), writes of events as having two parts: a development and a culmination (= *telos*), so that processes are simply events with no culmination, accomplishments include both a development part and a culmination.

Here are some examples and discussion:

A. *States*

(33) John is in London.

States resist construal with the frame *it took* . . . [duration expression], *to* . . . or with the progressive

(34) It took John three years to be in London.
Here we have to understand (34) to mean come to be in London or the like. Similarly with (40) below: to come to know the answer. (37) is a stative sentence. Construed with the progressive as in (38) we have to understand the sentence as being about a temporary state. Locative be (33) is most resistant to any kind of reconstrual as an activity or process (35). Compare (35) with the activity be + Adjective (36) [Partee, 1977].

(35) ?John is being in London.
(36) John is being difficult.
(37) Mary lives in London.
(38) Mary is living in London.
(39) ?Sally is knowing the answer.
(40) ?It took Sally three hours to know the answer.

B. Processes (activities, atelic events)

Processes contrast with states; note the acceptable (41)-(46). (The term activity has the difficulty that it connotes agentive involvement, while process is neutral in this dimension.)

(41) Harry is mowing the lawn
(42) Harry mows the lawn.
(43) Ed was mowing the lawn.
(44) Rose was running.
(45) Rose ran.
(46) Rose was running to the store.

C. Accomplishments (protracted telic events)

(47) Jamison crossed the street.
(48) Jamison was crossing the street.
(49) Jamison was crossing the street, when the truck hit him.

Much discussion, started by Dowty [1977], centered on sentences like (47)-(49), illustrating the “imperfective paradox” or puzzle: though (48) seems to entail the truth of (47), (48) can be true without (47) being true, as shown by (49).
D. Achievements? (instantaneous telic events)

(50) Harry realized something.

(51) Harry was realizing something.

(52) As soon as Bill got up, he was up.

Unlike the previous categories, achievements are not as robust. Originally, they were supposed to be impossible to use in the progressive, and this was supposed to be because they were instantaneous while accomplishments were supposed to take time. In our opinion the closest we can come to really instantaneous events are mental ones like realizing (50). Nevertheless we can think of contexts where even such happenings can be construed as processes. Imagine we are watching Harry’s brain with a device (a “chronoscope”) that can slow down events to perceptible rates. Then (51) seems perfectly understandable. But it seems that there is nothing in our language that prevents us from talking about instantaneous events, as in (52).

Here we will look at a different puzzle, centering on sentences about accomplishments:

(53) I fixed the fence.

(54) ?I fixed the fence, but I didn’t finish it (i.e. fixing the fence).

An English sentence like (53) entails that the accomplishment indicated was successful, that is, that it reached a culmination [Parsons, 1990]. It has been claimed that this entailment does not hold in all languages, most recently in various Salishan languages [Bar-el et al., 2005], but that a weaker notion of implicature is correct. Hence, the implicature can be canceled, so that sentences corresponding to ones like (54) are apparently completely unproblematical.

A related puzzle has to do with the interpretations of plain past tense sentences in English and (among other languages) Dutch [Landman, 2009]:

(55) Ik sliep.

(56) I slept

(57) I was sleeping.

(55) can be interpreted either as (56) or (57). The main question here is whether there is a genuine ambiguity here or whether there is an interpretation of the simple past in such languages that covers the denotations that are split up in English under some unified field, just as the denotation of common nouns in languages that do not have obligatory plurals can be understood as denoting something like the union of interpretations for singular and plural nouns in English-type languages.
4.5 Unifications and Parallels

Parallelisms across the various areas discussed have been noted for some time, for example the domains of count-mass-plurality and verbal aspect (see [Bach, 1986a] and literature cited there).

Recently, the parallels have been extended and sharpened in the work of Lucas Champollion [2010], who has posited a formal and substantive theory that unites three areas: verbal aspect (Aktionsart) (58), measurement (59), and distributivity (60):

(58) John left. / John was leaving.

(59) 5 kilos of apples / six feet of snow / *six degrees of snow

(60) all the boys lifted the piano / each of the boys lifted the piano

Champollion provides a semantic theory that gives a unified account for a wide range of facts about interpretation, acceptability, and logical properties of natural language expressions in these three domains. The theory is couched in a mereological (part-whole) frame rather than the more usual set-theoretical base. The central concept that makes the unified theory possible is that of stratified reference. The basic property that is at the basis of the unification is boundedness, as exhibited in these examples from Champollion’s work, with the usual tags used for the contrasts in the literature:

(61) (a) John ran for five minutes. atelic
    (b) *John ran to the store for five minutes. *telic

(62) (a) thirty pounds of books plural
    (b) thirty liters of water mass
    (c) *thirty pounds of book *singular

(63) (a) The boys each walked. distributive
    (b) *The boys each met. *collective

Champollion constructs a general theory that posits two parameters that can be set to account for the parallelism and differences across the several domains: dimension and granularity. The empirical basis of the work draws primarily from the three construction types illustrated in examples (61)–(63) and from a host of other facts from English and other languages.

An important part of Champollion’s theory is that it allows for the possibility of parametrized and context-dependent aspects of meaning. For example the granularity option can reflect the difference between talk about an hour of dancing and centuries of political change.
4.6 Ontological Sources for Grammar

The rich ontological zoo, of which we have just caught some glimpses, plays a crucial role in offering options for what can be talked about in natural languages, but also provides a matrix for grammar. Many categories in syntax and morphology draw upon the distinctions we have mentioned but many others as well. We give some examples, framed in terms of the grammar, in each case taking up the question of how such grammatical categories might play a role in natural language metaphysics. (Information about the kinds of classifications that languages make in their grammars can be drawn from many handbooks and grammars for individual languages. We draw attention to Greville Corbett’s excellent surveys of several such domains, listed in our references.)

4.6.1 Gender / Noun Classes, Classifiers

Gender systems that are familiar from Indo-European and Semitic languages based on sex of an individual are but one of a number of such classifications: masculine and feminine; masculine, feminine, and neuter; common and neuter; animate and inanimate. These are some of the more familiar schemes. In all such systems that we are aware of there is a certain amount of slippage and arbitrariness. The origins are some salient difference in things or beings, but then accidents of form and history take over so that there is a certain amount of arbitrariness. This extends even to systems that seem to be semantic in nature. Whorf [1945] records such “covert” categories for English nouns: for example, she referring to boats of a requisite importance or size. The noun classes of Bantu languages fall under this type as well, again with a very loose correlation with semantic categories. These nominal classifications make themselves felt primarily in agreement systems as expressed on: determiners, adjectives, subject and object marking on verbs.

Related to the category of gender, classifier systems prototypically have to do with counting, and we can see a faint bit of it in English: five head of cattle, six pieces of furniture and related measure phrases. Again the characteristics for the classifications have some basis in reality but with arbitrary or capricious features as well, sometimes the result of historical accidents, as when Japanese hon ‘book, volume’ is used as a classifier for long cylindrical objects like bottles.

Thai has one of the most elaborate systems of classifiers, with several hundred. One classifier (t’uā or dtua) is used for larger animals, furniture like tables and chairs, suits of men’s clothing — so far, things with arms and legs — but also germs and fish!

4.6.2 Grammatical Number

Familiar number categories in grammar are singular and plural, but systems with dual are not uncommon. Number is sometimes obligatory and general for all nominals (and agreeing determiners, verbs, adjectives) as in English, but a widespread
trait is for number to be restricted to human or perhaps animate things, and optional. Many languages do not distinguish singular and plural nominals.

Number enters into lexical distinctions as well. It is fairly common for languages to use completely different roots for verbs with singular and plural subjects or objects. Coast Tsimshian, for example, expresses ‘run’ with \textit{baa} for singular, but \textit{k’ol} for plural subjects.

We close this section, which could go on to many other areas where languages draw on semantic domains to people their grammars, reversing Quine’s well known quote: philology recapitulates ontology. ([Quine, 1960: viii] “Ontology recapitulates philology” playing on Haeckel’s “Ontogeny recapitulates phylogeny”.)

5 FROM NATURAL LANGUAGE METAPHYSICS TO REAL METAPHYSICS

A central semantic problem in this probing of metaphysics from natural language metaphysics is the issue of possible individuals. If there are individuals that are only possible but not actual, the concept of \textit{domain} of individuals used to define a model will need to contain them but this is an issue on which it would be unethical for us as logician or linguist (or grammarian or semanticist, for that matter) to take a stand [Montague, PTQ, footnote 8].

We have drawn a distinction between what kinds of things natural language or natural languages seem to need for semantic theories and “real” metaphysics. Nicholas Asher [1993] accepts this distinction and using Kamp’s \textit{Discourse Representation Theory} locates the step to real metaphysics at the juncture where \textit{discourse representations} are embedded into a model, the point where truth enters into the interpretation [Kamp, 1981; Kamp and Reyle, 1993].

Some philosophers will say that what we are doing just \textit{is} metaphysics. Other philosophers are interested in revisionist accounts, that is, ridding our language of various erroneous or extravagant machinery (for example, doubting whether we should countenance intensional entities like groups). That is not our aim. We believe that this kind of endeavour can be carried out under the usual strategies of inquiry followed in other linguistic domains, syntax and phonology, for example. A recurrent pair of questions is then: how much of what we are claiming about the things and distinctions we posit are universal, common to all languages, and what are the limits and possibilities of variation across languages.

Whatever the answers to such questions at this level of granularity, we hope to have shown how an articulated theory allows various options besides the strictly denotational account for coping with differences of interpretation within languages and language.
The Metaphysics of Natural Language(s)

BIBLIOGRAPHY


MEANING AND USE

Robert van Rooij

1 INTRODUCTION

This paper deals with the meaning of natural language expressions, and how meanings of expressions are used in communication. The two disciplines that talk most about meanings of expressions are linguistics (semantics and pragmatics) and philosophy. This paper is about topics discussed in both disciplines. The first part of the paper is more philosophical in nature and discusses what is meaning in the first place, and how it is related with reference. The second part is concerned with the relation between semantics and pragmatics.

Although certainly not uncontroversial, there can be little doubt that what is known as ‘formal semantics’ is the most productive brand of natural language semantics within linguistics (cf. the popular introductions, [Chierchia & McConnell-Ginet, 1990; Heim & Kratzer, 1998]). In formal semantics, the meaning of a sentence is its truth conditions. Once we adopt a truth conditional concept of meaning, it is natural to think of reference as depending on meaning and that semantics is consistent with the Chomskyan cognitive, and individualistic, program in linguistics. In the first part of this paper we will first discuss two well-known problems for a particular way to combine these ideas: ‘Putnam’s paradox’ and Kripke’s counterexamples. Next, we discuss how a causal theory of reference can overcome these problems when framed within a two-dimensional conception of meaning, and what this means for the interpretation of the latter framework.

The second part of the paper investigates the relation between semantics and pragmatics. First it discusses what is communicated with the use of a sentence on top of its semantic meaning: Gricean conversational implicatures. The discussion will be limited to implicatures generated by Grice’s maxim of Quality and his first submaxim of Quantity. Speech acts will be discussed afterwards. First assertions, and the idea to think of a presupposition of a sentence as a felicity condition for the appropriate use of the sentence. Then questions, focussing on the issue whether Searle [1969] was right in claiming that the meaning of the interrogative sentence ‘Is the door open?’ is the same as that of its declarative analogue ‘The door is open’, the difference being just the way (speech act) in which the sentence is used. Finally we will deal with imperatives and permissions. In this part it will be discussed whether a performative or an assertive analysis of disjunctive permission sentences is most suitable to account for their free choice inferences.
2 MEANING AND REFERENCE

2.1 Meaning determines reference

The perhaps most ‘natural’ conception of ‘meaning’, at least in its point of departure, identifies ‘meaning’ with naming. The meaning of an expression is that what the expression refers to, or is about. What meaning does is to establish a correspondence between expressions in a language and things in the (model of the) world. For simple expressions, this view of meaning is natural and simple. The meaning of a proper name like ‘John’ or definite description like ‘the number of major planets’, for instance, is the object or number denoted by it, while the meaning of a simple declarative sentence like ‘John came’ could then be the fact that John came. Beyond this point of departure, things are perhaps less natural. What, for example, should be the things out in the world that common nouns and a negated sentence like ‘John didn’t come’ are about? One can, of course, assume that they refer to real existing properties and negative facts, but these assumptions make our initial hypothesis immediately less appealing. Except for conceptual worries, this referential theory of meaning gives rise to a serious empirical difficulty as well: the substitution problem. Assuming, by the principle of compositionality, that the meaning of a complex sentence depends only on the meanings of its parts and the way these parts are put together, it follows that if two expressions have the same meaning, one can substitute the one expression for the other in a complex sentence without change of meaning. But because there are 9 major planets in our solar system, on the theory of meaning at hand the expressions ‘9’ and ‘the number of major planets’ refer to the same thing, and thus have the same meaning. Still, we cannot substitute the expressions ‘the number of major planets’ for the number 9 in the sentence ‘It is necessary that 9 is bigger than 7’ without changing its truth value. The natural conclusion is that the meaning of an expression like ‘the number of planets’ should not be identified with its referent.

It seems natural to assume that to be a competent speaker of English one has to know what it means for ‘John came’ to be true or false. So a minimal requirement for any theory of meaning seems to be that one knows the meaning of a declarative sentence if one knows under which circumstances it is, or would be, true. The proposal of formal semanticists to solve our above conceptual problems is to stick to this minimal requirement: identify the meaning of a declarative sentence with the conditions, or circumstances under which the sentence is true. These circumstances can, in turn, be thought of as the ways the world might have been, first order models, or possible worlds. Thus, the meaning of a sentence can be thought of as the set of models, or possible worlds, in which it is true. This latter set is known in possible worlds semantics as the proposition expressed by the sentence.

Possible world semantics can account for the substitution puzzle when it assumes that the different expressions have different referents in at least some possible worlds. The most natural way in which this can be accounted for is to follow
Frege [1898] and make a distinction between the meaning and reference, or denotation, of an expression. In possible world semantics this distinction can be modeled by assuming that the denotation of an expression in a possible world is simply an object, and that its meaning is a non-constant function from possible worlds to its denotation in that world. This is natural for definite descriptions like ‘the number of major planets’, but what about other types of expressions for which such substitution puzzles can arise, like proper names (‘Hesperus is Phosphorus’), and common nouns (‘Water is $H_2O$’)?

Well, it seems that with the use of a proper name, or common noun, we associate a cluster of predicates or properties. The natural suggestion then is to define the meanings of these expressions in terms of these predicates or properties: they (or a sufficient number of them) give the list of necessary and sufficient conditions that an object, or stuff, has to satisfy in order to be denoted by the proper name or common noun.\(^1\) Obviously, the above substitution puzzles for proper names and common nouns do not arise on such a view. For instance, although the names ‘Hesperus’ and ‘Phosphorus’, or ‘Cicero’ and ‘Tully’, actually refer to the same individual, one can propose that they don’t have the same meaning, and thus cannot always be substituted for each other in a sentence without change in meaning.

The truth-conditional tradition in semantics has its source in the work of logicians and philosophers like Frege who held a rather anti-psychologistic view towards meanings. However, once we adopt the above possible world semantics and think primarily of truth-conditions rather than of truth, and of reference as depending on meaning, we can think of semantics as being consistent with the Chomskyan cognitive, and individualistic, program in linguistics. Specifying the meaning of an expression should be consistent with knowing the meaning of that expression, and this, of course, is true if one takes the meaning of a proper name, or common noun, to be the set of properties associated with that expression. Moreover, the truth-conditional view on sentence meaning seems natural as well, because from a cognitive perspective this means that knowing the meaning of a sentence is to know under which circumstances it is true. In this way model theoretic, or possible world, semantics can account for the primary task of natural language semantics, at least when seen from a Chomskyan perspective: to account for pretheoretical judgements of speakers concerning semantic relatedness of expressions of a particular language, in particular of the relation of entailment.

More in general, specifying meanings as in standard possible world semantics is consistent with the computational model of the mind which has become fundamental for cognitive science. The view that the meaning of an expression, or of an internal state, determines, but is independent of, what that expression or

---

\(^1\)This description theory of reference is only one example of a cluster theory of meaning. Many scholars, ranging from philosophers like Wittgenstein [1953] and Searle [1958], linguists like Katz & Postal [1964], and psychologists like Rosch [1978], have proposed that it are clusters of characteristic properties that (help to) identify the denotation of an expression in a particular context of use, or what a thought is about.
state is about, is compatible with the computational model of the mind which sees interpretation and the explanation of behavior as involving only internal states, or of internal models of external states of affairs. It is true that according to some cognitive scientists meanings just are internal states, and why should model theory be of relevance here? But, then, the computational model of the mind favors a functional view of internal states: it is not an internal state all by itself that has a meaning, but rather the (abstract) function that state has to explain an individual’s overall behavior. Model theory is well suited to account for such abstract functions.

2.2 Problems for the standard conception of meaning

However appealing and natural this combination of possible world semantics and the cluster theory of reference might be, it gives rise to at least two problems, one conceptual and one empirical in nature. In the following subsection I will discuss both problems, and some suggested solutions to them. As we will see, both problems are, in fact, independent of possible world semantics as such, and concern only the cluster theory of reference.\footnote{This is not to say that possible world semantics by itself doesn’t give rise to conceptual and/or empirical problems. It certainly does, but I will ignore those problems in this paper.}

2.2.1 Intended interpretation and meaning holism

A first problem concerns the predicates, or properties, used in the description that is supposed to identify the referent. According to the cluster theory of reference, a speaker refers with ‘N’ to a because a is the unique individual or stuff that satisfies the set of predicates that the agent associates with ‘N’. We can think of this set of predicates as the speaker’s representation of a. So, this analysis explains the speaker’s reference of a by ‘N’ in terms of the reference of the predicates associated with ‘N’. But that only gives rise to the questions what those predicates themselves refer to, and why they do so. The standard cluster theory of meaning doesn’t seem to do more than explaining one part of the language in terms of other parts — the terms in which the descriptions are given. Obviously, our problem of why one type of expression refers to what they do is not really solved, but only replaced by the same problem for another type of expression. But perhaps we should not think of these other type of expressions as belonging to the same external language; perhaps they are expressions of an internal language, or some other kind of internal representations of an agent. It all doesn’t matter much: in whatever way we represent the speaker’s meaning of ‘N’ in terms of a set of internal representations of general terms, it always gives rise to the further question of why these internal representations of general terms are about what they are in fact about.

One way to get out of our above regress problem is to propose that we can’t interpret the terms of a language individually, but that we have to do so simultaneously for the language as a whole. The idea would be that the terms refer to
Meaning and Use

whatever things, properties, and relations that do the best job of making the set of sentences true that speakers in fact consider to be true. Unfortunately, Putnam [1981] has showed that this picture as such is not constrained enough to fix the meaning of the expressions of a language in the intuitively correct way. He elaborates on the model theoretic fact that for any consistent set of sentences a model can be constructed with a domain of individuals that is not of the wrong size: we can always come up with different sets of objects or different interpretation functions that make the same set of sentences true. From this fact, Quine [1960] concluded to the indeterminacy of reference: knowing the truth value of a collection of sentences doesn’t mean that you know the references of its constituents. Putnam [1981] generalized this argument to intensional languages: even if one knows the truth value of a sentence in every possible world (its intension), this doesn’t necessarily mean that one knows the intuitively correct intensions of its constituents. For instance, it is possible to formulate highly counterintuitive intensions for expressions like cat and mat, so that in the actual world they refer to trees and cherries, respectively, without affecting the intension of The cat is on the mat. To determine the meaning of the terms of our language, knowing the truth value or intension of a collection of sentences is not enough, because the terms of the language can be assigned weird and ‘unintended’ interpretations.

Though Lakoff [1987] has argued otherwise, this argument obviously does not rule out model theoretic linguistic semantics as such. In practice, it doesn’t even seem be of any importance for model theoretic linguistic semantics at all: it typically doesn’t care much about the meaning of basic terms, and certainly not about why these expressions have the meaning they have. Natural language semanticists that use model theory to account for the meaning of expressions are interested only in how the meanings of more complex expressions can be explained in terms of the meanings of simple terms and some expressions with a logically fixed meaning. Still, if we believe that the world is at it is independently of our conceptions of it, Putnam’s argument gives rise to two general questions: first, why are only some of all possible interpretation functions compatible with the way English is spoken; second, also why are only some of all possible interpretation functions compatible with the way basic common nouns are used in any possible natural language? The first of these questions asks why English expressions have the meanings they actually have, and how the meaning of a proper name and common noun is determined in the first place, while the second one asks for more general constraints on how meanings could be assigned to expressions. Both of these problems can be solved if we are able to supplement model theoretic semantics with natural constraints on reference. But where could those constraints come from?

Perhaps we should not limit ourselves to behavior that involves verification or falsification of sentences, but should consider behavior in general, and how this is related with the beliefs and desires of the agents. We might then propose that meanings are assigned primarily to attitudes of agents, and that such an attitude is about, or directed to, an object, stuff, or state of affairs because the agent is disposed to perform actions that involve this object, stuff, or state of affairs.
Unfortunately, assignment of beliefs and desires such that it fits, or explains, the behavior of agents won’t be enough:

What makes an assignment of a system of belief and desire to a subject correct cannot just be that his behaviour and behavioural dispositions fit it by serving the assigned desire according to the assigned beliefs. The problem is that fit is too easy. The same behaviour that fits a decent, reasonable system of belief and desire also will serve countless very peculiar systems. Start with a reasonable system, the one that is in fact correct; twist the system of belief so that the subject’s alleged class of doxastic alternatives is some gruesome gerrymander; twist the system of desire in a countervailing way; and the subject’s behaviour will fit the perverse and incorrect assignment exactly as well as it fits the reasonable and correct one. Thus constitutive principles of fit which impute a measure of instrumental rationality leave the content of belief radically underdetermined. [Lewis, 1986, p. 38]

At this point Quine’s or Davidson’s principle of charity, or of humanity seems a natural extra constraint. This principle demands that we should not attribute too much irrationality to a person in order to explain his behavior. Lewis [1984] argues that making use of such a principle involves making additional constraints on what the meanings and/or references of expressions and internal states could be. He proposes³ that the intended interpretation function of our language, or of any natural language, is not as free as Putnam presupposes, because the meaning, or intension, of simple lexicalized predicates like ‘cat’ and ‘mat’ must refer to ‘well-behaving’ or ‘natural’ properties, and he proposes some constraints (mostly involving a notion of similarity) on what such natural properties and relations could be. Lewis suggests that when we limit ourselves to interpretation functions that map the simple predicates we use to ‘natural’ properties, there is no longer any guarantee that (almost) any world, or model of it, can satisfy (almost) any collection of sentences, and thus meaning indeterminacy might be tackled. Although I believe that Lewis’ proposal makes sense (if interpreted without his realist’s baggage) to answer the second of our above problems, I cannot see how it could account for the intended interpretation of English. Lewis’ proposal still seems to leave open too many interpretation functions.

Before we will discuss another possible way to solve Putnam’s paradox, let us first discuss a second problem for the cluster theory of reference.

2.2.2 Empirical problems

The second problem for the cluster theory of reference is empirical in nature. Donnellan [1970] and Kripke [1972] have convincingly argued that this theory of reference leads to counterintuitive results for proper names. More in particular,

³See also [Gärdenfors, 2000] for a similar proposal to solve Putnam’s paradox.
they have shown that speakers can refer, and even can intend to refer, to particular individuals without being able to describe or identify those individuals. First, speakers can successfully refer to a particular individual without having a uniquely identifying set of descriptions in mind. Second, even if they have such a description in mind, they sometimes still refer to an individual that doesn’t satisfy this description.

By very much the same kind of arguments, Kripke [1972] and Putnam [1975] have convincingly argued that the set of properties that speakers or agents associate with natural kind terms should also not be equated with the meaning of the noun. This is made very clear by the ‘Twin Earth’ stories given by Putnam [1975] and others. These stories always involve a comparison between two almost identical persons (twins): one in the actual world and one in a counterfactual world, Twin Earth, minimally different from the actual world. In Putnam’s story, the stuff that the inhabitants of the counterfactual situation call water is superficially the same as the stuff we call water, but its chemical structure is not $H_2O$, but $XYZ$. If, then, both the earthling and his twin assert ‘Water is the best drink for quenching thirst’, intuitively they have said something different. But how can this be if they associate exactly the same description with the word and if speaker’s description determines reference? A similar ‘Twin Earth’ story invented by Burge [1979] shows that the problem is not limited to a small set of terms. In fact, stories can be invented for almost any expression to show that it is not the description that the speaker associates with an expression that determines its extension. The reason is that the linguistic practices of members of the agent’s community are crucial in determining the extension of a term.

Perhaps what counts, then, is not so much the descriptions the speaker, or the relevant agent, associates with it, but rather the set of descriptions that most people, or the specialists in the relevant linguistic community associate with it. It is then this set of descriptions that determines the reference. However, Donnellan and Kripke have argued that this, too, give rise to counterintuitive predictions for proper names, while Putnam [1975] shows the same for natural kind terms. Putnam’s demonstration involves the same ‘Twin Earth’ story, but now set in 1750. Specialists on Earth and Twin Earth are not yet able to see any difference between $H_2O$ and $XYZ$. But intuitively, even if a typical Twin-Earthian (twin-) English speaker utters ‘Water is the best drink for quenching thirst’ on Earth, he is not talking about $H_2O$.

2.3 The causal theory of reference

Kripke and Putnam claim that the meaning of at least proper names and natural kind terms is not the set of descriptions associated with them, but simply what they refer to. But this gives rise to the question of why these expressions have the references they in fact have. At this point, Kripke proposed his causal theory of reference. Kripke [1972] argues that proper name ‘N’ can refer to a, only if, and because, a is the entity that is the source of the reference-preserving link
from the initial baptism of the expression to the speaker’s use of the name. Evans [1973] was perhaps the first to propose that the causal theory of reference should be based on a causal theory of belief, or of information. He argued with Kripke that a causal link for proper names is necessary, but that this causal link should not be between the initial naming and the speaker’s current use of the name, but rather between the body of information, or superficial properties, relevant to the speaker’s use of the proper name on a particular occasion and the object that is the dominant causal origin or source of this body of information. An object can be the dominant source of a particular body of information even if it does not fit this information very well. It follows that if \( P \) is one of the properties we associate with ‘N’, we still do not know that the sentence ‘N is P’ is true by necessity. This causal theory of aboutness can also explain why Oscar, but not his twin, talks or has beliefs about \( H_2O \) if he uses, or considers, the term ‘water’ in Putnam’s [1975] Twin Earth story.

The causal theory of reference, or of meaning, seems also the natural candidate to limit the possible interpretations of the expressions of ‘our’ language, or of our thoughts, so as to solve Putnam’s paradox.\(^4\)

The causal account of meaning is not without problems. It is not clear how to cash out the causal account in a completely naturalistic way and there are problems of how to account for our intuitions that we can have false beliefs.\(^5\) Moreover, it is unclear how a causal theory could ever determine the meaning of functional words, or of prepositions like ‘in’. But it seems that the causal account of content leads to unsolvable problems even if the above problems can be accounted for. Once we accept that the content of an intentional state or expression is just the causal source of the state, or of the use of the expression, we are confronted again with many old problems. If the meanings of ‘Hesperus’ and ‘Phosphorus’ are just their referents, the substitution puzzle arises again: ‘Hesperus is Phosphorus’ is predicted to express the necessary true proposition. But, then, how can we account for the fact that agents seriously doubt that such statements are true? Thus, the causal theory seems to predict a notion of content that is sometimes not fine-grained enough to account for our intuitions. At other times, however, the causal account of content seems to predict a notion of content that is too fine-grained, or too specific. For instance, it seems to predict that attitude ascriptions can no longer do the job commonsense psychology tells us they do. A common sense explanation of why the Earthling and his counterpart drink so much of the stuff that in their

\(^4\)Putnam [1981] claims that making use of this causal story is just adding more statements to our consistent set of sentences. But then it doesn’t solve the problem because the predicates of these additional sentences might be interpreted in unintended ways as well. But with Lewis [1984] I think that this is only the case if one thinks of the causal theory in a ‘descriptive’ way, as a set of sentences that has to be made true by the interpretation function. But the causal story is not intended to be incorporated within the semantic content of what is said with the name, it rather determines the content itself from a more external point of view.

\(^5\)One way to solve both of these problems involves making use of so-called ‘normality conditions’. But in order for the resulting analysis to be wholly naturalistic, we need a naturalistic analysis of such conditions. A natural candidate that suggests itself to provide such an analysis is Millikan’s [1984] biosemantics. I am not sure, though, whether this theory can do the full job.
respective communities is called ‘water’ if they are thirsty is that they think that what they call ‘water’ is the best drink for quenching thirst. The problem is that according to the causal conception of content it seems that the belief attribution ‘Oscar believes that water is the best drink for quenching thirst’ is more specific than we want, because we know that Oscar cannot distinguish $H_2O$ from $XYZ$. This problem is also of relevance to linguistics, because on the causal story Oscar doesn’t even know what he himself is talking about when he is using the term ‘water’. This seems to be incompatible with Chomskyan linguistics.

2.4 Two-dimensional semantics

It is an obvious observation that what is expressed by a sentence is context-dependent: in different contexts the same sentence can express different propositions. For instance, the proposition expressed by ‘I am living in Amsterdam’ depends on who is the speaker in that context. In Kaplan’s [1989] theory of context dependence, contexts consists of certain aspects of a world, like speaker, hearer, time, etc, and sometimes also the world itself. A context partially determines what is said by a sentence, and this is still modeled by a set of possible worlds.

Kaplan’s theory of context dependence can explain why there are two ways people can disagree about the truth value of a statement. Suppose that the speaker claims something by uttering a sentence, and the hearer disagrees. They can disagree because the hearer has misunderstood the speaker. The hearer has made a wrong guess about the context of utterance the speaker was in, and thus about the context-dependent proposition expressed by the speaker. It is also possible that they agree about what is said, but disagree about the facts that determine the truth value of what is said.

If both context and possible world are relevant for determining the truth value of a sentence, we might say that the meaning of a sentence is a relation between them, a two-dimensional intension. Following Kaplan, we can call this kind of meaning the character of a sentence. The character of a sentence is compositionally determined by the characters of its parts. If $E$ is an expression, we might call $[E]$ the character of $E$. Given a context, $c$, $[E](c)$ is the content or intension of $E$. $[E](c)(w)$, finally, is the extension of $E$, if $w$ is a possible world. The content of a sentence is a proposition, and its extension a truth value.

Kaplan’s theory of context dependence allows us to distinguish different reasons why a sentence is ‘necessary’ true. First, what a sentence expresses in context $c$ can be true in every relevant world, $[A](c) = K$, where $K$ is the set of all relevant worlds. Sentences like ‘Hesperus is Phosphorus’ and ‘I am John’ used by John are necessary in this way, because the contents, or intentions, of proper names and indexicals are constant functions. But it might also be the case that a sentence is true in every context in which it is expressed. If $w(c)$ gives us the world of $c$, this means that for all $c : w(c) \in [A](c)$ holds. For instance, an English sentence like ‘I am here now’ is necessary true for this reason. We can think of the set of contexts
in which a sentence is true as a semantic object as well, and we might call it the diagonal.⁶ Important about this diagonal is that if a sentence contains a context dependent expression, it might be that the sentence expresses a necessary truth, although its diagonal doesn’t contain of all contexts.

Consider John’s uttering of ‘I am John’, for instance. We have seen that this sentence is necessarily true — i.e., its content is the set of all worlds — because both noun phrases refer to the same individual. Still, the sentence can, intuitively, be informative, because the hearer might be ignorant of the identity of the speaker, or at least doesn’t know that he is called ‘John’. This intuition can be accounted for within two-dimensional semantics by making use of the diagonal: the diagonal consists of some, but not all, contexts, because the hearer is unsure whether the actual context is one where the speaker is called ‘John’.

The examples discussed so far are rather straightforward, and involve all obviously context dependent expressions. But the two-dimensional analysis has been used to account for the other problems as well: it has been used to account for the fact that people can doubt whether the identity statement ‘Hesperus is Phosphorus’ is really true, and to explain why the belief attribution ‘Oscar believes that water is the best drink for quenching thirst’ is intuitively true, although Oscar cannot distinguish ‘real’ water, i.e, $H_2O$, from $XYZ$. The reason why this can be done is that not only the reference of expressions like ‘I’ and ‘you’ depend on contingent features of the context, but this is also true — at least according to the causal theory of reference — for proper names, natural kind terms, and, if we may believe Burge, in fact, for any other type of expression. But how would this go? Can we assume that the reference of a proper name is world-dependent, but not just because of the fact that objects could have been called differently?

The causal theory predicts that, in a sense, statements like ‘Hesperus is Phosphorus’ indeed only say something about the semantic rules of English. Still, it predicts that we can learn something non-linguistic if we are informed that Hesperus is Phosphorus. This is the case because even if the exact referent of an expression used in a conversation is not clear, we normally do have a pretty good idea about what properties the referents of terms being used have. Thus, if we receive the information that the sentence ‘Hesperus is Phosphorus’ is true, we learn not only some facts about the semantics of English, but also some astronomical facts. We learn that the most salient heavenly body seen in the morning sky is identical with the most salient heavenly body seen in the evening sky, because we already believe and presuppose that we are in a world in which the referents of the relevant expressions have those properties. The same is true for a belief attribution like ‘The Babylonians didn’t believe that Hesperus is Phosphorus’. This sentence can be used to attribute a belief about a (partly) astronomical fact to the Babylonians, because we know all too well what information the Babylonians associated with the expressions.⁷

---

⁶After Stalnaker [1978].
⁷The two-dimensional framework has also been used to explain what Oscar and his twin have in common when they say to themselves ‘Water is the best drink for quenching thirst’. 
Notice, though, that there is a difference between the sense in which the reference of these expressions depends on context. The expression ‘I’ is context dependent, because *in English*, ‘I’ always refers to the speaker, and the same expression of English might be uttered by different speakers. The reference of ‘Phosphorus’ and ‘water’, on the other hand, are context dependent only because in different worlds they have a different meaning, or causal origin. But, of course, in that sense the meaning of ‘I’ is context dependent as well: in another world it might be that the meanings of the pronouns ‘I’ and ‘you’ are interchanged. If we assume that a language, or grammar, determines both what are expressions of a language and what these expressions mean, then it will follow that when the same expression has a different meaning (and not just reference) as it has in the actual world, the one who utters that expression in that other world uses a different language. But the same would be true when we consider proper names and natural kind terms: if we assume that the meanings (and not just references) of proper names and natural kind terms depend on their causal origin, we have to conclude that an expression used in a world with a different causal origin as in our world is part of a different language. Assuming that we speak a particular language, it follows that we sometimes don’t know the meanings of the expressions we use. Though this might feel to some like a contradiction in terms, others will take this conclusion to be equally innocent as a philosopher’s worry whether the tree in front of him is a ‘real’ tree, and not just a holographic image of it: only in case the language user or philosopher is contemplating such skeptical thoughts it has any practical consequence.

On the emerging picture, we cannot simply think of ‘Phosphorus’ as an expression of a language that might have different causal origins in different worlds. But why not?, you might wonder. Why not just think of the meaning of ‘Phosphorus’ as something like ‘whatever is the causal origin of our use of the expression ‘Phosphorus’”? One reason is that on this view the causal theory of reference cannot by itself solve Putnam’s paradox. We have explained the meaning of one expression in terms of the meaning of others and have thereby turned the causal theory of reference into a description theory that involves causal talk that itself might be interpreted in unintended ways. The basic point is that on this new analysis we still cannot put sufficient constraints on how to interpret expressions. Think of the analogy with indexical pronouns again. Even if we already know somehow that ‘I’ is an indexical pronoun, the meaning of ‘I’ would under a similar view not be much more then ‘depending on the actual convention of the language, ‘I’ refers to either the speaker or the hearer’. It is even worse for expressions of whom a speaker

---

8 See especially [Stalnaker, 1997] for a defense of this view.
9 See especially [Stalnaker, 2001] for this type of argument.
doesn’t know its type, or of an expression whose reference is not determined by its causal history, but — if we may believe Burge [1979] — still depends on external factors. For such type of expressions the meaning doesn’t seem to be any more specific than, ‘whatever this expressions means’, which basically comes down to the view that the meaning of an expression is nothing but the expression itself, or the internal representation associated with it.

2.5 Vagueness and context dependence

Truth conditional semantics assumes that the meaning of a sentence is given by its truth conditions. The phenomenon of vagueness is a potential threat to this framework. Consider the sentence John is tall uttered in a situation where John is 1.80 meters tall. Is this sentence true or false in this situation? This is hard to tell.

Vagueness is standardly defined as the possession of borderline cases. The borderline cases of tall are normally said to be those individuals of which we cannot really say whether they are tall or not: a man who is 1.80 meters in height is neither clearly tall nor clearly non-tall. In three-valued logics one can handle this phenomenon by saying that such a man is neither in the positive extension of tall, nor in its negative extension. These positive and negative extensions are given by a partial valuation function. If John is a man who is 1.80 meters high, John falls in the gap of the positive and negative extension of tall and the sentence John is tall is predicted to be neither true nor false. A well-known problem of this analysis is that too many sentences are predicted to be of this fate: Both John is tall or John is not tall and John is tall and John is not tall are predicted to be neither true nor false as well, although the former, and certainly the latter, intuitively have a classical truth value: true and false, respectively. To get rid of this problem, Fine [1975] and Kamp [1975] proposed to make use of, next to a partial valuation function, also a set of total valuation functions. A total valuation function doesn’t allow for gaps: each individual is either tall or not tall. Thus, a total valuation can make a partial valuation function more precise. If John is tall is neither true nor false according to the original partial valuation function, it will be either true or false according to each (accessible) total valuation function, because such total valuation functions have a specific cut-off point, or delineation, from whereon a man is counted as tall. However, different valuation function will have different such cut-off points. Crucial in supervaluation theory is the notion of supertruth: a sentence is supertrue iff it is true according to all (accessible) total valuation functions of the given partial valuation function. Similarly for the notion of superfalsity. Much of the appeal of supervaluations is that John is tall or John is not tall is predicted to be supertrue: although some total valuations count the former disjunct true and others the latter, they each make one of them true. Similarly for John is tall and John is not tall, which comes out as superfalsely. Supervaluation theory makes a difference between local and global notions of validity, defined in terms of the notions truth and supertruth.
Just as it is supertruth that behaves classically, so it is for the global notion of validity. $\phi$ superentails $\psi$ iff for all models $M$ and partial valuation functions $s$, if $\phi$ is supertrue in $M$ and $s$, $\psi$ should be supertrue in $M$ and $s$ as well.

Supervaluation theory assumes that the partial valuation function with which we start gives the semantics of English (in the actual world). Thus, it assumes that we should treat vagueness within semantics. But one might want to think of vagueness in a conceptually somewhat different way by just reinterpreting the partial and total valuation functions used in supervaluation theory: instead of saying that languages are vague, one can also say that it is our use, or knowledge of language that is imprecise. Lewis [1969], for instance, suggests that languages themselves are free of vagueness but that the linguistic conventions of a population, or the linguistic habits of a person, select not a point but a fuzzy region in the space of precise languages.\footnote{Lewis [1970] takes this analysis of vagueness to be very similar to (what is now called) a supervaluation account. Burns [1991] argues (unconvincingly, we think) that the two are very different.} A very similar view is taken in the epistemic approach of Williamson [1994]: English, or its valuation function, is precise, but agents don’t know exactly what this valuation function is. On such a view, borderline cases of tall are not individuals that are neither definitely tall nor definitely not tall in English, but rather individuals that some speakers of a language consistent with the linguistic convention of English consider to be tall, while others don’t (on Lewis’s meta-linguistic account), or individuals of which an agent doesn’t know whether they count as being tall or not, although the agent knows their precise length (on the epistemic account). Notice that both analyses of vagueness are very much in line with supervaluation theory: the partial valuation function from which the standard analysis starts still plays a role, if properly re-interpreted: it does not represent the semantics for English in the actual world, but rather what the population of speakers agrees on, or what an agent knows about, the actual interpretation function. Language users agree that the actual interpretation function is total, but disagree on, or are ignorant about, which one it is.

Notice that according to this re-interpretation, a total valuation can be thought of as a world that not only determines how the facts are (e.g. whether John’s height is 1.80 meters or 1.70 meters), but also how a vague predicate like tall should be interpreted: whether somebody who is 1.80 meters should be considered to be tall or not. Worlds fulfill two roles, and those roles are exactly the roles a world can play according to Stalnaker’s [1978] two-dimensional view on language discussed in section 2.4. If we fix the meanings of the expressions, a sentence expresses a (horizontal) proposition, represented by a set of worlds, and if the actual world is a member of this set, what is said by the sentence is true, false otherwise. But if what is expressed by a (token of a) sentence depends on context, we can think of the world (together with the expression token) as determining how the expressions should be interpreted, and then it might be that in different worlds
something different is said (i.e. different (horizontal) propositions are expressed) by the same sentential token. But if worlds fulfill the two roles suggested above, it seems natural to assume that they always fulfill the two roles at the same time. It follows that if we interpret a sentential token of a sentence $\phi$ in world $w$ of which we consider it possible that it is the actual world, we use $w$ both to determine what is said by $\phi$, (denoted by $[[\phi]]_w$), and to determining whether what is said by $\phi$ in $w$ is true in $w$, i.e., whether $w \in [[\phi]]_w$. The set of worlds denoted by $\{w \in W : w \in [[\phi]]_w\}$ is called the diagonal proposition by Stalnaker [1978]. The diagonal proposition expressed by a sentence is crucial to explain the so-called evaluative meaning of vague predicates. As noted by Barker [1992], if one says that ‘John is tall’, one can make two kinds of statements: a descriptive one saying that John is above the relevant cut-off point (if it is clear in a context what the cut-off point for being tall is) and a metalinguistic one saying that the cut-off point for being tall is below the length of John (if it is clear in a context what John’s height is). The latter involves the evaluative meaning of tall and can be accounted for straightforwardly in terms of diagonalization.

3 MEANING AND USE

3.1 Speech acts

In the sections until now we have concentrated on declarative sentences, and assumed and defended a truth conditional analysis of sentence meaning. This is in accordance with the assumption that the only, or at least primary, aim of language is to represent and communicate factual information, information that can be true or false. Around the middle of the last century, however, this assumption was seriously challenged by philosophers like Wittgenstein and Austin. Both stressed that making factual statements is only one thing we do with language, and that we should study our use of language from a more general perspective of human action and behavior. The moment we leave the realm of declarative sentences, this seems an obvious move. By using imperative and interrogative sentences like ‘Close the window!’ and ‘Is the window closed?’ we don’t describe an actual state of affairs, but rather make a command or ask a question in order to influence the behavior of our interlocutors. But, as pointed out by Austin, even for a whole class of declarative sentences the truth conditional analysis already seems unnatural. For instance, if John the judge says to the suspect ‘I sentence you to death’, to his brother ‘I bet that Ajax will win’, or to his wife ‘I promise to take care of our child’, his main purpose doesn’t seem to describe a state of affairs. What he seems to have done, rather, is to change the world: by his use of the sentence a sentence, a bet, or a promise came into existence that wasn’t really there before. Although Austin argued that for such so-called performative (uses of) sentences the question of truth or falsity doesn’t even arise, he noted that they can be used felicitously only if some appropriateness conditions are met. The utterance of ‘I sentence you to death’, for instance, only gives rise to a real sentence in case John makes it as a
judge, and in a country where the death penalty is in use. But, of course, not only do Austin’s performative sentences give rise to such appropriateness conditions for their successful use, imperative and interrogative sentences do so as well. The utterance of an imperative sentence like ‘Close the window!’, for instance, won’t be very successful in case the speaker has no authority over the hearer. Thus, the main point of all these sentences is to change, rather than to describe, the actual state of affairs, but this can be done successfully only if certain conditions are met. But as realized soon by Austin and others, this holds not only for imperative and interrogative sentences, together with the explicitly performative sentences of the type mentioned above, it is also true for standard declarative sentences like ‘The window is closed’. Although this sentence has truth conditions, a speaker uses it in an assertion only to make a point, and he can’t be successful in doing so when certain appropriateness conditions are not met. For instance, the speaker won’t make a point with his assertion in case it is already common knowledge among his conversational partners that the window is closed.

We have stated above that we make commands or ask questions in order to influence the behavior of our interlocutors. This seems to be true for making assertions as well, although perhaps more indirectly by influencing the hearer’s beliefs. So we might say that a command, question, or assertion is successful if and only if the hearer indeed performs the effect intended by the speaker. The intended effects for commands and questions would then be, most naturally, complying to the order and answering the question. But this intended effect depends very much on context. For instance, it might well be that one intended effect of my particular assertion using the sentence ‘It is cold’ is that you not only believe that it is, in fact, cold, but also that you close the window. In an influential article, Grice [1957] tried to determine, or define, what the speaker means by a sentence in terms of this intended effect in a hearer by means of the recognition of this intention. It is clear, however, that complying to the order, or answering the question, does not follow automatically when the hearer recognizes the speaker’s intention. If I tell you to close the window, you can refuse to do so, and still, intuitively, understand what I meant. The same is true for assertions: you can understand what I mean when I say ‘It is cold’ without closing the window, or even accepting that it is, in fact, cold. So, if we want to determine what the speaker means by a sentence in terms of the automatic effect in a hearer that follows from the recognition of this intention, we have to think of a specific kind of effect: what Austin and Searle call the illocutionary effect. What could such an illocutionary effect be? Well, if I tell you to close the window and you don’t, I still have communicated something when you recognized my intention, namely that I make it public, between you and me, that I want you to close the window. Similarly for my assertion that it is cold: even if you don’t close the window, or believe what I say, if you recognized my intention I still have made it public between you and me that I want you to believe (and so make it common ground) that it is cold. As the examples illustrate, commands and assertions have different illocutionary effects, the one involves what I want you to do, the other what I want you to believe, or become common ground. For
this reason they are called different illocutionary acts, or speech acts.

The traditional problem for speech act analyses was to find interesting types of speech acts, and to find necessary and sufficient conditions for the successful performance of the act. In these traditional analyses it was assumed that one could make a strict separation between what is expressed by a sentence, and the speech act performed by it. For instance, it was assumed that the sentences ‘Close the window!’, ‘Is the window closed?’, and ‘The window is closed’ all express the same proposition, namely that the window is closed, but that this proposition is used in different speech acts: a command, a question, and an assertion, respectively. Searle [1969] claims that assuming this hypothesis has many advantages. For instance, it allows us to make a distinction between propositional and illocutionary negation. When a negation is applied to a proposition, it just results in another proposition but leaves the character of the illocutionary act unchanged. When a negation is applied to the illocutionary act, on the other hand, the proposition remains the same, but the illocutionary act changes. A negation of an assertion, for instance, gives rise to a denial, and the negation of a promise gives rise to the refusal to make a promise.

More recently, speech act theorists concentrated themselves on the essential effects of speech acts, and these effects are analyzed in terms of how the speech act changes the conversational situation. In the remainder of this section I will discuss assertions, questions, and commands and permissions, but will highlight within those discussions three separate issues. In the discussion of assertions I will take issue on whether we really can separate what is expressed by a sentence, its content, from its illocutionary force. Here I will also discuss some appropriateness conditions for making a successful assertion. When talking about questions I will discuss whether the content of an interrogative sentence is really as close to the content of an assertive sentence as traditional speech act analysis suggests, and what it means for the standardly assumed autonomy of semantics with respect to pragmatics. When looking at commands and permissions, finally, I will discuss whether permission sentences should best be treated as assertions or as imperatives in order to account for their well-known free choice performative effects.

### 3.2 Presupposition as a felicity condition

On the assumption that the primary aim of language is to represent factual information, all that counts for the interpretation of a sentence is its truth value (in a world). We have assumed above that the sentence ‘John came’ is true (in a world) if the referent of ‘John’ actually came, and false if this referent did not come. But what if the name ‘John’ has no referent? One natural reaction is to widen the concept of falsity: the sentence is true if the referent of ‘John’ actually came, and false otherwise. Unfortunately, as already observed by Frege [1898], this gives rise to the counterintuitive prediction that the negation of the sentence would not be ‘John did not come’, but rather ‘John did not come, or the name ‘John’ (as used by the speaker) has no reference’. Strawson [1950] famously proposed to
solve the puzzle by claiming that if the referential expression has no reference, the sentence is neither true nor false. In order for ‘John came’, or ‘The king of France is bald’, to have a classical truth value (1 or 0), the referential terms occurring in it are required, or presupposed, to have a reference. In case of reference failure, the sentence in which the term occurs has a no classical truth value, though perhaps a non-classical one (i.e., *). After Strawson, linguists extended the notion of presupposition from referential terms to other types of expressions, including factual verbs like ‘regret’ and ‘know’, aspectual verbs like ‘stop’, and particles like ‘even’ and ‘too’. Sentences like ‘Mary knows that John came’ and ‘John came too’, for instance, are said to presuppose that John came, and that somebody different from John came, respectively. These sentences would neither be true nor false, in case their presuppositions are not met.

So far, we have looked at simple sentences, but what about complex ones? If the truth value of a complex sentence involving a truth conditional connective is determined from the truth values of its parts, the most natural way of dealing with truth value gaps is to claim that any complex sentence inherits the true value gap of any of its parts. It is trivial to extend a two-valued logic into a three valued one which would have this result. But thinking of presupposition failure as having a truth-value gap gives rise to the prediction that whenever a simple sentence gives rise to a presupposition, any complex expression in which this simple sentence occurs gives rise to this presupposition as well. This prediction, however, is not in accordance with our intuitions: ‘Mary came and John came too’ doesn’t give rise to a presupposition, and, indeed, the sentence seems false in case nobody different from John came. One way to solve this problem is to come up with a new three-valued logic, where conjunction, for instance, does not give rise to a symmetric truth table, but to an asymmetric one instead: ‘A and B’ can be said to be false, rather than neither true nor false, when ‘A’ is false and ‘B’ neither true nor false. Although such an account would be more in accordance with our intuitions, this way of solving the problem seems to be rather ad hoc. Furthermore, it seems rather dubious whether we should account for presuppositions solely in terms of a third truth value. For one thing, because some people have argued that some sentences (like ‘Even John came’, or ‘Mary does not regret that John came’) can be true although their presuppositions are not met. For another, because it is rather doubtful whether we have firm, and theory neutral, intuitions that tell us whether a sentence is neither true nor false in a particular circumstance at all.

Both of these problems can be met when we think of language, and of presupposition, from a more general perspective. Once we assume that the primary aim of using declarative sentences is to communicate, rather than just to represent, factual information in order to influence the beliefs or actions of one’s conversational partners, we can think of presupposition as just a special kind of felicity, or appropriateness, condition for the successful use of a speech act. In Stalnaker’s [1978] classical analysis, an assertion of a declarative sentence ‘φ’ is successful just in case it increments, or updates, what is presumed to be commonly believed with

12This is Peters’ [1977] truth table of conjunction.
the content of the assertion. Thus, an assertion of ‘\(\phi\)’ is made with respect to the context of what is taken to be commonly believed, represented by \(K\), and its pragmatic effect is that this context is updated from \(K\) to \(Upd(\phi, K)\). If we assume that the context can be represented by a set of possible worlds, this means that \(Upd(\phi, K) = K \cap [\phi]\), where \([\phi]\) is the proposition denoted by ‘\(\phi\)’. If the aim of an assertion is to update the context, this context has to satisfy certain conditions in order for the assertion to be successful. For instance, the context should not yet entail the proposition expressed by ‘\(\phi\)’, because then the assertion would not change the context, and thus have no pragmatic effect. But a further constraint now follows naturally. Notice that on a common sense of ‘presupposition’, what is presupposed by a speaker is just what he takes to be common ground between the participants of a conversation. On this view, it is primarily speakers that presuppose something. However, a sentence might presuppose something as well: we can say that sentence ‘\(\phi\)’ presupposes \(P\) just in case ‘\(\phi\)’ can be appropriately uttered by a speaker only if he presumes it to be common ground that \(P\) is the case. But this means that the assertion of ‘\(\phi\)’ puts a constraint on the contexts in which it can be used appropriately: it has to be a context \(K\) that already entails, or satisfies, \(P\).

This speech act analysis of presuppositions can, arguably, solve the problems discussed above for a truth-conditional analysis of presuppositions (cf. Stalnaker, 1974). First, it can account for the fact that a complex sentence like ‘Mary came and John came too’ doesn’t give rise to a presupposition, although the second conjunct does. The reason is that it seems reasonable to assume that the context of interpretation of the second conjunct is not the initial context, \(K\), but rather the initial context updated with the content of the first conjunct. Because in this updated context the presupposition of the second conjunct is satisfied, even if this is not the case in the initial context, the utterance of the conjunctive sentence doesn’t put any constraint on initial contexts, and thus doesn’t give rise to a presupposition. Second, the speech act analysis of presuppositions can, in principle, account for the intuition that a sentence can be true (or false) in the actual world, although in this world the presupposition is not met. The reason is that the speaker might presuppose something that is actually false, and so the actual world is not an element of the context.

But is the above picture not simply mistaken? Isn’t it obvious that a sentence can be used that, intuitively, gives rise to a presupposition, although the speaker does not take it to be common ground that this ‘linguistic’ presupposition is true before he made the assertion? Indeed, Mary can say that she regrets that John did not come to convey the new information that John did, in fact, not come. But perhaps not all information that is conveyed by a sentence should be accounted for by semantics. Perhaps what is called presuppositional inference is one such case. And this makes sense from the pragmatic point of view: In case it is commonly known that a sentence \(\phi\) can normally be asserted appropriately only in case certain information \(\psi\) is already taken for granted by the participants of the conversation, it becomes possible to exploit this knowledge by using \(\phi\) to pretend that \(\psi\) is already
assumed. A speaker can pretend to take something to be already common ground and thereby, indirectly, convey new information. For this to be possible, however, it is required that this pretense is the exception, rather than the rule. So we see that our appealing pragmatic picture of presuppositions can be appropriate to the extend that in most, or at least in typical, conversational situations in which the speaker uses a sentence that gives rise to a presupposition, this presupposition is already common ground.

3.3 Questions and the autonomy of semantics with respect to pragmatics

Just like an assertion, also a question is a speech act: it is something we do with a sentence. However, the sentences we typically use to ask a question differ from the sentences we typically use to make an assertion. While assertions correspond with declarative sentences, questions correspond with interrogative sentences. But this correspondence is not complete: declarative sentences, for instance, can be used not only to make assertions, but also to ask questions. This is typically the case for declarative sentences with rising intonation. Suppose that we can determine whether a sentence is used as a question or not. Then we need to know what is the meaning of the sentence, and what is the pragmatic effect.

So, what is the meaning of a question? It seems natural that a question like ‘Did John come?’ also involves a proposition, but how should this involvement be spelled out? In traditional speech act theory it was assumed that the question simply expresses the proposition that John came, i.e., that the meaning of a question is just a proposition, but that this proposition is just used differently than in an assertion. So how is this proposition used in a question, i.e., what is the pragmatic effect of ‘Did John come?’ What the speaker expresses with the sentence is that he wants to know what is the correct and satisfying true answer. But what is a satisfying answer to a question? For yes-no questions this seems obvious: ‘yes’ and ‘no’, or just the proposition expressed by a yes-no question and its negation. So what the speaker then wants to know is whether, of all possible worlds that he takes to be live options, the actual world is one where the proposition expressed by the question is true, or not. But this means that the essential effect of a yes-no question is to introduce to the context the issue of whether the proposition expressed by the sentence is true. Thus, what the question does is to divide the worlds of the context into those where the proposition holds and those where it does not, i.e., it partitions the context.

Let us say, following traditional speech act theory, that the meaning of the question ‘Did John come?’ is the set of possible worlds in which John came. Equivalently, this is just the following function from worlds to the value ‘true’ if John came in that world, and the value ‘false’ otherwise: \( \lambda w[\text{John came in } w] \). The pragmatic effect of the question with respect to context \( K \), \( Upd(\text{Did John come}, K) \), can now simply be modeled as the following partition of \( K \).
\[ \{ v \in K : \lambda w [ \text{John came in } w (v) = \lambda w [ \text{John came in } w (u)] ] \mid u \in K \} , \]

Because of the correspondence between partitions and equivalence relations, the pragmatic meaning of ‘Did John come?’ can also be given by the equivalence relation \( \lambda v \lambda w [ \text{John came in } w \text{ iff John came in } v ] \) on \( K \).

To account for \( wh \)-questions along the same lines, it seems we need propositional functions and not complete propositions, because in contrast to yes-no questions a speaker asking a \( wh \)-question does not express a complete proposition. So let us assume that the meaning of a question like ‘Who came?’ is the function, \( \lambda w \lambda x [ x \text{ came in } w ] \), that when applied to a world and an individual assigns the value ‘true’ if the individual came in that world, and the value ‘false’ otherwise. To determine the pragmatic effect in a similar way as for yes-no questions, we have to know what counts as a satisfying answer to the question. In principle, many answers can be given to the question, but some are more natural than others. But we don’t care about natural answers here, but only about the kind of answers that would fully satisfy the questioner. A fully satisfying answer seems to be one where he learns the \textit{complete} answer, and knows afterwards who came. Groenendijk & Stokhof [1982] have argued that to know who came, the agent needs to know of each single individual whether he or she came. On this proposal, the pragmatic effect of a question is to give the set of all possible complete answers. Notice that such complete answers exclude each other, and, given that for each world there is one complete answer true in it, the pragmatic effect of ‘Who came?’ with respect to context \( K \), \( Upd \text{(Who came?, } K) \), thus gives rise to a set of propositions which partitions context \( K \) as well, or, equivalently, to the equivalence relation on \( K \) below.

\[ \{ v \in K \mid \lambda w \lambda x [ x \text{ came in } w (v) = \lambda w \lambda x [ x \text{ came in } w (u)] ] \mid u \in K \} \]

\[ \lambda v \lambda w [ \lambda x [ x \text{ came in } w ] = \lambda x [ x \text{ came in } v ]] \]

Notice that on this analysis, the pragmatic effects of questions all give rise to partitions. This allows us to define an entailment relation between questions, in terms of their pragmatic meanings, or effects. Suppose that \( Q^K \) and \( Q'^K \) are the pragmatic effects of two interrogative sentences with respect to context \( K \). Then we might say that the first interrogative sentence \textit{pragmatically entails} the second interrogative sentence just in case for every proposition that is an element of the partition \( Q^K \) there is a proposition that is an element of the partition \( Q'^K \) such that the former is a subset of the latter. More formally, \( Q^K \) pragmatically entails \( Q'^K \) iff \( \forall q \in Q^K : \exists q' \in Q'^K : q \subseteq q' \).\textsuperscript{13} This abstract characterization seems to make sense as well. For instance, it predicts that the question ‘Did John come?’ is pragmatically entailed by ‘Who came?’, because any (complete) answer to the

\textsuperscript{13}One might also abstract away from the context, or from the model, but I will leave that to the reader.
latter question also completely answers the former question. This prediction seems to be in accordance with our intuitions.

So it seems that the traditional speech act analysis of questions is quite appealing. Still, semanticists typically don’t adopt this analysis. The reason for this is that if we assume that the meaning of a sentence should be determined in terms of the meanings of their parts, the analysis can’t account for *embedded* sentences like ‘Mary knows who came’. On the most natural analysis of sentences like ‘Mary knows that John came’, the verb ‘know’ denotes a relation between an agent and the proposition expressed by the embedded sentence ‘John came’. Unfortunately, on the present analysis, the semantic meaning of ‘who came’ is a propositional function, rather than a proposition. So, either we have to assume that the meaning of the verb *know* is ambiguous between the standard one that involves a proposition, and a question-meaning that involves a propositional function, or we have to give up the assumption that the meaning of ‘Mary knows who came’ can be compositionally determined in terms of the meanings of its parts. In fact, a simple ambiguity between a propositional and a question-meaning of ‘know’ is not going to be enough, for on the standard speech act analysis the meaning of a question cannot only be a propositional function, but also a propositional relation, or a proposition, as in ‘Mary knows whether John came’. But even if we assume that the meaning of the verb ‘know’ is multiply ambiguous, it is still not clear how it could account for the intuitively correct truth conditions of sentences like ‘Mary knows who came’ and ‘Mary knows whether John came’. According to almost everybody’s intuition, the latter sentence is true if and only if Mary knows *that* John came if John in fact came, and Mary knows *that* John did not come if John did in fact not come. But how can we derive these truth conditions if we assume that the meaning of the embedded question is just a propositional function, rather than a proposition? Similarly for embedded *wh*-questions: if the semantic meaning of such a question is just a propositional function, how can we account for the intuition we have discussed above that ‘Mary knows who came’ is true if and only if Mary knows for each individual whether that individual came?

Fortunately, there seems to be a very straightforward solution to this problem. Just assume that to account for the truth conditions that involve embedded questions we shouldn’t look at the *semantic* meanings of these embedded questions, but rather at their *pragmatic meanings*. Remember that according to the traditional speech act analysis one might think of the pragmatic meaning of a question as a partition on, or an equivalence relation between, possible worlds. If ‘whether John came’ and ‘who came’ have such equivalence relations as their pragmatic meanings, these relations between worlds give rise to propositions if they are applied to a world. Now assume that to determine the truth conditions, to check whether ‘Mary knows Q’ is true in a world $w$ — where $Q$ is the embedded question — we just look at the proposition determined by the pragmatic meaning of $Q$ applied to $w$. Now we can assume that the meaning of ‘know’ is just a relation between individuals and propositions, and we correctly predict that ‘Mary knows whether John came’ is true just in case Mary knows *that* John came if John in fact came,
and Mary knows that John did not come if John did in fact not come. One can check that on this procedure also ‘Mary knows who came’ receives the intuitively correct truth conditions.

There is, in fact, not so much to say against this analysis, except for the following methodological complaint: once we take embedded questions into account, we cannot determine the semantic meaning of the whole sentence solely in terms of the semantic meanings of their parts and the way they are formed together. Instead, we have to take the pragmatic meaning into account as well. This, obviously, contradicts Searle’s explicitly mentioned assumption that we can determine the proposition expressed by a sentence without the mentioning of illocutionary force, and more in general it would lead us to give up the assumption that semantics is autonomous with respect to pragmatics.

Suppose that we want to keep semantics autonomous with respect to pragmatics, how should we proceed? Given the above suggested analysis, this is easy to see, and it gives rise to Groenendijk & Stokhof’s [1982] analysis. Just as it is normally assumed that you know the meaning of a declarative sentence when you know under which circumstances this sentence is true, Hamblin [1958] argues that you know the meaning of a question when you know what counts as a satisfying answer to the question. If we say, as we did above, that only complete answers count as satisfying answers, this means that the semantic meaning of a question on this analysis is just the same as the pragmatic meaning of the same question on the suggested analysis above. It follows that the meaning of a question is a partition, or an equivalence relation, and that we can account for our intuitions concerning entailments between questions in terms of their semantic, rather than their pragmatic, meanings.

But how should we now account for the pragmatic effect of questions, and what do we do with embedded questions? As for the pragmatic effect things are easy: we can just say that the effect of a question when used in a context is that it (further) partitions this context. As for embedded questions, we now say that they have the same semantic meanings as their unembedded counterparts, and that Mary knows whether John came, for instance, is true if and only if Mary knows the proposition denoted by the semantic meaning of the embedded question applied to the actual world. This obviously gives rise to the same, and thus intuitively correct, semantic meaning of the whole sentence as on the analysis described above, except that now we don’t have to give up the assumption that semantics is autonomous with respect to pragmatics.

\[14\] In fact, the above sketched pragmatic analysis was, of course, modeled on Groenendijk & Stokhof’s semantic analysis. To be sure, there are other semantic analyses of questions, and they give rise to different — and some would say better — empirical predictions. But my main concern here is methodological rather than empirical, and so I won’t discuss those alternative semantic analyses here.
3.4 Permissions and the free choice effect

3.4.1 The problem of free choice permissions

According to Austin's classical analysis of speech acts, sentences of the form *You must/may do* \( \phi \) are not used to describe a states of affairs. In terms of the language game between master and slave as described by Lewis [1970/9], they are typically used by one person, the master, to command or permit another person, the slave, to do certain things.

How should we account for these so-called performative effects of the sentences used by the master? One proposal might be to say that command and permission sentences are assertorically used, but that the performative effect is accounted for in an indirect way, due to the fact that we learn, or realize, more about the world. One of the things one might learn about the world is what is demanded and permitted. A truth conditional analysis of what is demanded and permitted is given in deontic logic. Standard deontic logic (SDL) was based on the same principles as classical modal logic.\(^{15}\) Where normal modal logic has the operators \( \Box \) and \( \Diamond \) standing for necessity and possibility, SDL has the two operators \( O \) and \( P \), standing for *ought* or *obliged* and for *permission*, respectively. Model theoretically, we say that \( O(\phi) \) is true in \( w \) iff in all ideal worlds accessible from \( w \), \( \phi \) is true, and that its dual \( P(\phi) \) is true iff \( \phi \) is consistent with this set of all ideal worlds, i.e. if there is at least one ideal world accessible from \( w \) in which \( \phi \) is true. The set of ideal worlds in \( w \) will be denoted by \( P(w) \), and is known as the *permissibility set*. We might now propose that the performative effect of command and permission sentences is due to the fact that only after a command or permission sentence is used by the master, the slave *knows* that he is obliged/permitted to do something, by having eliminated worlds with inappropriate permission sets, and acts accordingly.

This assertoric analysis seems appropriate for some uses of command and permission sentences, but it has always taken to be problematic whether the performative effect of all permission sentences should be accounted for in the epistemic way sketched above. Consider the sentence ‘You may take the apple or take the pear’. According to standard deontic logic, this sentence follows from both ‘You may take the apple’ and from ‘You may take the pear’, and neither of them follows from the first disjunctive permission. In a sense this is how things should be because, as observed by Kamp [1979], there is nothing problematic with the assertion of ‘You may take the apple or the pear, but I don’t know which.’ On the other hand, however, we can intuitively infer both ‘You may take the apple’ and ‘You may take the pear’ from the disjunctive permission sentence. How could we possibly account for this latter free choice inference if the latter also can be inferred from the former two?\(^{16}\) In the following I will discuss two proposed solutions to

\(^15\)There are other truth conditional analyses of deontic concepts, of course, but we won’t go into that here.

\(^16\)Many authors have discussed this puzzle, and this typically involves dropping the standard truth conditional analysis of ‘or’, or the standard analysis of modals such that the disjunctive
this problem: (i) a performative analysis, and (ii) an analysis that explains the free choice inference as a conversational implicature.

3.4.2 The performative analysis of imperatives

The natural alternative to the assertoric analysis of obligation and permission sentences is the performative one involving a master and his slave. According to the performative analyses of Lewis [1970/9] and Kamp [1973], command and permission sentences are not primarily used to make true assertions about the world, but rather they are made by the master to change what the slave is obliged/ permitted to do.\(^{17}\) With some feeling for Amsterdam rhetorics, we might say that according to the performative analysis, we know the meaning of an imperative sentence, when we know how imperatives change permissibility sets.

According to this Lewis/Kamp account, if the master commands John to do \(\phi\) by saying ‘You must do \(\phi\)’, or allows John to do \(\phi\) by saying ‘You may do \(\phi\)’, it is typically not yet the case that the proposition expressed by \(\phi\) is respectively a superset of, or consistent with, John’s permissibility set, \(P\).\(^{18}\) However, the performative effect of the command/permission will be such that in the new context what is commanded is a superset of, and what is permitted is consistent with, the new permissibility set. Thus, in case the command or permission is not used vacuously, the permissibility set, \(P'\), of the new context will be different from \(P\), so that the obligation/permission sentence will be satisfied.

But if knowing the meaning of an imperative means that you have to know how the imperative changes the permissibility set, our problem is to say how command and permission sentences govern the change from the prior permissibility set, \(P\), to the posterior one, \(P'\).

For commands this problem seems to have an easy solution. If the command ‘You must do \(\phi\)’ is given by the master, the new, or posterior, set of permissible futures for John, \(P'\), is simply \(P \cap [\phi]\), where \([\phi]\) denotes the proposition expressed by \(\phi\).\(^{19}\) However, things are more complicated for permission sentences. It is clear that if \(\phi\) is allowed, \(P'\) should be a superset of \(P\) such that \(P' \cap [\phi] \neq \emptyset\). It is not clear, however, which \(\phi\)-worlds should be added to \(P\). Obviously, we cannot simply say that \(P' = P \cup [\phi]\). By that suggestion, giving permission to \(\phi\) would allow everything compatible with \(\phi\), which is certainly not what we want. But how then should the change from \(P\) to \(P'\) be determined if a permission is given? This is Lewis’s problem about permissions.

\(^{17}\)Although Lewis [1970/9] and Kamp [1973] account for the effect of permission sentences in rather different ways, both might be called performative analyses in the sense that their effect is to change the permissibility set.

\(^{18}\)From now on I will assume in most of this paper that there is only one (global) permissibility set around.

\(^{19}\)What if the new command is incompatible with one or more of the earlier ones? In that case we might make use of change by revision to be discussed below.
One possible way to solve Lewis’s problem about permissions is to assume that we not only have a set of best, or ideal, worlds, but also an ordering that says which non-ideal worlds are better than others. Thus, to account for the performative effects of commands and permissions, we need not only a set of ideal worlds, but rather a whole preference, or reprehensibility, ordering, \( \leq \), on the set of all possible worlds. On the interpretation that \( u \leq v \) iff \( v \) is at least as reprehensible as \( u \), it is natural to assume that this relation should be reflexive, transitive, and connected.\(^{20}\) In terms of this preference order on possible worlds we can determine the ideal set \( P \) as the set of minimal elements of the relation \( \leq \):

\[
P \overset{\text{def}}{=} \{ v \in W | \forall u : v \leq u \}
\]

In terms of this set of ideal worlds we can, as before, determine of course whether according to the present state \( \phi \) is obligatory or just permitted. For instance, \( \phi \) is obligatory iff \( P \subseteq [\phi] \).

But this ordering relation contains more information than just what the set \( P \) of ideal worlds is, and in terms of this extra information we can determine the new permissibility set \( P' \). If the master permits the slave to make \( \phi \) true, we can assume that \( P \) contains no \( \phi \)-worlds, i.e. none of the \( \phi \)-worlds is ideal. But some \( \phi \)-worlds are still better than other \( \phi \)-worlds. We can now propose that the effect of allowing \( \phi \) is that the best \( \phi \)-worlds are added to the old permissibility set to figure as the new permissibility set. The best \( \phi \)-worlds are the worlds ‘closest’ to the ‘ideal’ worlds \( P \) where \( \phi \) is true. This set will be denoted as \( P^*_{\phi} \) and defined in terms of the relation \( \leq \) as follows:

\[
P^*_{\phi} \overset{\text{def}}{=} \{ u \in [\phi] | \forall v \in [\phi] : u \leq v \}
\]

To implement this suggestion, we can say that the change induced by the permission \textit{You may do } \phi \textit{ is that the new permission set, } \( P' \), is just \( P \cup P^*_{\phi} \).\(^{21}\) Thus, according to this proposal, command and permission sentences change a context of interpretation as follows (where I assume that John is the relevant agent, and \( P \) his permission state):

\[
\begin{align*}
\text{Upd(Must}(\text{John}, \phi), P) &= P \cap [\phi] \quad \text{Upd(May}(\text{John}, \phi), P) = P \cup P^*_{\phi}
\end{align*}
\]

Note that according to our performative account it does not follow that for a permission sentence of the form ‘You may do \( \phi \) or \( \psi \)’ the slave can infer that according to the new permissibility set he is allowed to do any of the disjuncts. Still, the performative analysis can give an explanation why disjuncts are normally

\(^{20}\)A relation \( R \) is reflexive if for all \( w : R(w, w) \), it is transitive if for all \( w, v \) and \( u \): if \( R(w, v) \) and \( R(v, u) \), then \( R(w, u) \), and it is connected if for all \( w \) and \( v \), \( R(w, v) \) or \( R(v, w) \).

\(^{21}\)This analysis of permission sentences was assumed by Kamp [1979] in his discussion of the performative analysis of permissions.

\(^{22}\)This is, in fact, just \( P^*_{\phi} \), where it is assumed that \( \phi \) is compatible with \( P \).
interpreted in this ‘free-choice’ way. To explain this, let me first define a deontic preference relation between propositions in terms of our reprehensibility relation between worlds, \( \leq \). We can say that although both \( \phi \) and \( \psi \) are incompatible with the set of ideal worlds, \( \phi \) is still preferred to \( \psi \), \( \phi \leq \psi \), iff the best \( \phi \)-worlds are at least as close to the ideal worlds than the best \( \psi \)-worlds, \( \exists v \in [\phi] \) and \( \forall u \in [\psi] : v \leq u \). Then we can say that with respect to \( \leq \), \( \phi \) and \( \psi \) are equally reprehensible, \( \phi \approx \psi \), iff \( \phi \leq \psi \) and \( \psi \leq \phi \). Because, as it turns out, it will be the case that \( P^*_{\phi \lor \psi} = P^*_\phi \cup P^*_\psi \) iff \( \phi \approx \psi \), we can now explain why normally disjunction elimination is allowed for permission sentences.\(^{23}\) For simple disjunctive permission sentences like ‘You may do \( \phi \) or \( \psi \)’, it is not unreasonable to assume that when performatively used, the master has no strict preference for the one above the other. If we make the same assumption for command sentences, it follows that from ‘You may/must take the apple or the pear’ we can conclude that the speaker may take the apple and that he may take the pear.

This performative analysis gives rise to a number of problems, but for reasons of space I will consider only two. First, the analysis might be natural for permissions that are, intuitively, used performatively, but still can’t account for the intuition that the ‘free-choice’ effect results even if the sentence is used assertively. So, even if the analysis is correct, one still needs an analysis that makes the same predictions for assertively used permissions. But, then, perhaps, such an assertive analysis is all that one needs. The second problem is very similar: it seems that other sentences involving disjunction have ‘conjunctive’ readings as well, although for these examples it is neither performativity nor modality that seems to be crucial:\(^{24}\) from ‘Several of my cousins had cherries or strawberries’ we naturally infer that some of the cousins had cherries and some had strawberries. These problems suggest that we (at least also) need a general assertive analysis of free choice inferences. In the following section I suggest accounting for the inferences as conversational implicatures.

### 3.5 A pragmatic analysis of Free Choice

The fact that there is nothing wrong with the assertion of ‘You may take the apple or the pear, but I don’t know which’ suggests that the free choice permission inference is cancellable, and should be accounted for as a Gricean conversational implicature.

#### 3.5.1 Gricean implicatures

Traditionally, the semantic meaning of natural language expressions like ‘and’, ‘or’, ‘every’, ‘some’, ‘believe’, and ‘possibly’ has been analyzed in terms of their intuitive analogs in classical logic: \( \wedge \), \( \lor \), \( \forall \), \( \exists \), ‘\( \square \)’, and ‘\( \Diamond \)’, respectively. However, in

\(^{23}\)It is well possible to give a performative analysis of permission sentences where the free choice effect comes about without requiring that both disjuncts are equally reprehensible, but I won’t go into that here.

\(^{24}\)I learned these examples from Regine Eckhardt.
Meaning and Use

many contexts these expressions receive interpretations that are different from what is predicted by this approach to their semantics. In most circumstances, for instance, we infer from the assertion of ‘John is came or Mary came’ that John and Mary didn’t come together, and from ‘It is possible that John came’ that it is not necessary that John came.

How should these inferences be accounted for? Grice [1967] argued that the above inferences should not be accounted for within a semantic analysis, but should be accounted for in terms of general principles of rational communication. Grice assumes a theoretical distinction within the ‘total significance’ of a linguistic utterance between what the speaker explicitly said and what he has merely implicated. What has been said is supposed to be based purely on the conventional meaning of a sentence, and is the subject of compositional semantics. What is implicitly conveyed belongs to the realm of pragmatics and depends also on facts about the utterance situation, the linguistic context, and the goals and preferences of the interlocutors of the conversation. What is implicitly conveyed, or conversationally implicated can be determined, or so is proposed, on the basis of Grice’s cooperative principle: the assumption that speakers are maximally efficient rational cooperative language users. Grice comes up with a list of four rules of thumb – the maxims of quality, quantity, relevance, and manner — that specify what participants have to do in order to satisfy this principle. They should speak sincerely, relevantly, and clearly, and should provide sufficient information.

Over the years many phenomena have been explained in terms of the Gricean maxims of conversation. Horn [1972] and especially Gazdar [1979] proposed to formalize Grice’s suggestions in order to turn informal pragmatics into a predictive theory. They concentrated on Grice’s maxim of quality and his first submaxim of quantity. Grice’s maxim of quality says, roughly speaking, that the speaker always knows (or believes) what he says, while his first submaxim of quantity (and Relevance) assumes that the speaker makes his contribution as informative as required. Obviously, to implement these maxims, we need to take the knowledge state of speakers into account.

Formalizing that the speaker obeys quality is not that difficult: If our designated speaker utters \( \phi \), we simply assume that the speaker’s knowledge state entails \( \phi \), and thus that \( K\phi \) is true. Thinking of \( S \) as the set of knowledge states, quality demands that a speaker of ‘\( \phi \)’ is in one of the following knowledge state: \( \{ s \in S | s \subseteq [\phi] \} \). To account for the first subclause of the maxim of quantity that demands speakers to convey all (relevant) information they possess, we are going to select among those states where the speaker knows her utterance to be true the states where she has least additional relevant knowledge. This is formalized by defining an order on epistemic states and then select minimal elements of this order. The order compares the relevant knowledge of the speaker and we select minimal elements in the set \( \{ s \in S | s \subseteq [\phi] \} \). How much relevant knowledge a speaker has is taken to be represented by how many of a class of alternative sentences she knows to hold. Let us assume that if \( \phi = [\text{John}\text{smokes}] \), for instance, the set of alternatives contains sentences like ‘John smokes’, ‘Mary smokes’ and ‘Bill
smokes’ as well as the conjunctive and disjunctive combinations of them. Now we say that the speaker has less relevant knowledge in state $s$ than in $s'$, $s <_{\text{Alt}(\phi)}^{K} s'$, iff the set of alternative sentences known in the former state is a proper subset of the set of alternative sentences known in the latter state:

**Definition 1. (Ordering knowledge states)**

$$s <_{\text{Alt}(\phi)}^{K} s' \iff \{\psi \in \text{Alt}(\phi) : s \subseteq [\psi]\} \subseteq \{\psi \in \text{Alt}(\phi) : s' \subseteq [\psi]\}.$$  

Now we define the Gricean interpretation of $\phi$ as the set of minimal models where the speaker knows $\phi$ with respect to the set of alternatives $\text{Alt}(\phi)$.

**Definition 2. (A Gricean Interpretation)**

$$[\text{Grice}]^S(\phi, \text{Alt}(\phi)) = \{s \subseteq [\phi]^S : \forall s' \subseteq [\phi]^S : s <_{\text{Alt}(\phi)}^{K} s'\}.$$  

According to this interpretation function, if the speaker utters ‘[John]$_F$ came’ we conclude that the speaker knows that John came, but not that Mary came, and if she utters ‘[John or Mary]$_F$ came’ we conclude that the speaker does not know of anybody that he or she came. This is a nice result, but in many cases we conclude something stronger: in the first example that Mary, Bill, and all the other relevant individuals did not come, and the same for the second example, except that now this is not true anymore for Mary. How do we account for this extra inference in terms of our richer modal-logical setting?

In van Rooij & Schulz [2004] it is shown that this can be accounted for by assuming that speakers, in addition to obeying the Gricean maxims, are maximally competent (as far as this is consistent with obeying these maxims). This can be described by selecting among the elements of $[\text{Grice}]^S(\phi, \text{Alt}(\phi))$, the ones where the competence of the speaker is maximal. To account for this we need a new order that compares the competence of the speaker. This order is described in definition 3.

**Definition 3. (Ordering by consistency statements)**

$$s <_{\text{Alt}(\phi)}^{P} s' \iff \{\psi \in \text{Alt}(\phi) : s \cap [\psi] \neq \emptyset\} \subset \{\psi \in \text{Alt}(\phi) : s' \cap [\psi] \neq \emptyset\}.$$  

The minimal models in this ordering are those states where the speaker knows most about the alternatives. Now, finally, we define the function $[\text{Comp}]^S(X, \text{Alt}(\phi))$ ($\text{Comp}$ stands for competence) by selecting the minimal elements in $X$ according to the ordering $<_{\text{Alt}(\phi)}^{P}$:

**Definition 4. (Maximizing competence)**

$$[\text{Comp}]^S(X, \text{Alt}(\phi)) = \{s \in X : \neg \exists s' \in X : s' <_{\text{Alt}(\phi)}^{P} s\}.$$  

If we now apply $[\text{Comp}]^S$ to $[\text{Grice}]^S(\phi, \text{Alt}(\phi))$, where $\phi$ is a sentence like ‘[John]$_F$ came’ or ‘[John]$_F$ came or [Mary]$_F$ came’, we see that from the first we

25The same idea can be found also in [Spector, 2006].
can conclude that the speaker knows that Mary and Sue did not come, while from the second that the speaker knows that Sue did not come, but also that it is not the case that John and Mare came together. So, in this way we have accounted for the exclusive reading of ‘John came or Mary came’.

One can show (cf. [Van Rooij & Schulz, 2004; Spector, 2006]) that if we apply $\text{Comp}^S$ to $\text{Grice}^S(\phi, \text{Alt}(\phi))$ we derive exactly the same implicatures as we can derive using exhaustive interpretation. The exhaustive interpretation of $\phi$ with respect to its alternatives $\text{Alt}(\phi)$ is defined as follows:

$$\text{exh}(\phi, \text{Alt}(\phi)) = \{ w \in [\phi]^W : \neg \exists v \in [\phi]^W : v <_{\text{Alt}(\phi)} w \},$$

where $v <_{\text{Alt}(\phi)} w$ iff $\forall \psi \in \text{Alt}(\phi) : v \in [\psi]^W \rightarrow w \in [\psi]^W$.

How does our Gricean analysis account for the inference that it is not necessary that John came, if the speaker asserted ‘It is possible that John came’? This is quite straightforward. We can just assume that one of the alternatives of a sentence of the form ‘$\Diamond \phi$’ is the sentence ‘$\Box \phi$’. It is easy to see that from $\text{Grice}^S(\Diamond \phi, \text{Alt}(\Diamond \phi))$ we can conclude that the speaker doesn’t know that $\Box \phi$ is true. If we then assume that, in addition, the speaker is maximally competent on the alternatives, it follows that the speaker knows that $\Box \phi$ is false, and thus that $\phi$ is, in fact, not necessary the case.

### 3.5.2 A pragmatic analysis of free choice permission

Consider again sentences of the form ‘You/John may take the apple or the pear’, both represented by $\Diamond (\phi \lor \psi)$. Let us assume, for simplicity, that the alternatives of the embedded clause are just $\phi$ and $\psi$ itself, together with their conjunction and disjunction. Let us also assume that the set of alternatives to a sentence of the form ‘$\Diamond \phi$’ is just given by the set $\{ \Diamond \psi : \psi \in \text{Alt}(\phi) \} \cup \{ \Box \psi : \psi \in \text{Alt}(\phi) \}$. If we now apply the Gricean interpretation rule of the previous section, it is easy to see that things go wrong: we see that for each alternative to $\Diamond (\phi \lor \psi)$ (except the sentence itself, of course), there is a model that makes $\Diamond (\phi \lor \psi)$ true but not this alternative, but that there is no model that falsifies all these alternatives together: neither $\Diamond \phi$ nor $\Diamond \psi$ has to be true in order for $\Diamond (\phi \lor \psi)$ to be true, but they cannot be false both. There are various ways to go to solve this problem: either change what counts as an alternative, or the Gricean interpretation rule $\text{Grice}$. According to the first alternative, proposed by Schulz [2003], the set of alternatives to a sentence of the form ‘$\Diamond \phi$’ is just given by the set $\{ \Box \psi : \psi \in \text{Alt}(\phi) \} \cup \{ \Box \neg \psi : \psi \in \text{Alt}(\phi) \}$. First, notice that by applying $\text{Grice}$ to a sentence of the form ‘$\Diamond (\phi \lor \psi)$’ it immediately follows that the speaker knows neither $\Box \neg \phi$ nor $\Box \neg \psi$, in formulas, $\neg K \Box \neg \phi$ and $\neg K \Box \neg \psi$. What we would like is that from here we derive the free choice reading: $\Diamond \phi$ and $\Diamond \psi$, which would follow from $K \Box \neg \phi$ and $K \Box \neg \psi$. Of course, this doesn’t follow yet, because it might be that the speaker does not know what the agent may or must do.\footnote{Notice, though, that this inference does follow if ‘$\Box$’ and ‘$\Diamond$’ stand for epistemic must and...}
that the speaker is competent on this.\textsuperscript{27} Intuitively, this assumption means that the speaker thinks it is possible that the agent can or must do \textit{a} if and only if the speaker knows that the agent can or must do \textit{a}. In formulas: \( \Diamond \phi \equiv K \Box \phi \) and \( \Diamond \phi \equiv K \phi \). This assumption is completely natural for performatively used permission sentences, because in that case the speaker is the authority on what is permitted. But for assertively used sentences it is sometimes natural to make this assumption as well. Remember that after applying Grice, the minimal models falsify \( K \Box \neg \phi \) and \( K \Box \neg \psi \), which means that \( P \neg \Box \neg \phi \) and \( P \neg \Box \neg \phi \) have to be true. The latter, in turn, are equivalent to \( P \Diamond \phi \) and \( P \Diamond \psi \). By competence we can now immediately conclude to \( K \Diamond \phi \) and \( K \Diamond \psi \), from which we can derive \( \Diamond \phi \) and \( \Diamond \psi \), because knowledge implies truth. Thus, following Schulz’ [2003] minimal modal analysis, we get the free choice effect as a pragmatic inference.\textsuperscript{28}

One natural step would be to say that the speaker is competent on who John thinks might have passed the examination. In that case, the above minimal state disappears, and we will end up with \textit{two} minimal states according to the \( \preceq_L \) ordering: One where the speaker knows that only \( \Diamond_j P(a) \) is true and one where the speaker knows that only \( \Diamond_j P(b) \) is true. But — as we have seen in the beginning of this section — ‘only knowing’ doesn’t make sense in case we have more than one minimal state, so something has to be done. Perhaps what we do is to pragmatically reinterpret the sentence by first eliminating the above minimal states from the set of information states that we take as input (because these states can be expressed more economically by alternative expressions with a stronger meaning), and then apply the pragmatic interpretation function Grice to this more reduced set. In that case we end up with the desired result: we again have a unique minimal state, and in this state the speaker knows neither \( \Diamond_j P(a) \) nor \( \Diamond_j P(b) \), but she does know both \( \Diamond_j P(a) \) and \( \Diamond_j P(b) \).

\textit{Epistemic might.} This is so, because for the epistemic case we can safely assume that the speaker knows what he believes, which can be modeled by taking the epistemic accessibility relation to be fully introspective. This predicts correctly, because from ‘Katrin might be at home or at work’, it intuitively follows that, according to the speaker, Katrin might be at home, and that she might be at work.

\textsuperscript{27}Formally this is done by making a constraint on models: consider only models where the speaker knows the (deontic) accessibility relation of the agent.

\textsuperscript{28}It is easy to see that this analysis can account for the ‘free choice’-inference of the existential sentence as well: that from ‘Several of my cousins had cherries or strawberries’ we naturally infer that some of the cousins had cherries and some had strawberries. First we assume that the sentence is represented by something like \( \exists x[P x \land (Q x \lor R x)] \). Then we take the alternatives of this existential formula to be universal formula: \( \forall x[P x \rightarrow Q x], \forall x[P x \rightarrow \neg Q x], \forall x[P x \rightarrow R x] \), and \( \forall x[P x \rightarrow \neg R x] \). Applying Grice to these alternatives means (among others) that the speaker knows none of them. In formulae, this means that \( \neg K \forall x[P x \rightarrow Q x], \neg K \forall x[P x \rightarrow \neg Q x], \neg K \forall x[P x \rightarrow R x], \) and \( \neg K \forall x[P x \rightarrow \neg R x] \). This is equivalent to saying that the following formulae are true: \( P \exists x[P x \land \neg Q x], P \exists x[P x \lor Q x], P \exists x[P x \land \neg R x], \) and \( P \exists x[P x \lor R x] \). To strengthen this inference, we apply competence again. The relevant notion of competence now, of course, is that the speaker knows which \( P \)-individuals have property \( Q \) and/or \( R \). Making use of this competence assumption we can strengthen the possibility statements into knowledge attributions: \( K \exists x[P x \land \neg Q x], K \exists x[P x \lor Q x], K \exists x[P x \land \neg R x] \) and \( K \exists x[P x \lor R x] \). Because knowledge entails truth we infer (among others) to the conjunctive reading: \( \exists x[P x \land Q x] \) and \( \exists x[P x \land R x] \).
But, you will wonder, why does this procedure not also work for ‘p ∨ q’? Does this procedure not predict that from ‘p ∨ q’ we can conclude that the speaker knows both p and q? Indeed, on assuming competence, also these examples give rise to two minimal states. Eliminating those states without giving up competence now results in \( K(p \land q) \). For non-embedded disjunctions, however, we assume with others that their conjunctive counterparts are alternative expressions. But that means that in contrast to \( \Diamond_j(Pa \lor Pb) \), conveying such information could be done more transparently by semantically stronger alternative expressions, so in these cases, these interpretations are not allowed (notice that we use here some type of bidirectional interpretation procedure, taking also into account how the speaker would have expressed his information state). Giving up competence is the only rescue now, which is possible for ‘p ∨ q’.

BIBLIOGRAPHY


For more explicit bidirectional analyses of free choice permissions, see [Franke, 2010; van Rooij, 2010].


INTRODUCTION

For a long while, lexical semantics developed independently from formal semantics. Apart from a few, daring forays into the formal world (e.g., [Dowty, 1979]), lexical semanticists worked largely in isolation from formal semanticists, who were focused on sentential or discourse meaning. To make the point in a somewhat charicatural fashion, lexical semanticists investigated argument structure, verbal diathesis (shifts in meaning due to shifts in argument structure), polysemy, and meaning decomposition all within various “cognitive” systems lacking rigour and a tie to model theoretic semantics; meanwhile, formal semanticists paid little attention to matters of lexical meaning for open class terms like ordinary nouns and verbs—the meaning of a word $x$ was typically rendered as $x'$. Valuable work was done in both areas but there was something of a missed opportunity in which neither camp profited from the insights of the other.

1.1 From Discourse to the Lexicon

As work on formal semantics and, in particular, discourse semantics progressed, the need for a formal specification of the meanings of open class terms became more and more pressing to build up meanings of discourses compositionally. Formal semanticists working in discourse were no longer able simply to avoid thinking about the meanings of open class terms. The reasons why lexical meanings are so important to the composition of discourse meaning are a little involved. To explain them, I need to give a sketchy introduction to what the interpretation of a discourse involves.

An interpretation for a discourse depends not only on a compositional semantics for sentences but also on what one might call a “binder” rule.\footnote{The modeling of continuation style semantics for programs using the notion of a monad in category theory is due to [Moggi, 1991]. [Barker and Shan, 2006] and [de Groote, 2006] show the applicability of continuation methods for the semantics of natural language discourse.} Let the structure of a text be the sequence of its constituent sentences. Given a category of sentences $S$ and a “combinator”, $\cdot\cdot$, we define the category of texts $T$ as:

- $S \rightarrow T$
- $T.S \rightarrow T$
Thus, a sentence is a text and a text combined with a sentence is also a text. Where $\|T\|$ is the meaning (or meaning representation) of $T$ and $\|S\|$ is the meaning of a sentence whose meaning is to be added to $\|T\|$, a binder rule is an operation $b$ that takes a text meaning, combines it with a sentence meaning and returns a new text meaning that can be integrated with further text meanings:

$$\begin{align*}
(1) \quad b: \|T\| \times \|S\| & \rightarrow \|T\|
\end{align*}$$

All theories of discourse semantics have some form of binder rule. In a Stalnakerian semantics for discourse, where each sentence denotes a set of possible worlds, the operation $b$ is set theoretic intersection. In Hans Kamp’s dynamic semantics, *Discourse Representation Theory* (DRT), the operation $b$ is an operation of merge over *discourse representation structures* (DRSs), DRT’s meaning representations. In dynamic semantic systems like Predicate Dynamic Logic [Groenendijk and Stokhof, 1990], where the meaning of a sentence is a relation between an input assignment and an output assignment (relative to a fixed model), $b$ is the operation of relational composition. For the continuation based discourse semantics of [de Groote, 2006] and [Barker and Shan, 2006], the binder rule is a bit more complicated but fits into the general pattern.

For dynamic semantic theories, such as Segmentized Discourse Representation Theory (SDRT), that assign texts a meaning involving a rich discourse structure, the way $\|S\|$ combines with $\|T\|$ will sometimes depend on details of the lexical entries of the words in $\|S\|$. The basic premise of a discourse semantics like SDRT is that the way $\|S\|$ will combine with $\|T\|$ will depend on the rhetorical or discourse function that $\|S\|$ has in the context of $\|T\|$. Discourse functions affect many aspects of discourse meaning, including the resolution of anaphoric expressions and ellipsis, temporal structure, presupposition, and the interpretation of adverbials [Hobbs, 1979; Asher, 1993; Lascarides and Asher, 1993; Hobbs et al., 1993; Asher and Lascarides, 2003; Vieu et al., 2005]. To compute these components of interpretation, we need to compute the discourse functions of discourse constituents (which for the moment we may continue to think of as sentences or clauses).

Sometimes, a relatively small class of adverbs or adverbial phrases, what [Knott, 1995] and others have called *discourse connectors* or *discourse markers*, suffices to determine the discourse functions and hence a method of combination in SDRT. Syntactic constructions may also yield important clues as to discourse structure and how a new sentence, or rather discourse constituent, must combine with the text’s meaning. But sometimes the method of combination will depend on open class words like verbs, their arguments and their modifiers. To illustrate consider:

$$\begin{align*}
(2) \quad & a. \quad \text{John fell. He slipped.} \\
& b. \quad \text{John fell. He got hurt.} \\
& c. \quad \text{John fell. He went down hard, onto the pavement.}
\end{align*}$$

---

2DRSs are pairs of sets, the first element of which is a set of discourse referents which represent entities that the discourse talks about and the second element of which is a set of formulas over those discourse referents. The merge of two DRSs $(U_1, C_1)$ and $(U_2, C_2)$ is: $(U_1 \cup U_2, C_1 \cup C_2)$. For details see [Kamp and Reyle, 1993].

3These are conjunctions like *but*, *because* and *for*, adverbs like *also* and *too*, and adverbial phrases like *as a result*, and *and then*. [Knott, 1995] contains a long list of such markers, and computational linguists like Manfred Stede have constructed lists for other languages.
In each of (2a-c), the second sentence has a different rhetorical or discourse function, which is reflected in the way SDRT integrates its content with the discourse context. For example, SDRT would relate the discourse constituents denoted by the two sentences in (2a) by Explanation—John fell because he slipped. The constituents in (2b) would be related by Result, while the constituents in (2c) would be related by Elaboration. In each case, the rhetorical function can be traced to the meanings of the verb phrases.

Each of these discourse functions is defeasibly inferred from the way interpreters understand the combinations of the open class words and how their meanings combine in the predications. To make generalizations about the binder rule for a theory of compositional discourse interpretation, like SDRT, we need to have a lexical theory about words and how these words interact within predication. In particular, we need to group words into general types that would furnish the appropriate generalizations. However, in order to account for the diverse array of discourse relations these general types must be much more specific than the ones assumed by most compositional semantics, in which all common nouns and intransitive verbs have the same type $e \rightarrow t$, where $e$ is the type of entities and $t$ the type of truth values.

1.2 From the lexicon to discourse

In fact there is a two-way interaction between semantics and the lexicon. In the previous section, I argued that a good type-driven, lexical semantics is needed for a good discourse semantics. In the present section I will argue that a good, type-driven lexical semantics is dependent on discourse semantics and on a sophisticated account of semantic composition.

Many theories of word meaning countenance a rich typology, at least in principle, but these views still await a proper formal and conceptual analysis. Taking types seriously in one’s lexical semantics brings with it complexities and puzzles, some of which I want to bring out here. In general, these puzzles involve context dependency of a sort most compositional semanticists have ignored. For these semanticists, context sensitivity typically stops with anaphoric pronouns and indexical expressions; I believe, however, that context dependence pervades the lexicon.4

One of the intriguing but not well-understood observations about the composition of meaning is that when word meanings are combined, the meaning of the result can vary from what standard compositional semantics has led us to expect. In applying, for instance, a property term ordinarily denoting a property $P$ to an object term ordinarily denoting an object $a$, the content of the result sometimes involves a different but related property $P'$ applied to an object $b$ that is related to but distinct from $a$. While the choice of words obviously affects the content of a predication, the discourse context in which the predication occurs also affects it, where by discourse context I mean not only the predicational environment but also the discourse context to date. An important theme of current

---

4I should add, however, that in dynamic semantics especially DRT and in some philosophical circles, authors have argued that other linguistic elements have at least some sort of context sensitivity. These elements include modals, attitude verbs like believe, want and know and tense. See [Asher, 1986; Kamp, 1985; Roberts, 1989; Kamp and Rohrer, 1983; Kamp, 1979; Kamp, 1990; Veltman, 1996].
lexical and compositional semantics is how to make sense of this interaction. I illustrate with three types of context/lexical interactions.

**Discourse intrusions**

Prior discourse can sometimes affect how lexical meanings interact.

(3) All the children were drawing fish. Suzie’s salmon was blue.

In (3) we understand the relationship between Suzie and salmon in a complex way: Suzie is drawing a picture of a salmon that is blue. This interpretation is due not only to the genitive construction but also to its discourse environment. Here is an example of a different construction with the same moral.

(4) a. Julie began with the kitchen, proceeded to the living room and finished up with the bedrooms.

   b. Yesterday Julie cleaned her house. Julie began with the kitchen, proceeded to the living room and finished up with the bedrooms.

   c. Last week, Julie painted her house. Julie began with the kitchen, proceeded to the living room and finished up with the bedrooms.

As I argue in [Asher, 2011], the discourse in (4a) is not very felicitous because we don’t know what Julie is doing with the kitchen and the other rooms. It’s like trying to interpret an utterance of she’s nice in a context with no salient antecedent. However, once discourse specifies an activity, (4b) and (4c) are completely felicitous.

We will need something like SDRT’s rich notion of a discourse context and text meaning to account for discourse intrusions, something I will come back to in section 7. The problem is to specify in a precise and detailed way how this interpretation comes about without going so far as to say that any meaning shift is possible when made salient by the context.

**Concealed questions**

In a very interesting paper on concealed questions, [Percus, 2010] investigates meaning shifts concerning concealed questions. Consider these examples.

(5) a. John didn’t know how much the vase cost. John asked the price of the sales clerk. (what the price was).

   b. John didn’t know who Sam’s parents were. # He asked Sam’s mother of Julie (who Sam’s mother was).

(5b) is unintelligible whereas (5a) is fine. (5a) shows that in combination with certain noun phrases or DPs, ask can shift the meaning of its direct object or internal argument DP to have the meaning of an indirect question. But (5b) shows that it cannot do this for all DP internal arguments. Thus the meaning shift operation, however, it is to be implemented, cannot be a general operation over syntactic categories like DP; it must actually pay close attention to the meaning or semantic type of the internal argument.
Aspectual coercion

Aspectual coercion, in which an aspectual operator is applied to a verb phrase denotation, which specifies an eventuality type *inter alia*, to produce another verb phrase denotation and eventuality type description, is another example of a meaning shift. Aspectual coercion is quite language specific, and thus is not the result of any general pragmatic operation such as that considered by Neo-Griceans or relevance theorists such as [Sperber and Wilson, 1986; Recanati, 2004]. Consider, for example, (6), which involves the progressive aspect.

(6) a. #John is knowing French.
   b. John is being silly.
   c. John is just being John.
   d. John’s being an asshole.

One of the truisms about the progressive aspect is that stative constructions don’t support it, as shown in (6a). Nevertheless, (6b-d), which are progressivizations of the stative constructions *John is silly*, *John is John*, and *John is an asshole*, are perfectly unproblematic.

Interestingly, aspectual coercion with the progressive appears to be a particular feature of the English progressive aspect morpheme. Languages like French that lexicalize progressive aspect do not seem to support this meaning shift:

(7) a. Jean est idiot.
   b. #Jean est en train d’être idiot.
   c. Jean est en train de faire l’idiot.

Aspectual coercion is a thus language specific phenomenon and thus cannot be the result of a general cognitive principle of strengthening or weakening due to Gricean or Neo-Gricean constraints on communication. Such meaning shifts must be a part of the linguistic system, due to the meaning of particular words.

Another language specific aspectual coercion concerns the application of a perfective aspectual operator to a verb phrase containing an ability modal. Consider the following French examples. (8) translates roughly as *Jeanne had to take the train* and (8a) and (9a) use the perfective aspect, while (8b) and (9b) have imperfective aspect.

(8) a. Jeanne a dû prendre le train. → Jeanne a pris le train.
   (Jeanne had to take the train. → Jeanne took the train).
   b. Jeanne devait prendre le train. → Jeanne a pris le train.
   (Jeanne was supposed to take the train. → Jeanne took the train.)

(9) a. Jeanne a pu prendre le train. → Jeanne a pris le train.
   (Jeanne was able to take the train. → Jeanne took the train.)
   b. Jeanne pouvait prendre le train. → Jeanne a pris le train.
   (Jeanne was able to take the train. → Jeanne took the train.)

The → signifies an *actuality entailment*. Were we to consider ability modals as true modals that we can symbolize with □ and ◇, the actuality entailments in (8a) and (9a) would translate, respectively, to (10a) and (10b):
(10) a. $\square \phi \rightarrow \phi$.

b. $\Diamond \phi \rightarrow \phi$ or $\phi \rightarrow \square \phi$.

which implies a collapse of the modality (Bhatt 1999). However, with the imperfective aspect, these inferences vanish, and there is no collapse. The puzzle is, how can an application of the aspectual perfect collapse the modality? This is unpredicted and indeed bizarre from a Montagovian view of composition.

Actuality entailments with certain verb forms, like coercion with the progressive aspect, is a phenomenon particular to certain languages. In English, for instance, the actuality entailment does not appear to exist:

(11) John was able to take the train.

(12) John had to take the train.

(13) ?John has been able to take the train.

None of these have the actuality entailment, though they might have what one could call an actuality implicature. Once again, the actuality entailment cannot be the result of some general cognitive but non linguistic principle of strengthening. It is a semantic and lexically constrained kind of inference.

Matters are still more complex when one considers how temporal adverbials interact with modality and aspect to produce actuality entailments.5

(14) Soudain, Jean pouvait ouvrir la porte.

(Suddenly, Jean could open the door.)

In (14) the actuality entailment holds, despite the fact that the imperfective aspect is used. This is explained by the general observation that adverbs like suddenly coerce the imperfective aspect into an incoherent one with a perfective meaning. But once again we have a shift of meanings.

I believe that the apparent meaning shifts discussed in 1.2–1.2 should receive as uniform a treatment as possible within a semantic/pragmatic framework of lexical meanings and semantic composition—that is, how lexical meanings compose together to form meanings for larger semantic constituents like propositions or discourses. But we can only address this issue adequately within a larger view of how context affects interpretation. To this end, I will review the outlines of how dynamic semantic frameworks, including theories like SDRT, view discourse content computation. This will give us the tools with which to understand context effects at the level of clausal content composition and apparent meaning shifts. I will then discuss a couple of classic meaning shift cases and spell out the general approach to these that I favor, comparing it to recent pragmatic as well as semantic accounts.

5These observations are due to Vincent Homer, a discussion of which can be found in his paper for Journées de Sémantique et Modélisation, 2010, Nancy.
2 TOOLS FOR THE LEXICON FROM DYNAMIC SEMANTICS

In the last 30 years, computer scientists and linguists have developed sophisticated ways for modeling the effects of context on interpretation. These ways include various kinds of dynamic logics for programs and dynamic semantics, and there is often a close correspondence between them. Dynamic semantics, of which there are several schools (Discourse Representation Theory or DRT, Dynamic Predicate Logic or DPL, update semantics, and even some versions of situation semantics (Heim, Elborne)), treats the meaning of a natural language sentence as a relation between information states, an input information state and the output information state. Thus, each sentence corresponds to an action on the input information state, just as elements in a program are actions on the input computational state. The input information state represents the content of the discourse context to date while the output information state represents the content of the previous discourse context integrated with the content of the formula.

Various versions of dynamic semantics differ as to what are the input and output states. DRT, for example, incorporates this relational conception of meaning at a representational level. The input and output states are representations known as Discourse Representation Structures or DRSs. DRT proposes to build the update and dynamics into the construction of logical form but not of its interpretation. This makes the semantics of DRT static and provably equivalent to a Tarskian semantics, but it means that the construction of the logical form is a relatively complicated affair. It is unclear how the construction process for a DRS is to be interpreted compositionally in [Kamp, 1981]. It can be stated in a bottom-up version of a dynamicized lambda calculus but the results are rather quirky [Asher, 1993]. In reaction to DRT, linguists and philosophers have invented many compositional versions of dynamic semantics. DPL and various relational versions of DRT define the update and build in the dynamics at the level of semantic content [Fernando, 1994; Asher and Lascarides, 2003].

The interpretation of a discourse in such versions of dynamic semantics involves the relational composition of constituent sentences’ relational meanings. In dynamic semantics for natural languages, as well as in the dynamic semantics for programming languages, the interpretation of a formula can either function as a test on the input context or can transform the context. For example, John is sleeping in dynamic semantics yields a formula that functions as a test on the input context, which we can think of as a set of elements of evaluation. If an element of evaluation verifies the proposition that John is sleeping then it is passed on to the output context; if it does not verify the proposition, it does not become part of the output context. Operators like conditionals form complex tests on an input context C: an element of evaluation e will pass the test defined by If A then B just in case any output o from A given e will yield an output from B (given o as an input). Some sentences, for instance those containing indefinite noun phrases, output a context that is distinct from the input one. They transform elements of the input context; in particular they reset or extend the assignment functions that are parts of elements of the context to reflect the information they convey. On a view that treats assignments as total functions over the set of variables, an indefinite has the action of resetting an assignment that is part of the point of evaluation for formulas, as in Tarskian semantics. On a view where
assignments are treated as partial functions, the interpretation of an indefinite extends the assignment with a value to the variable introduced by the indefinite in logical form. This reset or extended assignment becomes part of the output context.

Logical forms and logics of composition are easily constructable for DPL or relational versions of DRT, using a suitably dynamicized version of the lambda calculus (Dekker 1999, Muskens 1996, Amsili and Roussarie 1998). This is all to the well and good since this makes dynamic semantics more comparable to ordinary truth conditional semantics. This allows us to follow standard practice in semantics and take logical forms to provide truth conditions (of the static or dynamic variety) and thus to be a basic component of semantics. A theory of meaning composition must yield interpretable logical forms for meaningful clauses; and a theory of discourse structure must yield logical forms for discourses. So the move towards compositional treatments of dynamic semantics is important.

On the other hand, the composition logic and the validity notion that results from such dynamic semantic frameworks is quite non-standard. Types that involve assignments of objects to variables or some equivalent become part of the type system in the lexical theory. This makes it hard to evaluate lexical entries within dynamic semantics with entries in classical theories like Montague Grammar. However, recently, researchers have used more sophisticated tools from computer science known as continuations to build contextually sensitive semantics within the confines of classical higher order logic. Continuations permit a faithful embedding of dynamic semantics into higher order logic. Without getting into the technical details, the idea of continuations is to build into lexical entries a “left” context parameter, which provides elements of discourse context relevant to interpretation like available discourse referents, together with a “right” context of the discourse to come. The trick is to get, for instance, indefinites to “pass on” the discourse referents they introduce to subsequent discourse while remaining within the framework of classical logic, according to which we end up with a classically valued proposition for a discourse. This is what various versions of continuation style semantics manage to do (de Groote, 2006; Barker and Shan, 2008; Moortgat and Bernardi, 2010). Roughly, anaphoric expressions in continuation style semantics select elements from the left context while indefinites update the left contexts with an element and the updated left contexts are then passed on to the right contexts. It is continuation style semantics that allows us to build the simpler lexical entries with a standard interpretation that are comparable to those developed in Montague Grammar.

So far dynamic semantics, and continuation style semantics in particular, have principally restricted themselves to the interpretation of anaphoric expressions and anaphoric dependencies. However, while anaphoric expressions like pronouns, verb tenses and ellipsis constructions are the clearest examples of context dependent interpretations, they are by no means the only examples. I maintain that the sorts of meaning shifts introduced...
in the previous section are all examples of context dependent interpretation.

While only a few words in continuation style semantics introduce elements that affect the right context so far, there is a sense in which almost all words introduce constraints on the context to come when it comes to selectional restrictions. The verb *try*, for instance, imposes on the compositional context “to come” that its subject must be an intentional agent; a verb like *hit* imposes the restriction that its object or internal argument must be a physical object. Thus,

(15) Mary hit John’s idea.

is predicted to be difficult to interpret unless the context allows us to interpret *John’s idea* as some sort of physical object (perhaps it’s his child or some artifact that he created).

Despite the ubiquity of the use of selectional restrictions in lexical semantics, few have examined exactly what sorts of things these are. Selectional restrictions of an expression $\epsilon$ pertain to the type of object denoted by the expression with which $\epsilon$ must combine. However, this information about the type of argument is not of a piece with the asserted content of the predication. It is rather a type of presupposed content. In effect selectional restrictions are *type presuppositions.* Selectional restrictions resemble presuppositions because their satisfaction seems to be a prerequisite for any expression’s containing them having a well-defined semantic value. Their demands for satisfaction or justification percolate up the semantic construction very much in the way that ordinary presuppositions do. That is, (16a,b) make the same demands on the context of interpretation that the unembedded (15) does, patterning in a similar way to the presuppositional content of definite descriptions like *the present King of France.*

(16) a. Mary didn’t hit John’s idea.
   b. Did Mary hit John’s idea?
   c. John didn’t see the present King of France at the exhibition.
   d. Did John see the present King of France?

To understand better type presuppositions, I briefly survey the current status of presupposed content in dynamic semantics. In dynamic semantics, presuppositions constitute a particular sort of test on the input context. Consider a sentence like

(17) Jack’s son is bald.

The presupposition generated by a definite noun phrase like *Jack’s son* (namely, that Jack has a son) must be satisfied by the input context, if the interpretation of the rest of the sentence containing the definite is to proceed. One way of satisfying the presupposition of (17) is for it to be already established in the context of utterance of (17) that Jack has a son. This can occur, for instance, when (17) is preceded by an assertion of *Jack has a son.* Presuppositions can also be satisfied by contexts within the scope of certain operators, as in (18), even though it has not been established in the discourse context that Jack has a son:

(18) If Jack had a son, then Jack’s son would be bald.

---

7 That selectional restrictions are type presuppositions is a fundamental principle of [Asher, 2011].
8 One of the great successes of Dynamic Semantics has been to show that the behavior of presuppositions introduced by material within the consequent of a conditional follows straightforwardly from the conception of the conditional as a complex test on the input context and thus offers a solution to the so called *projection problem.*
In dynamic semantics the evaluation of the consequent is done against the background of an update of the prior context with the content of the antecedent of the conditional. It is in such a context that the presupposition generated by the definite description *Jack’s son* is evaluated. The satisfaction of the presupposition by the antecedent of the conditional in (18) means that the presupposition places no requirement on the input context to the whole conditional, the context of utterance. The satisfaction of the presupposition by elements of the discourse context entails that the presupposition does not “project out” as a requirement on the context of utterance; (18) is consistent with the assertion that in fact Jack has no son. On the other hand, dynamic semantics predicts that if we change (18) just slightly so that the antecedent does not provide a content that satisfies the presupposition, the presupposition will project out as a requirement on the input context to the whole conditional:

(19) If Jack were bald, then Jack’s son would be bald too.

Selectional restrictions act in the same way in similar contexts. For instance, to say something like

(20) The number two is blue.

is to invite at the very least quizzical looks from one’s audience, unless the context makes clear that *the number two* refers to some sort of physical object and not the only even prime. However, a counterfactual with an admittedly bizarre antecedent can satisfy the type presupposition projected from (20) in much the same way as the antecendent of (18) satisfies the presupposition of the consequent:

(21) If numbers were physical objects, then the number two would be blue.

What happens when a presupposition cannot be satisfied by the discourse context? It depends on what sort of presupposition is at issue. Some presuppositions, such as those introduced by definite noun phrases, are easily “accommodated.” In dynamic semantic terms this means that the input context is altered in such a way so that the presupposition is satisfied, as long as the result is consistent. Other presuppositions, such as that generated by the adverbial *too*, are much less easily accommodated. Given that operators like conditionals can add “intermediate” contexts between the context of utterance and the site where the presupposition is generated, we need a theory of where and how to accommodate in case the input context does not satisfy the presupposition. In some theories of presupposition that operate on semantic representations like that of [van der Sandt, 1992], accommodation simply involves adding a formula for the presupposition to an appropriate part of the representation for the discourse. Van der Sandt stipulates a particular procedure for handling the binding and accommodation of presuppositions: one attempts to bind a presupposition first, trying first to bind it locally and then in a more long distance way. If binding fails, one tries to accommodate at the outermost context first and if that fails then one tries to accommodate at the next outermost context. The constraints on accommodation are that the addition of the presuppositional material be consistent with the discourse context. So, for instance, one cannot add the presupposition that Jack has a son to a context where it is established that Jack does not have a son.

---

9In other theories like Heim’s [1983] the accommodation procedure is not really well developed; but see [Beaver, 2001] for a detailed account of accommodation in a Heimian approach to presupposition.
Something similar happens with selectional restrictions or type presuppositions. Type presuppositions in normal circumstances are bound or justified in that the type of the argument expression matches the type presupposition of its predicate. But they can sometimes be accommodated. Consider the noun water. It can combine with determiners that require either a noun that denotes something of type mass (22a) or with determiners that are intuitively count determiners:

(22) a. some water.

b. a water.

One way of accounting for this is that water itself does not determine its denotation to be either of a subtype of type mass or of type count. If that is the case, then we can accommodate the requirements of the determiner simply by applying the type count or mass to the type of the expression water— in simplified terms, a water ends up denoting a property of properties that have a non empty intersection with the collections of portions of water.¹⁰

3 FROM TYPE PRESUPPOSITION TO COERCION

Sometimes the argument does not satisfy the type presupposed by its argument and cannot be accommodated in the given context. In that case, semantic composition crashes and there is no well-defined value for the semantic composition, as in (15). It is not obvious what are the principles for accommodating type presuppositions,. They sometimes permit the rescue of a predication in those cases where the argument’s type does not satisfy the predicates type presuppositions but in other cases they don’t. Examining this issue takes us to the heart of coercion.

Consider the following example, discussed at length in the literature, in particular by [Pustejovsky, 1995].

(23) Julie enjoyed the book.

The intuition of many who work on lexical semantics is that (23) has a meaning like (24) with the doing something filled in by an appropriate activity

(24) Julie enjoyed doing something to (e.g., reading, writing, ...) the book.

The intuition is this: enjoy requires an event as its direct object as in enjoy the spectacle, enjoy the view. This also happens when enjoy takes a question as its complement, as in enjoy (hearing) what he said. When the direct object of a transitive use of enjoy does not denote an event, it is “coerced” to denote some sort of eventuality.

If the intuitions behind coercion are easy to grasp,, however, modelling coercion by formal means is rather difficult. Is it, for instance, just the transformation of the denotation of the noun phrase the book into some sort of eventuality denoting expression?¹¹ If that is

¹⁰Not all mass nouns work so well. A. Kratzer pointed out to me the word blood; in an out of the blue context, it is difficult to interpret a blood as opposed to some blood. It seems that one can apply a count determiner to what is ordinarily a mass noun, only in contexts where there is some contextually salient or conventionally determined portion of matter of the mass.

¹¹The account in [Pustejovsky, 1995] seems to adopt this line of attack toward the problem.
the case, then how can we access in subsequent discourse the referent of the book?\footnote{This problem makes the qualia story as developed in [Pustejovsky, 1995] a non starter as I argue in [Asher, 2011].}

(25) Julie enjoyed the book. It was a mystery.

These observations are familiar but show that we cannot shift the meaning of the book to some sort of eventuality. Or at least if we do, whatever process is responsible for the shift must also allow the book to retain its original, lexical meaning and its original contribution to logical form.

The other alternative explored in the literature is to shift the predicate, in this case the verb, to mean something like enjoyed doing something with.\footnote{[Nunberg, 1995], and others have pursued this line of attack.} At the very least, however, one will need some underspecified form of coercion to handle cases of gapping coordination like the one below.

(26) Julie enjoyed (reading) a book and Isabel (watching) a movie.

Anaphora tests using ellipsis are instructive here. The meaning of the activity is predicted on the predicate modification view to shift with the choice of direct object during semantic composition. But if that is the case, then the ellipsis predicts a peculiar reading of the second clause of (26). Things go worse when the ellipsis involves a non event DP and an event DP as in the following example:

(27) Julie enjoyed her book and Isabel her parade.

In (27) Isabel simply enjoyed the parade as an event. Or perhaps she enjoyed being in the parade. Isabel and Julie did different things with the book and the parade—and simply shifting the predicate doesn’t do justice to this intuition.

A quick survey of web based examples shows how much trouble the predicate shifting approach to coercion gets into:


b. Julie enjoyed a book and Justin watching a movie.

c. I actually enjoyed the book and her writing (Google).

d. I enjoyed the movie and the story line but of course, I wish that the movie had not glorified Buddhism so much. (Google search, http://www.christiananswers.net/spotlight/movies/2003/thelastsamurai.html)


g. I really enjoyed the book and the biological innovations and theories expressed in it. (http://www.powells.com/blog/?p=6863)

h. I really enjoyed the book and how it uses someone’s life in the story and gives ideas on how to stay unstressed and calm. (http://www.silverhillsontheroad.com/teens/teenbook)
While (28a-b) aren’t the best English, they are grammatical; and they, together with the attested examples above, all seem problematic for the predicate modification view. To handle the DP coordination in (28a), the predicate modification view predicts that Julie enjoyed doing something to watching a movie, which is uninterpretable. What would she enjoy doing to watching a movie? Did she enjoy watching watching the movie? Similarly, the predicate modification view would predict an uninterpretable reading for the sluicing example, (28b). All this is strong evidence that coercion does not involve a semantic adjustment to the predicate at the moment of composition.

One might suppose that in fact coercion and type presupposition accommodation in general is not part of semantics at all, but rather a pragmatic mechanism. The ellipsis facts seem to point to a pragmatic account according to which the appropriate enrichments to predicates and/or arguments occur after semantics has done its job, and hence phenomena like ellipsis have been resolved. But as I argued earlier, pragmatic accounts have difficulty accounting for the language specificity of many coercions. Pragmatic principles are supposed to follow from general principles of rational interaction between agents, and so they are expected to be universal; but we’ve already seen that coercions are typically language specific. In addition, we should expect pragmatic enrichment to allow coercions, say from objects to eventualities, whenever the grammar allows for an eventuality reading in the argument. But this isn’t true. Consider


b. #The book started at 10.

c. John started the book.

It’s perfectly straightforward to get the eventuality reading for the object of start (29c) in an out of the blue context. But it’s impossible to do this when the intransitive form of start is used with the book in subject position, even though an eventuality reading for the subject of intransitive start is available, as seen by (29a) and indeed mandatory.

So if you can’t shift the meaning of the predicate and you can’t shift the meaning of the argument and you can’t relegate the problem of coercion to the pragmatics garbage can, what is left? My answer is that you change the relation of predication that holds between the predicate and the argument. A type clash between the type of the argument and the type presupposition of the predicate induces not a type shift in either the argument or the predicate but rather a type shift on the predication relation itself, which is implemented by introducing a functor that is inserted around the argument. The meaning of the argument does not shift—it remains what it always was; the predicate also retains its original meaning, but what changes is how they combine. Given that the sluicing examples recover just the verb’s meaning, this account makes the right predictions for our examples.

Such shifts in the predication relation are lexically governed. It is the verb enjoy that requires an event but also which licenses a change in the predicational glue between it and its object when it is not of the right sort. In addition, as I argued in [Asher, 2011], these licensings are proper to certain arguments of the verb or rather to the syntactic realization of other arguments of the predicate. Once again we are not dealing with a phenomenon of general pragmatic strengthening or repair but rather with a problem at the syntax/semantics interface, which is how to account for the differences in (29).
My view requires the background assumptions that 1) verbs, and predicates more generally, distinguish between arguments that denote eventualities and those that denote, say, physical objects and 2) there is a clear distinction between physical objects and eventualities. But it seems to me that both of these assumptions are cogent and well supported linguistically across a wide spectrum of languages. In section 5, I sketch a formal system that works out my view technically. First, however, we have still not finished with the analysis of what sort of information type presuppositions are.

4 MORE ON TYPES

Types are semantic objects and they are clearly linked to denotations. Montague Grammar (MG), which uses Church’s conception of types [Church, 1936], according to which types are identified with the set of entities in the model of that type, countenances two basic types: the type of all entities \( e \), which is identified in a model with the domain of the model, and the type of truth values, \( t \). Further types are defined recursively: for instance, all nouns have the type \( e \rightarrow t \), which is the set of all functions from the domain of entities into the set of truth values. Types, however, do play a role even in Montague’s system. They have enough semantic content to check the well-formedness of certain predications. But the type system in MG is far too impoverished to be able to determine the semantic values of expressions and hence the truth conditions of sentences or discourses.

Given the sort of type checking relevant to examples of coercion that I have surveyed here, the type system must incorporate many more distinctions. For instance it must distinguish various distinct basic types that are subtypes of \( e \); it must distinguish the type for eventualities from the type for physical objects, as well as distinguish the type informational objects from these two. I’ve already indicated that even more fine-grained distinctions will be necessary in the type system: for instance, the distinction between states and events must be reflected there, and there are many other semantic distinctions that end up, on my view, as distinctions between types. This already complicates the task of doing lexical semantics considerably, because the standard Church/Montague conception of types as sets fails to give a sensible notion of subtyping for higher functional types. For example, the Church/Montague conception of types as sets predicts that the type of physical properties and the more general type of first order properties have no common inhabitants, because the set of functions from physical objects to truth values and the set of functions from all entities to truth values are disjoint.\(^{14}\)

Once one has embarked upon the task of providing a richer set of types, it seems that there is no natural stopping place for the taxonomy. Semantic well-formedness may depend on type distinctions that are so finegrained that the type system postulates almost a type for each lexical meaning. The question then arises, why have a denotational semantics in addition to a semantics given in terms of types? Those working in type theoretic semantics, like [Ranta, 2004; Cooper, 2005; Fernando, 2004] inter alia, argue that there is no need to keep these two kinds of semantics for expressions and have worked out semantic approaches that use types. Nevertheless, even in this much richer system of types,

\(^{14}\)For more discussion, see [Asher, 2011].
it is important to distinguish the task of checking for semantic well-formedness from the task of delivering truth conditions and truth at a point of evaluation. For one thing, the two tasks demand intuitively vastly different computational and cognitive resources. A competent speaker should be able to check whether a predication in a given context is grossly semantically ill-formed or not, whereas deciding whether an arbitrary sentence of English is true is impossible for most if not all speakers. The former is part of semantic competence, whereas the latter is not. Granted that there is a continuum of cases of semantic well-formedness and that there is some interaction between the two,\textsuperscript{15} it seems nevertheless clear that the poles of the continuum are distinct. Semantic competence requires speakers of a language $L$ to be able to judge whether sentences of $L$ are semantically well-formed, but it in no way requires speakers of $L$ to be able to decide which $L$ sentences are true. Fleshing out this line of thought yields two semantics for $L$ expressions: one for checking well-formedness, one for determining truth. The former is a semantics internal to the linguistic or at least conceptual system, whereas the latter is external to the conceptual system, linking expressions to their real world denotations, represented model-theoretically. As the foundation of an internal semantics for the language, types have a natural characterization as proofs or computations that the conceptual system can carry out. For instance, the type of first order properties, which is standardly understood as $e \Rightarrow \tau$, defines a computation which given any object of type $e$ (an entity) furnishes an object of propositional type. This is exactly what is required, say, when checking whether a first order property can combine with a determiner, since the latter requires as an input that is just such a computation and outputs the familiar generalized quantifier type, now also interpreted as a computation.

While the internal semantics cannot determine truth conditions in the external semantics for well-known externalist reasons,\textsuperscript{16} there are still very interesting connections. For instance, connections of subtyping in the internal semantics translate into analytically true statements in the external semantics; for instance, if $\text{lion}$ is a subtype of $\text{animal}$ in the type system, this means that $\text{lions are animals}$ is analytically true. It is a delicate question of what the subtyping relation is in the internal semantics, but there are some clear cases that add substance to the philosophically and empirically thin conception of analyticity, much maligned since Quine’s seminal “Two Dogmas of Empiricism” [Quine, 1951]. For example, the type $\text{physical object}$, which I abbreviate as $p$, and the type $\text{informational object}$, abbreviated as $i$, are clearly subtypes of the type of entities $e$, and so $\text{physical objects are entities, as are informational objects}$ would be an analytic truth. Similarly, if two types have incompatible properties, such as the types $\text{physical object}$ and $\text{informational object}$ (inhabitants of the latter can have multiple instantiations at distinct spatio-temporal locations whereas inhabitants of the former cannot), then $\text{physical objects are not informational objects}$ is also an analytic truth.

We can enrich the system of types further so that certain expressions say adjectives involve a type that is a function from types to types and whose value changes given the

\textsuperscript{15}Thanks to Ken Shan for this remark

\textsuperscript{16}By this I mean all the ink that philosophers have spilled on the Putnam-Kripke-Burge thought experiments to conclude that what $L$ speakers have in the head does not determine in many cases the denotations of $L$ expressions.
input type. A natural language example for which this might be an attractive option is the adjective flat. It is clear that flat changes its meaning rather dramatically depending on the noun it combines with:

(30) a. flat country.
    b. flat curvature.
    c. flat tire.
    d. flat beer.

Once such polymorphic types are countenanced (the type of the output depends on the type of the input), we can also handle morphological transformations like those in so-called psych verbs to furnish a rich list of analytic entailments. For instance, John angered Fred analytically entails that Fred was angry.17

5 A SKETCH OF A FORMAL THEORY OF LEXICAL MEANING

As I mentioned in the discussion of types above, we need a way of considering types other than the Church conception of types as sets of their inhabitants. Luckily, there are other models of types and the λ calculus that we can exploit such as those given by Type Theory [Martin-Löf, 1980; Luo, 1999] or Category Theory, according to which types are understood as proofs or computations. These theories furnish sensible notions of subtyping for both atomic and higher order or complex types. In this section, I detail a fragment of the type system TCL of [Asher, 2011], which includes simple types, functional types and the functional polymorphic types that will serve in the analysis of coercion. This will give a glimpse of the richness and complexity of lexical semantics.18

In providing a formal system, there are two things we need to do. We have to implement talk of type presuppositions in a formal framework, and then we have to sketch the rules of composition. As regards the first task, predicates impose two type constraints on the types of their arguments, one an absolute requirement and occasionally another which licenses a modification of the predicational environment between the predicate and its argument. This information flows to the argument. When the type presuppositions imposed by one or more predicates agrees completely with the type of an argument, the type presuppositions are justified and the term combining predicate(s) and argument(s) gets a well-formed type. If the type presuppositions are not directly satisfied, certain indirect justification or accommodation strategies are available that will enable the term to get a well-formed type—which will depend on what modifications of the predicational environment are allowable.

To allow the presupposition justification mechanisms to do their work, we need to separate the type presuppositions from the rest of the term. And this means complicating the lexical entry of all words to include a context parameter in which these type presuppositions can be encoded. In order to pass the presupposition from the noun to a modifier,
for instance, I will make use of this presupposition parameter, which is called $\pi$. $\pi$ is a list of presuppositions imposed by predicates on their arguments that gets added to as composition proceeds and that presupposition justification mechanisms can subsequently adjust. This dynamic aspect of the presupposition parameter makes it convenient to use a continuation style semantics in the style of [de Groote, 2006]. The presupposition parameter acts thus just like de Groote’s left context parameter, and it will figure in lexical entries in a similar way.

How does this parameter differ from the notion of a typing context already found in some approaches to the $\lambda$ calculus (a typing context can be understood as a function from terms to types)? In the standard $\lambda$ calculus, typing contexts encode types on terms. The context is fixed and has no particular effect other than the standard one which is to check the applicability of the rule of application. On the other hand, the type parameter in continuation semantics can be an arbitrary data structure on which various operations may be performed separately from the operations of $\beta$ reduction in the $\lambda$ calculus. This is the sort of flexibility we need to handle various operations of presupposition justification at the lexical level.

Before we can see how the use of the parameter $\pi$ transforms the rest of the type system, we have to revisit the question of how a noun and its modifiers combine and how the type presuppositions percolate from one to the other. This is a familiar problem to linguists in a new guise; in dealing with presuppositions, linguists have to determine what expressions or linguistic constructions trigger the presuppositions, and there has been a good deal of debate about that issue over the past 20 years or so. The same is true here, but the problem boils down to the question, Which is the predicate and which is the argument in a noun modifier construction? I follow tradition and assume that modifiers take first order properties as arguments of the sort traditionally associated with nouns. However, the vast majority of adjectives are subsective, and those that aren’t “preserve” in some sense the type of the noun. This observation suggests that nouns actually pass their type presuppositions to adjectives, not the other way around. That is, the adjective is in fact an argument to the noun, which means the lexical entries for nouns should take modifiers as arguments.

Thus, letting $P$ be a variable of modifier type, i.e. of type $1 \Rightarrow 1$ (where 1 is the general type of first order properties), the functional type from first order properties to first order properties, the lexical entry for $\text{tree}$ looks like this:

\[
\lambda P \lambda \lambda \pi P(\pi \ast \text{arg}^{\text{tree}})(v)(\lambda v . \lambda \pi' \text{tree}(v, \pi')).
\]

In this lexical entry, the noun takes the modifier as an argument using $P$.\footnote{I suppose that when no adjective is present, $P$ in (32) applies to the trivial subsective adjective, $\lambda P \lambda \lambda \pi P(\pi)(\lambda x)$ and so gives a predictable type for a simple NP. I deal with multiple adjectival modifiers through type raising to form a complex modifier; the exact logical form for such raised NPs will be a matter of meaning postulates. The issue of multiple adjectival modification gets complicated when we have a subsective combined with a non-subsective adjective as in $\text{fake, yellow gun}$ or $\text{yellow, fake gun}$. It seems that both of these NPs have the reading of denoting an object which is a fake gun and yellow; nevertheless, standard compositional accounts will not be able to provide this reading for $\text{fake, yellow gun}$ without resorting to meaning postulates as well. For more details see [Asher, 2011; Partee and Borschev, 2004].} The modifier is then applied to something that looks like the standard entry for $\text{tree}$, but in addition it has a presupposition parameter as an argument with the type requirement provided by the noun.
In this way the noun’s type presuppositions are passed to the modifiers. Presupposition parameters range over lists of type declarations on argument positions for predicates. In the lexical entry (31), the lexical entry for tree adds a type presupposition to the input list of type declarations that is of the form: \( \text{arg}_{\text{tree}} \cdot : p. \). This is a type declaration that the first argument of tree must be of physical object type, or of type \( p. \). This argument position is always associated with the referential or denoting argument of a noun. \( \pi^* \cdot \text{arg}_{\text{tree}} \cdot : p. \) says that whatever fills this first argument position must have a type that justifies the type \( p. \). This includes the variable \( v \) in the lexical entry (31), but it will typically also include a variable that is an argument of the modifier. So summing up, (31) says that a noun like tree takes a modifier as an argument in order to pass along its type presupposition requirements to the modifier. In order for the modifier to combine with the noun, it must justify the presupposition that is appended via \( \pi^* \) to its presupposition parameter argument.

This way of doing things may look complicated in comparison to the usual type declarations in the \( \lambda \) calculus, but the flexibility provided by this formalism allows us to keep things modular: the \( \lambda \) calculus rules are separate from the presupposition justification rules. This helps considerably in reducing notational clutter and allows researchers to tinker with the presupposition adjustment mechanisms, which may vary across linguistic expressions, without messing with the basic composition rules. To save on notational clutter, I’ll often just write \( \text{arg} \) when the predicate is obvious, or just list the type when the argument position and predicate is obvious. By integrating this \( \lambda \) term with a determiner and a verb, more type presuppositions will flow onto the presupposition parameter \( \pi \).

I haven’t said what the type of \( x \) is explicitly in (31). To make the justification of presuppositions as flexible as possible, the type presupposition of \( x \) should be very weak, and so I shall take it to be of type \( e. \) To avoid clutter, I’ll sometimes use the standard notation of the \( \lambda \) calculus for readability, but these presuppositions should be understood as generated by the predicate and stored in the local presuppositional parameter. Thus, the final entry for a tree will be:

\[
(32) \quad \lambda P: \text{mod} \\lambda x: e \quad \lambda \pi \quad P(\pi^* \cdot \text{arg}_{\text{tree}} : p.)(x)(\lambda v. \lambda \pi' \cdot \text{tree}(v, \pi')).
\]

In practice, I will omit the typings of terms when they are obvious. Given these abbreviations, the general type of nouns shifts from the standard \( e \Rightarrow t \) to \( \text{mod} \Rightarrow 1 \), where \( \text{mod} \) is the type of modifiers. NPs have the relatively simple type schema, \( \alpha \Rightarrow (\Pi \Rightarrow \tau) \), where \( \alpha \) is a subtype of \( e. \)

Let’s look briefly at determiners. All singular determiners require their head noun phrases to be first order properties. Most determiners, however, also carry type presuppositions on the sort of individuals they quantify over. For instance, the mass determiner expressions in English much and a lot of require that the elements in the domain of quantification be masses or portions of matter, another subtype of \( e. \) The determiner \( a \) on the other hand requires the variable it binds to range over quantized or countable objects. These presuppositions interact with the presuppositions conveyed by the NPs with which determiners combine. Consider

\[
(33) \quad \text{I’ll have a Chardonnay (a water, a coffee).}
\]

(33) means that I’ll have a particular quantity of Chardonnay, water or coffee, a glass, a cup, perhaps even a bottle, or thermos. Conversely, mass determiners impose portion of matter readings:
There was a lot of rabbit all over the road. (34) means that there was a considerable quantity of rabbit portions of matter all over the road. This leads me to suppose that the determiner imposes a type presupposition on the individual argument that interacts with the type presupposition of the NP. In the next section, we’ll see how. Linguistically, this is interesting, because it appears that many NPs are in fact underspecified with respect to the mass/count distinction; the mass/count distinction is imposed by the determiner in English (and in other languages, like Chinese, by the classifier).  

These observations lead to the following entry for $a$:

$$\lambda P: 1 \lambda Q: 1 \lambda \pi \exists x (P(\pi \ast \text{ARG}_P^\pi; \text{COUNT})(x) \land Q(\pi)(x)).$$

With this in place, we are almost ready to look at a simple derivation of a DP meaning within the formal system. But we first have to set out the basic rules of composition. Besides the normal rules of the $\lambda$ calculus, we have the following basic operations to justify presuppositions. The triggering configuration for these rules is a pair of type requirements on a given argument.

The rule Binding Presuppositions provides an account of binding with respect to lexical type presuppositions encoded in $\pi$. Suppose our context $\pi$ contains the information that type $\alpha$ is supposed to be placed on some argument position $i$ of a predicate $P$ but the variable that must occupy that argument position already has the type $\gamma$ assigned “locally” by an argument of $P$. That is, the argument of $P$ actually has a type requirement $\text{ARG}_j^Q: \gamma$. When this occurs I will write $\text{ARG}_j^Q \mapsto \text{ARG}_i^P$.  

Sometimes presuppositions can’t be bound. With clausal presuppositions, when a presupposition cannot be satisfied, its content is accommodated or added to a place in the discourse structure if it is consistent to do so. Something similar holds in TCL for type presuppositions. Suppose that an argument position $i$ for some predicate $P$ has type $\alpha$ but there is a presupposition imposed on it from another term $Q$ that it have type $\beta$. In this case, the local presupposition parameter $\pi$ will look like this: $\pi \ast \text{ARG}_j^Q: \alpha \ast \text{ARG}_j^Q: \beta$. If $\alpha \cap \beta \neq \bot$, then the type presupposition will simply be added to the typing of $t$. This is what the rule of Simple Type Accommodation states:

\[ \text{Simple Type Accommodation} \]

$$\gamma \sqsubseteq \alpha, \text{ARG}_j^P: \alpha, \text{ARG}_j^Q: \gamma, \text{ARG}_j^Q \mapsto \text{ARG}_j^P \\
\text{ARG}_j^P: \gamma$$

20 In this TCL agrees with the “exoskeletal” approach of [Borer, 2005a; Borer, 2005b].

21 In practice, the local type assignment will always be the last on the string of type assignments appended to the local presuppositional parameter, or at least outside those typing requirements imposed by other terms.
Simple Type Accommodation

\[
\frac{\alpha \cap \beta \neq \bot, \text{ARG}^P_1: \alpha, \text{ARG}^Q_1: \beta, \text{ARG}^Q_1 \rightarrow \text{ARG}^P_1}{\text{ARG}^P_1: \alpha \cap \beta}
\]

Binding is a special case of Simple Type Accommodation (since \(\alpha \cap \beta = \alpha\), if \(\alpha \sqsubseteq \beta\)).

Let us look at an example of adjectival modification involving the intersective adjective heavy. Heavy has the lexical entry in (38a). Note that it does not impose special typing requirements on its property argument \(P\) or on \(x\):

\[(38)\]

\begin{enumerate}
  \item \(\lambda P: 1 \lambda x: 1 \lambda \pi'': (P(\pi')(x) \land \text{heavy}(x, \pi'' \ast \text{ARG}^\text{heavy}_1 : \mathbb{P})).\)
  \item Applying the lexical entry for tree to (38a), we obtain:
    \(\lambda z \lambda x \lambda \pi'' (P(\pi')(x) \land \text{heavy}(x, \pi'' \ast \text{ARG}^\text{heavy}_1 : \mathbb{P}))\)
    \((\pi \ast \text{ARG}^\text{tree}_1 : \mathbb{P})(z)(\lambda u \lambda \pi' \text{tree}(u, \pi')).\)
  \item Using the normal rules of the \(\lambda\) calculus, we get:
    \(\lambda z \lambda \pi (\text{tree}(z, \pi \ast \text{ARG}^\text{tree}_1 : \mathbb{P}) \land \text{heavy}(z, \pi \ast \text{ARG}^\text{tree}_1 : \mathbb{P} \ast \text{ARG}^\text{heavy}_1 : \mathbb{P})).\)
  \item Binding now reduces the presuppositions and we get the finished result:
    \(\lambda z \lambda \pi (\text{tree}(z, \pi \ast \text{ARG}^\text{tree}_1 : \mathbb{P}) \land \text{heavy}(z, \pi \ast \text{ARG}^\text{tree}_1 : \mathbb{P} \ast \text{ARG}^\text{heavy}_1 : \mathbb{P})).\)
\end{enumerate}

Just as with clausal presuppositions (see (39) where his father is bound to the presupposed content of the proper name John), one presupposition can justify another: the type presupposition on the adjective heavy justifies, indeed satisfies, the type presupposition of the noun tree in (38d).

(39) John’s son loves his father.

As an example of Simple Type Accommodation, consider:

(40) I’ll have a Chardonnay.

The type presupposition of the determiner is count, but Chardonnay is neither mass nor count and in the appropriate predicational context it can be either. So in this case we get for the DP meaning:

\[(41)\]

\(\lambda Q: 1 \lambda \pi \exists x (\text{Chardonnay}(x, \pi \ast \text{ARG}^\text{count}_1 : \mathbb{P} \cap \mathbb{P} \land Q(\pi)(x)).\)

In (38b), we had two compatible typings on the same variable. When we attempt the same derivation for the modification heavy number, we get an irresolvable type clash in the presuppositions.

\[(42)\]

\(\pi \ast \text{ARG}^\text{number}_1 : \text{ABSTRACT} \ast \text{ARG}^\text{heavy}_1 : \mathbb{P}.\)

We cannot justify the type presuppositions of the noun and the adjective with any of our rules. Binding doesn’t work, and the presuppositions cannot be accommodated because the intersection of the types presupposed by number (1 or informational object) and heavy (\(\mathbb{P}\) or physical object is empty). So the derivation crashes and no well-formed lambda term corresponds to this noun phrase and it has no proof-theoretic interpretation.

Just as with complex clausal presuppositions, there is an order in which these presuppositions must be integrated into logical form. In particular, the presupposition introduced by the determiner must be satisfied or accommodated by the NP prior to dealing with the
presupposition introduced by the predicate that makes up the nuclear scope of the determiner. This leads to an explanation of the judgements in (43).

\[(43) \quad \begin{array}{ll}
\text{a.} & \text{A water has 12 ounces in it.} \\
\text{b.} & \ast \text{Much water has 12 ounces in it.}
\end{array} \]

Once we have integrated the mass presupposition on the variable bound by the determiner in (43b), it can no longer go with the obligatorily count property in the VP.

Further evidence for the ordering of presuppositions and their justification comes from looking at complex cases of coercion. Consider,

\[(44) \quad \begin{array}{ll}
\text{a.} & \text{The Chardonnay lasted an hour.} \\
\text{b.} & \ast \text{Much Chardonnay lasted an hour.}
\end{array} \]

The preferred reading of (44a) is that a particular quantity of chardonnay participated in some event (presumably drinking) that lasted an hour. We provide the whole noun phrase some sort of event reading only after the determiner’s type presupposition has been integrated into logical form and has been justified by the predicate in the restrictor. The fact that lasted an hour requires for successful coercion a quantized event leads to uninterpretability when the bound variable has received the type mass from the determiner.

As I’ve said, in general predicates must pass their typing presuppositions onto their arguments. So a VP should pass its typing presuppositions to its subject DP, and a transitive verb should pass its typing requirements to its object DP. So for a transitive verb like hit, the presuppositions of the predicate percolate to the internal (direct object) arguments as well as the external (subject) arguments. This leads to the following lexical entry for hit,

\[(45) \lambda \Phi \lambda \Psi \lambda \pi. \Phi(\pi \ast \text{hit}_1. \pi)(\lambda x: \pi \lambda \pi' \Phi(\pi' \ast \text{hit}_2. \pi)(\lambda y: \pi \lambda \pi' \text{hit}(x, y, \pi'))).\]

The lexical entries chosen determine how presuppositions percolate through the derivation tree, and they predict that presuppositions will be justified typically locally to the argument’s typing context (the \(\pi\) that determines the typing of the argument). For instance, when we combine a determiner and an NP to form a DP that is an argument to another predicate \(\phi\), then \(\phi\) conveys presuppositions to the NP or the DP’s restrictor. If such a presupposition is justified in the restrictor of a DP, it will also be justified in the nuclear scope. The converse, however, does not hold—leading to a notion of “global”, restrictor,—or “local”, nuclear scope, justification. As we shall see, there is a preference for binding like justifications of presuppositions at this “global” level, but this preference can be overridden when it is impossible to justify a presupposition at that site.

We are at long last ready to look at how a standard coercion works in the system. Let’s look at an example where enjoy applies to a DP like many books as in (46). enjoy has the same basic form as hit. In the logical form below, I’ve integrated the internal argument DP already into the logical form, though it is not yet normalized. For book I use the type book, which is not very informative; but to go into the exact nature of objects that have a dual nature would be a whole other chapter.\(^\text{22}\) The type agent is imposed on the

\[\text{22In fact in [Asher, 2011], I take book to have the complex type } p \bullet i. \text{ This type has generated considerable discussion [Pustejovsky, 1995; Asher and Pustejovsky, 2006; Asher, 2011; Luo, 2010] and has different metaphysical implications from the types used in coercion. But I won’t go into the details of this type constructor here.}\]
variable associated with the DP \( \Phi \), while \( ag \) is a function picking out the agent of an eventuality.

(46) George enjoyed many books.

Constructing a logical form for the DP and applying it to the entry for \( enjoy \) gives us:

\[
\lambda \Phi. \lambda \pi (\pi \ast \text{agent})[\lambda v. \lambda Q \text{many}(x) (\text{book}(x, \pi \ast \text{arg}_2^{\text{enjoy}}): \text{event} - \varepsilon(\text{hd}(\Phi), \\
\text{book} \sqcap \text{ct} ) * \text{arg}_1^{\text{book}}: \text{book} \sqcap \text{ct} ), Q(\pi)(x) (\lambda y_1, \lambda \pi_1 (\text{enjoy}(v, y_1, \pi_1) \\
\land ag(y_1) = v(\pi_1)))]
\]

A little more commentary is required to understand the type associated with \( enjoy \) in (47). The verb requires of its internal argument that it be of type \( \text{event} \), and in fact it must be an event in which the subject of the verb can participate. However, it also allows us to postulate an eventualty whose type is determined by \( \text{hd}(\Phi) \), which finds the most specific type of the variable bound by the DP and by the type of its actual, syntactically given argument, which in this case is book—which is what is meant by \(-\varepsilon(\text{hd}(\Phi), \text{book} \sqcap \text{ct})\).

Let us assume that such type presuppositions prefer a local justification, near the verb.

Abbreviating our type constraints on \( x \) and \( y_1 \), we get:

\[
\lambda \Phi. \lambda \pi (\pi \ast \text{agent})(\lambda v. \lambda Q \text{many}(x) (\text{book}(x, \pi) , Q(\pi \ast \text{evt} - \varepsilon(\text{hd}(\Phi), \text{book} \\
\sqcap \text{ct} ))(x))(\lambda y_1, \lambda \pi_1 (\text{enjoy}(v, y_1, \pi_1) \\
\land ag(y_1) = v(\pi_1)))]
\]

Continuing the reduction, we get:

\[
\lambda \Phi. \lambda \pi (\pi \ast \text{agent})(\lambda v. \lambda Q \text{many}(x) (\text{book}(x, \pi) , \text{enjoy}(v, x, \pi \ast \text{evt} - \varepsilon(\text{hd}(\Phi), \text{book} \\
\sqcap \text{ct} )))) (\lambda y_1, \lambda \pi_1 (\text{enjoy}(v, y_1, \pi_1) \\
\land ag(y_1) = v(\pi_1))]
\]

The type presuppositions in the nuclear scope of the quantifier cannot be satisfied as they stand. But this particular verb licenses a transformation of the predicational context.

This transformation introduces a functor over types (and which gives distinct output types for distinct input types). It is thus a "polymorphic" type functor. This functor will apply to the \( \lambda \) abstract in the consequent given by the verb, \( \lambda y_1, \lambda \pi_1 (\text{enjoy}(v, y_1, \pi_1) \land ag(y_1) = v(\pi_1)) \). For type presuppositions, this is a general procedure for presupposition justification. The functor introduces a predicate related to the polymorphic type. For example if the polymorphic type maps \( \text{cigarette} \) to an event of type \( \text{smoke(agent, cigarette)} \), then the predicate \( \text{smoke}(e, x, y) \) will be integrated into logical form. When the polymorphic type is underspecified and of the form \( \varepsilon(\alpha, \beta) \) we use the predicate \( \phi_{\varepsilon(\alpha, \beta)}(e, x, y) \). The functor instantiated for this example looks like this:

\[
\lambda \Phi. \lambda \pi (\pi \ast \text{agent})(\lambda v. \lambda Q \text{many}(x) (\text{book}(x, \pi) , \exists z_1: \varepsilon(\alpha, \beta, \pi')(z) \land \\
\phi_{\varepsilon(\alpha, \beta, \text{book} \sqcap \text{ct} )}(z, z_1, u, \pi'))]
\]

Applying the functor on the designated \( \lambda \) term within (49) and using the rules of the \( \lambda \) calculus together with Binding, we get:

\[
\lambda \Phi. \lambda \pi (\pi \ast \text{agent})(\lambda v. \lambda Q \text{many}(x) (\text{book}(x, \pi) , \exists z_1: \varepsilon(\alpha, \beta, \pi')(z) \land \\
ag(z) = v \land \phi_{\varepsilon(\alpha, \beta, \text{book} \sqcap \text{ct} )}(z, z_1, x, \pi'))]
\]

We can now integrate the subject into (51) and exploit the fact that \( ag \) is a function to get the finished result:

\[
\lambda \pi. \exists y (y = g(\pi) \land \text{many}(x) (\text{book}(x)(\pi), \exists z: \varepsilon(\alpha, \beta, \pi')(y, z) \land \\
ag(z) = y \land \phi_{\varepsilon(\alpha, \beta, \text{book} \sqcap \text{ct} )}(z, y, x, \pi'))]
\]

The type of functor in (50), which I call the \( \varepsilon \) functor, suffices to handle all cases of event coercion with verbs whose type presuppositions are sensitive to both the type of the subject and object. \( \varepsilon \) functor is licensed whenever the given argument has a type \( \beta \) that
is inconsistent with the type presupposition of the predicate but the predicate allows the introduction of a polymorphic type like $\epsilon$ that takes $\beta$ as an argument and that satisfies the type presupposition of the predicate. We can generalize our accommodation strategy for event coercion to arbitrary type coercions in a natural way. It is not known at present what is the exact class of transformations a language licenses and whether these transformations are universal or not. The transformation from entities to events in which they participate does seem to be part of the semantic baggage of most languages, however.

Why should such transfer principles and type shift from objects to eventualities be sound? The answer has to do with the presuppositions of the particular words that allow for this morphism, like, e.g., the aspectual verbs and *enjoy*. Enjoying a thing, for instance, presupposes having interacted in some way with the thing, and that interaction is an event. Similarly, one can’t finish an object unless one is involved in some activity with that object, whether it be creating it or engaging in some other activity towards it. That is why such transformations are lexically based; it is the lexical semantics of the words that license the coercion and that makes the rules sound. On the other hand, an object’s starting doesn’t have any such presupposition, and so the prediction is that (53a) should sound much worse than (53b), which it does:

(53) a. The book starts at 10am.
   b. The reading of the book (the book reading) starts at 10am.

Curiously for philosophers interested in natural language metaphysics, we don’t have nominal or verbal coercion licensing constructions with arguments of physical type and with type presuppositions for some abstract type (or vice versa). So there seems to be no coercion able to save our noun phrase *heavy number* and least in its literal reading.23

The verb *enjoy*, however, doesn’t specify what that event is. The event could be just looking at the object as in *enjoy the garden* or perhaps some other activity. So semantics gets us only so far. We now need to specify the underspecified formula $\phi_{\epsilon(HUMAN,BOOK)}$, associated with the type $\epsilon(HUMAN,BOOK)$. This we can do by adding to the type system axioms like this that defeasibly specify underspecified types. These may be considered an extended part of the lexicon.

- $(\alpha \subseteq HUMAN \land \beta \subseteq BOOK) > \epsilon(\alpha,\beta) = READ(\alpha,\beta)$.
- $(\alpha \subseteq AUTHOR \land \beta \subseteq BOOK) > \epsilon(\alpha,\beta) = WRITE(\alpha,\beta)$.
- $(\alpha \subseteq GOAT \land \beta \subseteq BOOK) > \epsilon(\alpha,\beta) = EAT(\alpha,\beta)$.
- $(\alpha \subseteq JANITOR \land \beta \subseteq P) > \epsilon(\alpha,\beta) = CLEAN(\alpha,\beta)$.

Let’s now go back to the difficult coordination examples. Consider again (27a):

(27a) Julie enjoyed a book and watching a movie.

23There is of course the somewhat vernacular American English expression *heavy number*, which might be used, say, to describe a song. In this case, however, *number* picks out a particular track on an album, not an abstract object, and *heavy* doesn’t refer to a physical magnitude but rather to things having strong emotional import. Some objects, namely those with the complex type $P \bullet I$ encode something like a realization relation between abstract objects and concrete ones, but this is something particular to particular kinds of things, not a general principle.
Example (27a) poses some difficult questions for any approach. It turns out that my account predicts that examples like (27a) in which a plural DP is formed through coordination requires an inherently distributive approach. To see this, let’s consider the type of a conjoined DP. Some conjoined DPs have a natural interpretation as a group forming operation ([Krifka, 1991] was the one of the first to point this out in formal semantics): for instance, John and Mary in

(54) John and Mary lifted the sofa.

has a salient interpretation according to which John and Mary as a group lifted the sofa. So the coordinated DPs should shift us from a generalized quantifier over individual objects to a generalized quantifier over plural objects, assuming that plural entities receive a distinct type in the theory. The plural type should be a pluralization of a common type that the two constituent DPs are defined on. In this case the two constituent DPs have the type person or the usual type raised version of this, and the plural DP shifts to a quantifier over the type pluralization of person, persons.

The difficulty with (27a) is that the conjuncts involve distinct types of individuals, whose least upper bound is the type $e$. In (27a), the gerund is a nominalization of a VP and denotes either an abstract entity (and thus is a subtype of $i$ or an eventuality. book on the other hand is neither of abstract type nor of eventuality type. Suppose we model coordination as operating over the plural sum and we take the join of the two constituent DP types as the basis of the pluralization. As the join of the types of the constituent DPs in (27a) is $e$, we can now simply accommodate enjoy’s eventuality type presupposition for its internal argument. But in this case there is no coercion at all and we get a bizarre reading for the example.

Such coordinations between objects of distinct and incompatible types are commonplace:

(55) a. John and the car arrived on time.

b. John and the winch lifted the piano.

(55a) means that John arrived with the car or they arrived separately, but not that John and the car arrived as a group. Similarly, (55b) doesn’t mean that John and the winch as a group lifted the piano but rather John lifted the piano using the winch. If we interpret coordinated DPs of unlike type distributively or via a manner interpretation as in (55), we can then use coercion where appropriate and get the salient readings. That is, for (27a), we get that Julie enjoyed reading a book and watching a movie.

This approach to meaning shifts is very powerful. Many other coercions fall under the general analysis proposed here. The sort of functors appealed to in Percus (2010) to account for concealed questions, for example, are straightforwardly implemented in TCL: ask or debate subtypes for a question in its theme argument but licenses a meaning shift from certain relational nouns to questions involving them. The same sensitivity to the actual word is also observed; just as start in its intransitive use doesn’t license the same polymorphic type and natural transformation as enjoy, so too wonder, which also subcategorizes for a question, doesn’t license the natural transformation from DPs to questions.

---

24This is a feature of $\bullet$ types, of which book is a subtype, something which [Asher, 2011] discusses at length.
6 MODALITY, ASPECT AND THE VERBAL COMPLEX

In this section, I briefly survey some issues about meaning shifts involving the verbal complex, the interpretation of the VP and its projections that include tense, modality and aspect. This is a rich area of study for linguists and which has philosophical implications. Many linguists, including [Dowty, 1979; Verkuyl, 1993; Smith, 1997] inter alia, have observed that certain meaning shifts occur to the type of object denoted by the verbal complex when aspect and tense are applied. Since Vendler’s work in the 1950s [Vendler, 1967], it has been customary to distinguish between different types of denotations of verbs and their projections. Vendler and most of those following him (except for Dowty) have talked of these denotations as types of eventualities, which include events of different types and states. Combined with [Davidson, 1968/69]’s treatment of action sentences, this has led to the received view of verbal modification by various projections, according to which verbal modification involves predication of an eventuality introduced by the verb (and bound by tense). If the modification applies at a node above Tense, for example with an adverb like allegedly or probably, then the eventuality is no longer available as an argument and so such modifications are customarily treated as modifications of some sort of abstract entity like an intension. On the other hand, nominal modification is typically thought to be much more heterogeneous, depending on whether the modifier is intersective, subsective or non-subsective.

Davidsonian event semantics is by and large quite successful. But there are cracks in the edifice. Davidsonian and Neo-Davidsonian semantics have a neat way of explaining verbal modification that crucially involves events. Basically, verbal modification via syntactic adjunction becomes a simple matter of applying the property provided by the modifier to an event variable that also serves as the event argument of the verbal complex:

- VP : VP MOD → λe(||MOD||e) ∧ ||VP||e).

However, there is some reason to treat some modifiers more like arguments of the verbal complex. That is the strategy I adopt in [Asher, 2011]. The simple strategy for interpreting adjunction would suggest that one can simply pile on more adverbials of the same basic type. Some modifiers resemble arguments more than adjuncts; a verb cannot have more than one modifier of a given type.25 Consider for example,

(56)  
   a. #Brutus killed Caesar with a knife with a hammer.  
   b. #Brutus killed Caesar with a knife with a sharp knife (no pause).  
   c. Brutus killed Caesar with a knife and a hammer.

With a knife and with a hammer are modifiers that fill in an instrumental role of the verb. One can coordinate instrumental modifiers as in (56c) and analyze these modifications using the mechanisms of copredication developed in TCL. But one cannot simply add them ad libidem. This is contrary to what a standard Neo-Davidsonian analysis would have us expect.

If we treat such modifiers as optional arguments of the verb, we get a better analysis of the data. I will suppose that a verb like kill takes an optional instrumental just in the way

25[Beaver and Condoravdi, 2007] make this point. It is worth noting that Montague’s syntactic treatment of modifiers also gets these facts wrong.
that a verb like \textit{wipe} takes an optional direct object, according to Kratzer and Levin. Such a semantic analysis is compatible with the standard syntactic analysis of PP modification as VP adjunction. But it could also point to a much more complex syntactic structure than usual (see, however, [Cinque, 1999]). Once that instrumental argument is filled by an explicit instrumental modifier, it cannot be filled again. This is what the $\lambda$ calculus derivation predicts. Similar observations hold for modifiers that provide a recipient role for a verbal complex:

\begin{enumerate}
\item John loaded the hay on the wagon on the train.
\item On the train, John loaded the hay on the wagon.
\item John loaded the hay on the wagon and on the train.
\item John loaded the wagon with the hay with the flour.
\item \#John wrote a letter to Mary to Sue.
\item John wrote a letter to Mary and to Sue.
\end{enumerate}

In (57a) \textit{on the train} does not describe a recipient role of the loading and hence not a modifier of the verb \textit{load} but is a modifier of the noun \textit{wagon}. Fronting this PP makes it a modifier of the verb but it furnishes a location not a recipient. The only way to have the wagon and the train both be recipients is to use coordination. A similar moral holds for the \textit{write} and its recipient role in (57e-f) and for the \textquotedblleft contents role\textquotedblright of \textit{load} type verbs in (57d). Neo-Davidsonian approaches have no explanation for these observations. If these PPs saturate optional arguments, then we have a ready made explanation of these facts. With regard to (57e) we have an infelicitous sentence: \textit{Mary to Sue} does not have the type required by \textit{write} (it should be \textit{agent} for its indirect argument), and so the only way to understand (57e) is that we are trying to saturate one optional argument twice, which we can\textquoteright t do in the $\lambda$ calculus.

Not all verbal modifiers saturate optional arguments. Some modifiers simply take the VP as an argument, as in Montague Grammar. Temporal and locative modifiers seem to fall into this class. We can have several temporal modifiers that simply narrow down the time at which the event described by the verbal complex took place. Locative modifiers work similarly.

\begin{enumerate}
\item On Monday Isabel talked for two hours in the afternoon between 2 and 4.
\item In Paris John smoked a cigarette on the train in the last second class compartment in seat number 27.
\end{enumerate}

The fact that temporal and perhaps other modifiers take the VP itself as an argument and are not arguments of the VP makes predictions in TCL. Predicates pass their type presuppositions onto their arguments in TCL, not the other way around. So TCL predicts that temporal adverbials can affect the type of the verbal complex, as seen in the examples (59). The temporal modifiers can change the aspect of the verbal complex from an achievement or accomplishment in (59a,c) to an activity in (59b,d).

\begin{enumerate}
\item John wrote a letter in an hour.
\item John wrote a letter for an hour.
\item John kissed Mary at 10 in the morning.
\item John kissed Mary for an hour.
\end{enumerate}
Furthermore, TCL predicts that temporal modifiers should not be affected by the type of the verb or the verbal complex.

In examining verbal modification in TCL, I have been speaking of modification of the verbal complex. But what is that, semantically and type-theoretically? With Neo-Davidsonians, we could stipulate that the verb projects an event variable to the tense and aspect projections. But some sentences intuitively don’t denote anything like an eventuality. Whatever is bound by the tense projection of the verbal complex is not an event or a state, at least if we take states and events to have some sort of spatio-temporal location (and if we do not, it’s unclear why we should call such entities states or events in the first place). Consider

(60) Two and two make four.
What (60) describes is not a state of the concrete physical world but rather a fact or a true proposition, a collection of possible worlds that contains the world of evaluation. It is important to note that such facts also have temporal modifications.

(61) Two and two make four, and two and two will always make four.
In addition, there are problems with the compositional picture on the event denotation view (discussed in [Asher, 1993]).

(62) No one danced at the party.
Intuitively this describes a fact too—the fact that there was no event of dancing at the party. Once again temporal modifications of such facts are commonplace:

(63) No one danced at the party for over two hours.
A natural move for Davidsonians to make in the case of negated event sentences like (62), or similar sentences containing quantifiers, is to claim that its denotation is a state, a property of the party that holds at a certain space-time region. But such a move is not really very palatable for (60), since no delimited space-time region is picked out of which a property is predicated. Some might object that facts aren’t temporally or spatially located. However, the data in this regard are relatively unambiguous, at least with the nominal fact. Facts are sensitive to time, as is the truth of propositions. If a fact is analyzed as a true proposition with de re characterization, then this is to be expected.

(64) a. For two years, it was a fact that you could cross the border without a passport. Now that’s no longer the case.
    b. Now suddenly it was a fact. An edict appeared offering amounts that descended as the rank descended: $5000 for a man-of-war’s captain;... (The Opium War 1840-1852).
    c. Suddenly, it was a fact of life. Like it or not, you have to go along. (New York Magazine, August 1989).
    d. She reached for a sudden fact. "It’s the largest town in England without a university." (Updike)
    e. As a result of the persecution, both state-sponsored and unofficial anti-Semitism became deeply ingrained in the society and remained a fact for years. (Wikipedia)
    f. post say QF dont recruit direct CSM’s yet 1 post (as a joke) says that they do the EVERYONE runs with it and its all of a sudden fact. ... (Quantas
What is perhaps more surprising is that facts can be spatially localized.

(65)  
   a. In Berkeley it’s a fact that you can get arrested for having a cigarette in a public place, but not in New York City.
   b. In Topeka but not in Greenwich, it is true that people go to church on Thursdays = It is true that people go to church on Thursdays in Topeka but not in Greenwich. ?= It is true in Topeka but not in Greenwich that people go to church on Thursdays.
   c. Everything depends on carefully establishing what, exactly, the facts in Georgia are. (‘Establish the facts in Georgia First’, www.theatlanticright.com).

Some of these examples, but not all, invite an analysis according to which the temporal or spatial modifications become part of the fact described; e.g., *in Berkeley it’s a fact that...* can be reanalyzed as *it is a fact that in Berkeley...* However, this strategy won’t work for the explicitly quantificational (65c). We can thus simply take these contents to demand that their realizers be facts like (60).

In TCL the verbal complex is a subtype of prop; the specific type is determined by the appropriate instance of the polymorphic type with its type parameters specified by the verb’s arguments. Thus, the type of an intransitive verb in (66) is a generalized function from subtypes of a DP type and a presuppositional context type to a subtype of prop. This is a refinement of the type in (66b) one might typically assign to VPs.

(66)  
   a. \( \exists x \subseteq \text{dp} (x \Rightarrow \Pi \Rightarrow iv(\text{hd}(x))) \).
   b. \( \text{dp} \Rightarrow \Pi \Rightarrow t \).

A similar story holds for transitive and ditransitive verbs.

The “tail” or the value of the verb type in (66a) may take on different fine-grained values for different values of its parameters; it is also subject to modification by operators that take the verbal complex in its scope. The strategy is to let various adverbials, tense and other modifiers modify this proposition; some modifiers force the introduction of a realizer of the propositional content, thus implementing Davidson’s intuition in a higher order setting. However, we can be rather agnostic as to what this realizer is. For simple action sentences, for example, modification by manner adverbials produces a realizing eventuality for the content given by the verbal complex. But for verbal complexes modified by negation, tense, or the presence of a locating adverbial like that in (62) may introduce a realizer that is a fact. Similarly, if the type of the verbal complex expresses simply a relation between informational objects as in (60), temporal adverbs or tense force a coercion that introduces a realizer of the content that must be a fact. This would predict that the temporal modification of (61) means something like it will always be true that 2 and 2 makes 4. Thus, most modifications involve a coercion from the verbal complex’s internal semantic value, which is a subtype of prop to an event or fact realizer.

Let’s now take a look at a derivation with a simple verbal modifier that acts semantically as an optional argument. Consider the verb phrase

(67) hit Bill with a hammer.

*With a hammer* is an instrumental, which is a kind of verbal modifier. The type specification logic contains axioms of the following form, where \( \text{ty}^+(\text{tv})(x, y) \) is the value of the
most specific instantiation of the polymorphic type of the transitive verb when applied to
type parameters $x$ and $y$:

$$(68) \text{ for } y, x \subseteq p, \text{ with}(\text{ty}^+(\text{tv}(x, y)), \text{hammer}) \subseteq \text{instrument}(\text{ty}^+(\text{tv}(x, y)), \text{hammer}).$$

This axiom suggests the proper formula with which to combine the modifier:

$$(69) \lambda u \lambda \pi' \exists x (\text{hammer}(x) \land \text{instrument}(x, u, \pi' * u: \text{evt}))$$

Similarly to the way TCL models modifiers for nouns, I add higher order arguments to
eventive verbal entries that are typed $\text{instrumental}$, $\text{manner}$, etc., which are all subtypes
of 1; the instrumental modifier is depicted in (70). If the instrumental argument isn’t
realized its $\lambda$ abstracted variable is applied to the identity property and no realizer is
involved. However, a non-empty entry for the modifier as in (69) instrumental argument of
the verbal complex forces the introduction of an eventuality realizing the verbal complex.
This features another use of EC, or event coercion, but this time from $\text{prop}$ to $\text{evt}$.

$$(70) \lambda \Phi \lambda P: \text{instrumental} \lambda \Psi \lambda \pi (\text{realizes}(w, \text{hit}(x, y, \pi')) \land P(w, \pi * \text{realize}: \text{evt})).$$

After constructing the VP using the coerced (70) and integrating the entry for the DP $Bill$,
we combine the modifier from (69) and allow tense to bind the resulting variable $w$.

$$(71) \lambda \Psi \lambda \pi \exists w: \text{evt} \exists t < \text{now} \left( \text{holds}(w, t) \land \text{realizes}(w, \text{hit}(x, b, \pi')); \exists \Phi (\lambda x. \lambda \pi'' \Phi (\text{realize}: \text{evt}))) \right).$$

This approach has several pleasing consequences. It predicts that two separate instrumental-
phrases cannot combine with a VP because the VP will have only one lambda
abstract for instrumentals; once that argument is saturated, we cannot integrate another
instrumental with that verbal complex. Second, it validates the Davidsonian simplifica-
tion inferences, if we assume, as we did for nominal modifiers, that empty verbal modifier
phrases are interpreted as the identity property. Third, it predicts that a verbal complex
may modify the type of an instrumental by passing to it type presuppositions. Some evi-
dence for this is the observation that different verbs lead to different interpretations of the
instrumental:

$$(72) \text{ paint a miniature with a brush.}$$

$$(73) \text{ scrub the floor with a brush.}$$

TCL also predicts that certain eventuality types may be derived from others. For in-
stance, $\text{walk}$ is an activity but $\text{walk to the store}$ with a goal PP is an accomplishment. The
system predicts that accomplishments should consist of an activity together with a
natural endpoint or telos (given by the goal PP). The value of the polymorphic type is an
information content but various modifiers can coerce this to an eventuality. In this way,
TCL takes a middle course between the views of Davidson and Montague on adverbial
modification.

On the other hand, verbal temporal modifiers take the whole VP as an argument, and
so can cause a local type justification of the verbal complex. $\text{For an hour}$ takes a VP like
$\text{hit Bill}$ as an argument and imposes the type presupposition that the variable of complex
and polymorphic type that it modifies must be of type $\text{activity}$. We now have a case of

---

$^{26}$\text{instrumental} is defined as the type $\text{evt} \Rightarrow \exists x \text{instrument}(\text{evt}, x)$. If we like, we can suppose that it is the
syntactic operation of adding an Instrument Phrase that introduces the additional $\lambda$ abstract over properties.
coercion since $hit(p, p)$ is a subtype of $\text{achievement}$. A version of EC licenses an accommodation of the $\text{activity}$ type presupposition by inserting an iteration operator over the VP, yielding an appropriate interpretation of $hit$ $Bill$ $for$ $an$ $hour$. TCL also predicts that temporal modifiers may lead to fine-grained shifts in meaning in the verbal complex. For example, consider:

(74)  

(a) She left her husband 5 minutes ago.

(b) She left her husband two years ago.

In (74b), we have a very different sense of $leave$ than in (74a). (74a) interprets $her$ $husband$ in a physical or locational sense whereas (74b) interprets $her$ $husband$ in a more institutional sense; that is, (74b) means that the subject has left her marriage. Conditional type constraints in the type specification logic can model the effects of the temporal adverbials on the predication.

Finally, this approach predicts temporal and spatial PP modification to be possible for all verb complexes, including sentences with negation on the VP or monotone decreasing quantifiers like $few$ $people$ in subject position. Temporal adverbials outside the scope of negation coerce the introduction of a fact that realizes $\neg \phi$, as does the application of Tense. Hence, TCL delivers a uniform account of temporal modification as predicates of realizers, unlike Davidsonian approaches.

The effects of adverbs, modals and aspect on the type of the verbal complex are examples of meaning shifts governed by grammatical means. TCL countenances finegrained types not only for words but also, thanks to the use of polymorphic types, for more complex expressions even clauses; and these types for the verbal complex need not all be eventualities. In fact, a more uniform approach is to take the type of the verbal complex when saturated with its syntactically given arguments to be a subtype of $\text{prop}$. Gerunds and other nominalizations provide a realizer of the general type $e$ of the material underneath its scope together with information about the temporal and or modal location of the realizer. The type of the realizer is polymorphic upon the fine grained type of its argument as well as on parameters of evaluation. Thus, $\text{realizer}$ is another example of a polymorphic type. In effect, it is unnecessary and incorrect to introduce Davidsonian event arguments into the lexical entries for verbs generally; stative sentences simply have propositional denotations (subtypes of $\text{prop}$) that are true at times and worlds. When needed as in nominalization, we can isolate an associated spatio temporal region via the realizer. This minimizes event promiscuity in one’s ontology and also offers a solution to the nasty problem of eventuality projection across quantifiers and operators like negation or modals as well as eventualities of such timeless sentences as $2 + 2 = 4$. Since there aren’t events, there isn’t any problem about projecting them through the logical structure of the asserted content.

6.1 Aspectual coercion

With this sketch of TCL’s view of verbal modification, let us return to aspectual coercion in the examples (6), one of which I repeat below.

(6b)  

John is being silly.
Because the types of the verbal complexes can be quite fine grained, we can distinguish between verbal complex types whose realizers are facts, those whose realizers are events (here I have in mind paradigmatic event sentences like *John kissed Mary*), and verbal complex types whose realizers are facts or states but have a close connection with eventualities and activities. In this last class fall statives that accept progressivization, copular sentences that involve some stage level predicate like *is silly, is stupid, is an asshole,...* 

(75) illustrate that these have a natural link to certain activities:

(75) a. John was ridiculous to insist on fighting that guy.

b. John was stupid/insane/silly to give his money to that woman.

c. John was an asshole in being so rude to that student.

This construction doesn’t work with other stative predications like

(76) #John knew French to give that speech/in making that speech.

These constructions indicate copular predications with stative adjectives form a particular subtype of *prop*. While these predications are usually classified as statives by the usual tests of adverbial modification, they are special in that they have a very tight connection with activities of which they describe the result. This subtype is an argument to the progressive and then produces a particular kind of realizer after tense is applied. What the progressive does ([Dowty, 1979; Asher, 1992] *inter alia*) is introduce a functor describing some process that leads at least in the normal instances to the appropriate realizer of the proposition given by the verbal complex. While the progressive does not apply to what one might call “pure statives” like *John know French*, the progressivization of this special subclass of statives introduces an eventuality “realizer” of the verbal complex given by a VP produced from a copula with adjective complement.

When it combines with an adjective, the copula passes the presuppositions of the predicate to to its DP argument. We could assume *John is silly* has a perfective or completed aspect; this forces the introduction of a realizer. This realizer is of type *state* because of the presence of the copula which affects the fine grained type of the verbal complex. I give the aspectual operator’s contribution first and then the end result. Below $P$ is the type of the adjectival VP and $\Phi$ as usual is of type $dp$.

(77) a. $\lambda P.\Phi \lambda \exists z: state realizes(z, \{\Phi(Pre(\arg P)(\pi))(P(\pi))\}).$

b. $\lambda \pi.\exists z: state realizes(z, \{silly(j, \pi)\}).$

Alternatively, we can follow the lead of Slavic languages in which there is no verb form for such sentences at all (and hence no aspect) and not introduce any realizer at all. This yields a very simple form for *John is silly*:

(78) $\lambda \pi silly(j, \pi)$.

Now let us turn our attention to (6b). After constructing the logical form of the VP, we apply the progressive operator in Aspect. The progressive aspect also introduces a realizer but it must be an event type that is non-stative. So it demands a realizer that is an activity. At this point local justification is attempted by introducing a realizing eventuality for the verbal complex. A coercion takes place when the aspectual information combines with the verbal complex, prior to Tense, but here the coercion is more complex. The verbal complex still requires that any realizer be stative (it is a type presupposition of the verbal complex itself), so we need Aspect, together with the fine grained type of the verbal
complex, which reflects the copula + adjective construction, to license a polymorphic type of the form $\text{activity}(\sigma, \alpha)$ whose parameters are $\sigma \subseteq \text{state}$ and the bearer of the state. The output or value of the polymorphic type is a type of activity or process involving an object of type $\alpha$ that results in a state of type $\sigma$. The associated functor for this polymorphic type is:

$$\lambda \pi \exists e : \text{activity}(e \circ \text{now}) \land \exists s (\phi(j, e) \land \text{result}(e, s, \pi) \land \text{realizes}(s, ^{\wedge} \text{silly}(j, \pi))).$$

We now use a version of EC to justify the progressive’s type presuppositions, and we get the following meaning for (6b):

$$\lambda \pi \exists e : \text{activity}(e \circ \text{now}) \land \exists s (\phi(j, e) \land \text{result}(e, s) \land \text{realizes}(s, ^{\wedge} \text{silly}(j, \pi))).$$

In words this says that John is doing some activity whose result state is $s$ and $s$ includes the temporal span of some aspect of John in which he is silly. The assumption that there is no aspect in $\text{John is silly}$ leads to essentially the same logical form, but this time we have a direct coercion to the result state interpretation from the propositional content $\land \text{silly}(j, \pi)$. These are the intuitively right truth conditions for such a sentence. Our discussion has shown how aspectual coercion falls within the TCL approach to coercion.\textsuperscript{27} This discussion also shows us what is behind the polymorphic types that we used for simple event coercion; they are introducers of event realizers for an underspecified verbal complex.

### 6.2 Modals and aspect

Let’s now turn to the interaction of modals and aspect for another illustration of meaning shifts. Once again we will be interested in the contributions of aspect, but in order to understand how these contributions interact with the semantics of modals, we will have to use a more expressive framework, that of TY2, in which world and time evaluation variables become explicit parameters in the logical form and, accordingly, the type system countenances atomic types for worlds and times.

What, first, is the position of aspect with respect to modality? Which takes which as an argument? The answer to these questions seems to depend, as Hacquard and others have suggested, on which modality we are interested in. Consider first epistemic modals like $\text{might}$ in English. Let’s assume that the verbal complexes in the first sentences of (81) license an event realizer. The mechanisms that discourse linguists have used to analyze modal subordination [Roberts, 1989] then predict a substantive difference between (81a) and (81b), which is born out.

$$\begin{align*}
\text{a. } & \text{John might run the Marathon tomorrow. It would take him at most 3 hours.} \\
\text{b. } & \text{John might run the Marathon tomorrow. It will take him at most 3 hours.}
\end{align*}$$

The second example is worse than the first, and this is what those familiar with examples modal subordination should expect. Assuming as is reasonable that the running of the Marathon is under the scope of $\text{might}$, modal subordination mechanisms predict that the pronoun under the scope $\text{would}$ can be linked to material under the scope of another

\textsuperscript{27}See [de Swart, 1998] and [Bary, 2009] for a more extensive discussion of uses of coercion to describe different uses of the passé simple and imparfait in discourse. As far as I can tell, all of these coercions are of a piece with the story for aspectual coercion that I have spelled out here.
modality like *might*, but that such material is not accessible to a pronoun that is not within the scope of a modal, as in (81b). The only difference between these examples and the classic examples motivating modal subordination is that the pronoun in the second clause here links to an event.

There is an additional question from the perspective of event anaphora as to whether the modality itself introduces some sort of eventuality. There is considerable evidence that modalities don’t introduce states, at least epistemic modalities. If we suppose that a state is introduced by the epistemic modal, we should be able to temporally modify it anaphorically. But that isn’t possible:

(82) #John might run the Marathon. That will last for a couple of years. (Where *that* should pick up the possibility of John’s running the Marathon.)

Adverbial modification also provides evidence that modals don’t introduce states. For instance, *for* adverbials, which provide one test for statehood of the (realization of) the verbal complex, are infelicitous with epistemic modals.

(83) a. John was sick for two weeks.
   b. John was sick at 2 pm.
   c. # John might finish his dissertation for two years.
   d. # John might finish his dissertation at 2.

(83c) sounds really bad to me, but conceptually it should make sense. The last example sounds fine but the adverbial modifies the VP under the scope of the modal, not a state introduced by the modal itself.

Spatial and temporal modifiers that hold of states also don’t seem to modify epistemic modals:

(84) a. John was sick at work.
   b. John might finish his dissertation at Jean Nicod.

The last example is fine but it doesn’t modify a state given by the epistemic modal but rather the event described by *finish*. These examples show that the data used to motivate the introduction of eventualities in classic action sentences doesn’t hold up for epistemic modals. The fact that it’s also difficult to pick these up anaphorically suggests that perhaps they aren’t there.28 Temporal and spatial modifiers of epistemic modals follow the same analysis as the temporal and spatial modifications of other facts: they contribute parame-

---

28Complicating this picture, however, is the observation that epistemic possibilities can shift with time, and thus are in some sense temporally located:

(85) Two years ago, we might have taken that option, but not now.

(86) Suddenly might we not need Google for much of our web browsing? (Benjamin Cohen on Technology 22 April 2010)

(87) Kendrick Meek suddenly might have a shot. (Atlantic Wire, May 11, 2010)

(88) And in his palm he might hold this flower, examining the golden dainty cup, and in him suddenly might come a sweetness keen as pain. Carson McCullers *Ballad of the sad cafe*

These examples show that epistemic modals can get situated in space or time, but not by VP adjoining adverbs only IP adverbs. Also they’re difficult to modify with ordinary tense in English. This shows, I think, with Homer that we can take epistemic modalities to have wide scope over tense and aspect, and over some higher projection of VP, and that the realizer of a modal statement is a fact.
ters of realization to facts. It looks then as though epistemic modals take very wide scope; all temporal and aspectual modification takes place within the scope of the modality.

The facts are quite different for ability modals. (the following examples are due to V. Homer p.c.).

(89)  
  a. Hier, Jean devait rendre son devoir demain, mais les choses ont changé : il doit maintenant rendre son devoir aujourd'hui.
  b. Pendant des semaines, Jean a dû rendre son devoir demain, mais les choses ont changé : il doit rendre son devoir aujourd'hui.
  c. Il n'y a qu'en France que les gens peuvent aller adopter un enfant au Mali, c'est interdit partout ailleurs.

The data show that spatial and temporal VP modifiers can clearly modify an ability modal claim. So this would suggest that ability modals are much closer to the root verb position and so would fall within the scope of tense and aspect.

Aspect is traditionally understood to bind the event variable introduced by a verb phrase. But perfective aspect in many languages takes on an evidential function, which has not to do with events but with propositions Faller (2002). I provide a framework in which this is natural. Aspect can bind a parameter in the modality or an eventuality, if one is coerced by, say, verbal modification. Note that in TY2, temporal modifiers automatically attach via a Davidson like rule to the time parameter, thus not requiring the introduction of any eventualities on that score. Perhaps one might also countenance, a space-time parameter directly in TY2 to account for spatial modifiers as well.

Given the type of worlds $s$ and the type of times $t$ as basic types, we have also have types of the sort $s \rightarrow s$, which is the type of a modal transition. This allows us to rewrite basic possibility and necessity modalities as:

- $\lambda w \Diamond \phi(w) := \lambda w : s \rightarrow s \lambda w (\phi(\rightarrow (w)))$.
- $\lambda w \Box \phi(w) := \lambda w : s \rightarrow s \lambda w (\phi(\rightarrow (w))) \land \neg \exists \rightarrow' \neg \phi(\rightarrow' (w))$.

When we are dealing with epistemic modalities, whose contribution does not fall under the effect of tense or aspect, we may assume these $\lambda$ bound variables to be existentially closed off. But when we are dealing with ability modals, aspect will contribute the existential closure. The payoff is that we will avoid the essentialist difficulties of Hacquard’s solution as well as its difficulties with interpretation under negation.

Within TY2, we can be more explicit as to what the realization predicate is than we can in the version of intensional logic used in TCL. Perfective aspect has the following entry where it realizes either an eventuality or a modal transition (of type $s \rightarrow s$) or a proposition. Let’s call the type of such a realizer has the type $\rho$. The contribution to logical form of perfective aspect is then the following:

(90) $\lambda P.t \lambda w \exists x; \rho (f_1(x,w) \subseteq w \land \text{at}(f(x,w),t)) \land P(f_2(x,w),t)$.

In the contribution to logical form, $P$ has the type of functions corresponding to propositions in the version of TY2 I am using here (a function from worlds and times to truth values); $w$ is a variable for worlds while $t$ is a variable for times and $f_1$ and $f_2$ are quasi-projection functions that combine $x$ with $w$, depending on the type of $x$. If $x$ is
an eventuality or fact, \( f_1(x, w) = x \) and \( f_2(x, w) = w \), whereas if \( x \) is a modal transition \( f_1(x, w) = x(w) = f_2(x, w) \).

Forgetting about tense for illustrative purposes, here’s what perceptive aspect does to something like \( (\Diamond) \):

\[
\lambda w \exists \leftarrow (\rightarrow (w) \subseteq w \land \text{take}(\text{train}, j)(\rightarrow (w))).
\]

This is the analysis for \( \text{Jeanne a pu prendre le train} \) (Jean was able to take the train).

This analysis involves no event essentialism because modality and perfective aspect doesn’t have to do with events but rather with realizations, which is a much more general notion. Perfective aspect collapses the modality underneath it in a compositional way. Negation and conditionals work standardly and as predicted. So for example \( \text{Jeanne n’a pas pu prendre le train} \) yields

\[
\lambda w \neg \exists \leftarrow (\rightarrow (w) \subseteq w \land \text{take}(\text{train}, j)(\rightarrow (w))).
\]

If we assume that the domain of \( \rightarrow \) also includes the identity map from \( w \) to \( w \), we get the desired inference.

For imperfective aspect, we have the following entry. It does not force the truth of the sentence under its scope at the actual world.

\[
\lambda P. \lambda t. \lambda w \exists x: \rho \ P(f_2(x, w), t).
\]

This also gives us the right predictions for French imperfective ability modal sentences. The imperfective, however, also seems to coerce the presence of an “inertial worlds” modality \( \text{la} \) [Dowty, 1979] in the absence of an explicit modal to capture the incompleteness of the action, as shown for the simple sentences below:

\[
\lambda w, t \exists t' \text{écrire une lettre}(j, w, t') \land t' < t).
\]

\[
\lambda w, t \exists t' \text{écrire une lettre}(j, w, t') \land t' < t).
\]

\[
\lambda w, t \exists t' \exists \rightarrow_i (\text{écrire une lettre}(j, \rightarrow_i (w), t') \land \neg \exists \rightarrow_i' \neg \phi(\rightarrow_i' (w), t') \land t' < t).
\]

7 DISCOURSE INTRUSIONS REVISITED

I’ve now sketched out how the predicational context can lead to apparent meaning shifts in a variety of domains. But the predicational context given by type presuppositions is only part of the context that affects interpretation. As we have seen in the Introduction, the phenomenon of discourse intrusions shows that discourse context can also affect interpretation. It’s time to revisit these. To do so, I’ll have to say a bit more about Segmental Discourse Representation Theory or SDRT, whose notion of discourse context will be important to the analysis of discourse intrusions.

For our purposes I will need the following features of SDRT and its notion of discourse context.\(^{29}\)

- SDRT’s semantic representations or logical forms for discourse, SDRSs, are recursive structures. A basic SDRS is a labelled logical form for a clause, and a complex

\(^{29}\)For more details, see, e.g., [Asher, 1993; Asher and Lascarides, 2003].
SDRS will involve one or more discourse relation predications on labels, where each label is associated with a constituent, i.e., a perhaps complex SDRS.

- An SDRS for a discourse is constructed incrementally within a logic of information packaging that uses several information sources and that is responsible for the final form of the SDRS. The logic of information packaging, which reasons about the structure of SDRSs, is distinct from the logic of information content, in which we formulate the semantic consequences of an SDRS.

- The rules for inferring discourse relations are typically rules that exploit a weak conditional $\succ$. They form part of the Glue Logic in SDRT, which allows us to “glue” new discourse segments together with discourse relations to elements in the given discourse context. This logic has exactly the same rules as the logic for specifying values for polymorphic types, though the language of types and the language for describing discourse logical forms are distinct. SDRT’s binder rule makes use of the results of the Glue Logic.

- The discourse relations used in SDRT, which have semantic (e.g. spatio-temporal, causal, etc.) effects, are binary and either coordinating (Coord) or subordinating (Subord). Examples of subordinating relations are Elaboration, where the second constituent describes in more detail some aspect of some eventuality or some fact described in the first constituent. Some coordinating relations like Narration (where constituents describe a sequence of events) and Continuation (where linked constituents elaborate simply on some topic) require a topic; i.e., there must be a simple constituent, a common “topic”, that summarizes the two related constituents and that is linked to them via the subordinating Elaboration relation. If this third constituent has not been explicitly given in the previous discourse, it must be “constructed”.

Discourse structure affects the way semantically underspecified elements are resolved. Sometimes the temporal structure of a discourse is more elaborate than what is suggested by a semantic analysis of tenses such as that found in DRT [Kamp and Reyle, 1993]. There are clearly temporal shifts that show that the treatment of tenses cannot simply rely on the superficial order of the sentences in the text. Consider the following discourse (from Lascarides and Asher 1993).\(^{30}\)

\[(98)\quad a. \quad (\pi_1) \text{ John had a great evening last night.} \\
 b. \quad (\pi_2) \text{ He had a great meal.} \\
 c. \quad (\pi_3) \text{ He ate salmon.} \\
 d. \quad (\pi_4) \text{ He devoured lots of cheese.} \\
 e. \quad (\pi_5) \text{ He then won a dancing competition.} \]

\(^{30}\)My apologies for the potential confusion on variables. SDRT uses $\pi, \pi_1, \ldots$ to denote discourse constituents and $\alpha, \beta, \ldots$ function as variables over constituents, while in TCL $\pi$ picks out a presuppositional parameter in the type system and $\alpha$ and $\beta$ range over types. I hope that context will make it clear which uses of these variables is in question.
(98c-d) provides ‘more detail’ about the event in (98b), which itself elaborates on (98a). (98e) continues the elaboration of John’s evening that (98b) started, forming a narrative with it (temporal progression). Clearly, the ordering of events does not follow the order of sentences, but rather obeys the constraints imposed by discourse structure, as shown graphically below. Thus the eventualities that are understood as elaborating on others are temporally subordinate to them, and those events that represent narrative continuity are understood as following each other. The relevant parameter for interpreting tenses is discourse adjacency in the discourse structure, not superficial adjacency. A theory like SDRT [Asher, 1993; Asher and Lascarides, 2003] provides the following discourse structure for (98) and this allows us to get a proper treatment of the tenses therein. Here π₆ and π₇ are discourse constituents created by the process of inferring the discourse structure.³¹ Note that π₁ and π₂ serve as topics for the Narrations holding between π₂ and π₅ and π₃ and π₄.

![SDRT graph for (98)](image)

Figure 1. SDRT graph for (98)

Temporal relations between events introduced by verbs with certain tenses are underspecified in a language like English, and discourse structure is an important clue to resolving this underspecification. SDRT predicts that discourse structure affects many types of semantic underspecification. Nearly two decades of work on ellipsis, pronominal anaphora, and presupposition has provided evidence that this prediction is correct (Asher 1993, Hardt, Busquets and Asher 2001, [Asher and Lascarides, 1998; Asher and Lascarides, 2003]). My hypothesis here is that discourse structure also helps resolve underspecification at the level of types and hence contributes to content in predication.

To see how this comes about, we need to examine discourse coherence and its relation to discourse structure. In SDRT, as in most theories of discourse interpretation, to say that a discourse is (minimally) coherent is to be able to derive a discourse structure for it. Discourse coherence is a scalar phenomenon, however. It can vary in quality. Following

Asher and Lascarides (2003), I say that an sdrs $\tau_1$ is more coherent than an sdrs $\tau_2$ if $\tau_1$ is like $\tau_2$, save that $\tau_1$ features strictly more rhetorical connections. Similarly, $\tau_1$ is more coherent than $\tau_2$ if $\tau_1$ is just like $\tau_2$ save that some underspecified conditions in $\tau_2$ are resolved in $\tau_1$. But for now, let’s focus on the perhaps simplistic position that discourse coherence is maximised by ‘maximising’ the rhetorical connections and minimising the number of underspecified conditions. We can define a principle that will govern decisions about where one should attach new information when there’s a choice. It will also govern decisions about how other forms of underspecification get resolved. And the principle is: the preferred updated sdrs always maximises discourse coherence or MDC (Asher and Lascarides 2003).

The degree-of-coherence relation $\leq$ thus specified is a partial ordering on discourse structures: other things being equal, the discourse structures which are maximal on $\leq$ are the ones with the greatest number of rhetorical connections with the most compelling types of relation, and the fewest number of underspecifications.

MDC is a way of choosing the best among the discourse structures. It’s an optimality constraint over discourse structures that are built via the glue logic axioms. [Asher and Lascarides, 2003] examine in detail how MDC works in picking out the intuitively correct discourse structure for (98), as well as many other examples. We won’t be much concerned here with exactly how discourse relations are inferred, but we will need from time to time to refer back to this background logic.

To get a feel for how MDC works in tandem with underspecification, consider the example from [Asher and Lascarides, 2003], (99):

(99) a. I met an interesting couple yesterday.

b. He works as a lawyer for Common Cause and she is a member of Clinton’s cabinet.

The pronouns he and she introduce underspecified formulas into the logical form for this discourse. They could be bound deictically to salient individuals in the context, but that would not allow us to infer a tight connection between (99a) and (99b). The discourse would lack coherence. On the other hand, if he and she are linked via a “bridging” relation to the DP an interesting couple, then we can infer a strong discourse connection between (99a) and (99b). MDC predicts that this anaphoric interpretation of the two pronouns is preferred because it leads to the preferred discourse structure.

Armed with SDRT’s notion of discourse structure, we can return to the examples with the aspectual verbs. I will use the speech act discourse referents $\pi_0, \pi_1, \pi_2, \ldots$ to isolate the minimal discourse units in these examples.

(100) a. ??Yesterday, Sabrina began with the kitchen ($\pi_1$). She then proceeded to the living room and bedroom ($\pi_2$) and finished up with the bathroom ($\pi_3$).

b. Yesterday Sabrina cleaned her house ($\pi_0$). She began with the kitchen ($\pi_1$). She then proceeded to the living room and bedroom ($\pi_2$) and finished up with the bathroom ($\pi_3$).

c. Last week Sabrina painted her house ($\pi_0$). She started with the kitchen ($\pi_1$). She then proceeded to the living room and bedroom ($\pi_2$) and finished up with the bathroom ($\pi_3$).
Roughly the story for these examples follows that for (99). Consider (100b). \((\pi_1), (\pi_2)\) and \((\pi_3)\) form a narrative sequence that jointly elaborates the information in \((\pi_0)\). Elaborations require that the events inferred via coercion from the aspectual verbs must all be part of the cleaning of the house. Of course, coercion underspecifies what the events involving the kitchen, living room and bedroom and the bathroom are; but the presence of the Elaboration and the concrete event in \(\pi_0\) in fact tells us what those events are: they were events of cleaning the bathroom, cleaning the living room and the bedroom, and so on. This discourse is predicted to be coherent. On the other hand (100a) lacks any specific event that \((\pi_1)-\(\pi_3)\) elaborate on, and there is no way of specifying the eventualities posited by the mechanisms of coercion and predication adjustment. So (100a) is predicted to sound incomplete and somewhat incoherent, just as statement \textit{she was wearing a nice dress} sounds to us incomplete and vaguely incoherent in a context where there is no specifiable antecedent for the pronoun.

8 CONCLUSION

Montague Grammar and Dowty’s use thereof for lexical semantics provided a paradigm for linguists for the last forty years. However, more recent developments have led to a reconceptualization of what lexical semantics should do. Lexical meanings were seen to have entries that depended upon a much richer typing system as well as upon discourse context. These developments put pressure on the MG framework and led to a general forgetfulness concerning formal issues and foundations in formal semantics, although the descriptive detail concerning lexical meaning deepened considerably. This chapter has sketched a framework in which foundational issues, both technical and philosophical can be addressed.

BIBLIOGRAPHY


TYPE THEORY AND SEMANTICS IN FLUX

Robin Cooper

1 INTRODUCTION

A frequent assumption in computational and corpus linguistics as well as theoretical linguistics is that words are associated with a fairly small set of meanings, statically defined in a lexical resource. [Jurafsky and Martin, 2009, Chap. 19 is a standard textbook reference presenting this kind of view.] This view is challenged by work in the psychology of language [Clark and Wilkes-Gibbs, 1986; Garrod and Anderson, 1987; Pickering and Garrod, 2004; Brennan and Clark, 1996; Healey, 1997, among others] where dialogue participants are regarded as creating meaning on the fly for the purposes of particular dialogues and this view has been taken up by recent approaches to dialogue semantics [Larsson, 2007b; Larsson and Cooper, 2009; Cooper and Larsson, 2009; Ginzburg, forthcoming]. [Cooper, 2010a] argues that a view of lexical meaning in flux is important for the lexicon in general, not just for the analysis of dialogue. Here we will explore the philosophical underpinning of this argument, in particular the kind of type theory with records (TTR) that we propose [Cooper, 2005a; Cooper, 2005b; Ginzburg, forthcoming].

The philosophical argument relates to two views of language that create a tension in the philosophy of language that has essentially remained unresolved since the middle of the last century. The conflict is represented in the contrast between early and late Wittgenstein, that is, the view represented in the Tractatus [Wittgenstein, 1922] as opposed to Philosophical Investigations [Wittgenstein, 1953]. We can think of the positivistic view of early Wittgenstein as somewhat related to the view of natural languages as formal languages expressed by Montague [Montague, 1974], even though Montague was reacting against the positivistic view of natural language as imprecise and informal. Montague’s application of formal language techniques to natural language does, however, give the impression of natural languages as being regimented with meanings determined once and for all by an interpretation. This is a view which is very different from that of the late Wittgenstein who talked in terms of language games and the creation of public language for specific purposes. [Cooper and Ranta, 2008] represents a sketch of an attempt to take something like the late Wittgenstein view without throwing away the immense advances that were made in twentieth century semantics by the application of Montague’s techniques. The idea there is that natural languages are to be seen as toolboxes (resources) that can be used to create limited languages for
Use in particular language games in the sense of late Wittgenstein. These limited special purpose languages may be formal in the sense that Montague had in mind. We will argue, however, that there is a lot of linguistic interest in trying to discover not only how natural languages provide these formal languages but also how agents using the language apply and develop these resources which are constantly in a state of flux as we use the language. We will argue that our particular kind of type theory is appropriate for such an analysis whereas the kind of semantics of the classical model theoretic approach represented by Montague does not provide us with enough structure to capture the notions of variation in meaning that appear to be necessary.

When people talk to each other they create new language suitable for discussing the subject matter they are addressing. Occasionally, people will create entirely new words to express a new concept that they are trying to convey to their interlocutor. More often, though, they will use a previously existing word but with a modified meaning to match the new concept. In order to analyze this we need an approach to meaning in terms of structured objects that can be modified. Sometimes innovation is asymmetric in the sense that the speaker uses a word in a way that is not innovative for her but the hearer either does not know the word at all or has not previously heard the word associated with the particular meaning intended by the speaker. The hearer processes and learns the new way of using the word by modifying the meaning he had previously associated with the word or, if the word is entirely new to him, possibly by modifying a similar meaning he associates with a different word. In order to analyze this we need an approach to meaning which allows a general notion of a similarity measure on meanings. This, like the modification of meaning associated with the learning of the innovative meaning, can be achieved by treating meanings in terms of structured objects where we can see, for example, how many components a pair of structured meanings share.

The classical notion of meaning from model theoretic semantics is that meaning is a function from possible worlds and contexts to denotations derived from the domain of the model. We will argue that record types provide us with feature structure like objects which easily admit similarity measures and structural modifications because they are structured into fields containing a label and a value. Similarity measures can be created by comparing fields in two objects and objects can be modified by adding or deleting fields or changing the value provided for a particular field.

The general view of language in which our discussion will be cast is that of language as action, speech events that can cause changes in the mental states of dialogue participants during the course of linguistic interaction. This view of language, though it might be seen as contrasting with the kind of formal language view presented by Montague [Montague, 1974] or even the general Chomskyan tradition, is not new. Apart from Wittgenstein, it has roots, for example, in speech act theory [Austin, 1962; Searle, 1969]. An early attempt to take an action or event-based view of all aspects of compositional semantics is [Barwise and Perry, 1983]. Two recent works that develop a linguistic view of interaction are
Type Theory and Semantics in Flux

Ginzburg, forthcoming and Linell, 2009, although these two books take very different approaches and have almost no overlap in the literature they refer to.

A frequent complaint against classical formal semantics is that it has nothing to say about the details of word meaning. If you have a superficial analysis of word meaning then it can appear that uses of words in many different situations have the same meaning. We shall argue that as you examine the details of word meaning, we see that the situation is much more like that proposed in the late Wittgenstein, that is, word meaning varies according to the use to which it is put in a particular communicative situation. In the following sections we will pursue the example discussed in Cooper, 2010a and show in detail how to construct a type theory to support the analysis we propose, based on a notion of frame deriving from Frame Semantics [Fillmore, 1982; Fillmore, 1985]. In what follows many sections are marked with a star. These sections may be omitted on first reading. By following the unstarrred sections the reader will obtain an emended version of [Cooper, 2010a]. Readers who dip into the starred sections will in addition get a technical account of what is discussed in the unstarrred sections as well as some more philosophical background. The unstarrred sections occur at the beginning of the main sections. Section 2 is concerned with how we can use TTR to represent frames in the sense of Fillmore’s frame semantics and developing the type theory we need to do this. Section 3 is concerned with how such frames could be exploited in the compositional semantics of verbs. Section 4 shows how this analysis can be used to solve a classical puzzle from formal semantics: the Partee puzzle concerning the rising of temperature and price. Section 5 looks more deeply into the lexical semantics of a single verb rise using Fernando’s string theory of events. Here we discover that looking more deeply at the lexical meaning of this verb suggests that meaning varies from situation to situation and that there is always an option for creating new meaning. In section 6 we place this observation in the context of the view of coordination that has been developed by Larsson. Finally in section 7 we draw some conclusions.

2 FRAMES

2.1 Representing frames in TTR

Frame semantics was introduced in Fillmore’s classic paper [Fillmore, 1982]. We will use semantic objects which are related to the frames of FrameNet. An important part of our proposal will be that these objects can serve as the arguments to predicates. We will use record types as defined in TTR ([Cooper, 2005a; Cooper, 2005b; Ginzburg, forthcoming]) to characterize our frames. The advantage of records is that they are objects with a structure like attribute value matrices as used in linguistics. Labels (corresponding to attributes) in records allow us to access and keep track of parameters defined within semantic objects. This is in

marked contrast to classical model theoretic semantics where semantic objects are either atoms or unstructured sets and functions.

Consider the frame Ambient_temperature defined in the Berkeley FrameNet\(^2\) by “The Temperature in a certain environment, determined by Time and Place, is specified”. Its core frame elements are given in (1).

(1) **Attribute** The temperature feature of the weather

**Degree** A modifier expressing the deviation of the Temperature from the norm

**Place** The Place where it is a certain Temperature

**Temperature** A quantity or other characterization of the Temperature of the environment

**Time** The Time during which an ambient environment has a particular Temperature

To make things of a manageable size we will not include all the frame elements in our representation of this frame. (We have also changed the names of the frame elements to suit our own purposes.) We will say that an ambient temperature frame is a record of type (2).

(2) \[
\begin{array}{ll}
  x & : \text{Ind} \\
  \text{e-time} & : \text{Time} \\
  \text{e-location} & : \text{Loc} \\
  c_{\text{temp-at-in}} & : \text{temp-at-in(e-time, e-location, x)} \\
\end{array}
\]

We will call this type *AmbTemp*. It is a set of four fields each consisting of a label (to the left of the colon) and a type (to the right of the colon). A record of type *AmbTemp* will meet the following two conditions:

- it will contain at least fields with the same labels as the type (it may contain more)
- each field in the record with the same label as a field in the record type will contain an object of the type in the corresponding field of the record type. (Any additional fields with different labels to those in the record type may contain objects of any type.)

Types constructed with predicates such as ‘temp_at_in’ have a special status in that they can be dependent. In (2) the type in the field labelled ‘c_{temp-at-in}’ depends on what you choose for the other three fields in the frame. Intuitively, we can think of such types formed with a predicate like ‘temp_at_in’ as types of objects which prove a proposition. What objects you take to belong to these types

\(^2\)accessed 25th Oct, 2009
depends on what kind of theory of the world you have or what kind of application you want to use your type theory for. Candidates would be events, states or, in this case, thermometer or sensor readings.

The notions that we need to define in our type theory in order to achieve this are:

- **basic types**, such as \( \text{Ind} \), \( \text{Time} \) and \( \text{Loc} \)
- **complex types** constructed with predicates
- **record types** based on basic types and complex types with predicates

We will see below that our construction of record types will in addition require us to introduce function types and a type \( \text{Type} \) of types which will lead us to stratify our type system. We will begin by presenting some philosophical background for the type theory.

*2.2 Type theory, mathematics and cognition*

The philosophical foundation of type theory (as presented, for example, by [Martin-Löf, 1984]) is normally seen as related to intuitionism and constructive mathematics. It is, at bottom, a proof-theoretic discipline rather than a model-theoretic one (despite the fact that model theories have been provided for some type theories). However, it seems that many of the ideas in type theory that are important for the analysis of natural language can be adopted into the classical set theoretic framework familiar to linguists from the classical canon of formal semantics starting from [Montague, 1974]. There is a risk in pushing this line of alienating both the type theorists (who feel that the philosophical essence of type theory is being abandoned) and the linguists (who tend to feel that if one is going to move in the type theory direction then one should probably be doing proof theory rather than model theory). Ultimately, the line between a proof theoretical approach and a model-theoretic approach that advocates structured semantic objects can be a hard one to draw when viewed from the perspective of a theory of natural language or human cognition. Both approaches are advocating the need for more structured objects than are provided by classical model theory, objects whose components can be manipulated by formal processes which are meant to model agents’ cognitive processes. In this section we will attempt to present a philosophical view of our type theory as an important component in a theory of cognition.

The notion of type that we are discussing is more general than the notion of type found, for example, in Russell’s theory of types as it was adapted to Montague’s semantics, that is, entities, sets, sets of sets, function from objects of one type to another, and so on. The kind of types we are discussing here correspond to what might be called properties in other theories. Types correspond to pretty much any useful way of classifying things.

While perception and typing are at the core of cognitive processing an important feature of cognitive systems is the ability to consider alternative typings which have
not be observed. While we perceive $a$ to be of type $T_1$ it is perhaps nevertheless conceivable that $a$ could have been of type $T_2$. This leads us to construct modal type systems with alternative assignments of objects to types.

In addition to basic types, cognitive agents perceive the world in terms of states and events where objects have properties and stand in relations to each other — what [Barwise and Perry, 1983] called situations. Thus we introduce types which are constructed from predicates (like 'hug') and objects which are arguments to this predicate like $a$ and $b$. We will represent such a constructed type as $\text{hug}(a, b)$. What would an object belonging to such a type be? According to the type-theoretic approach introduced by Martin-Löf it should be an object which constitutes a proof that $a$ is hugging $b$. For Martin-Löf, who was considering mathematical predicates, such proof objects might be numbers with certain properties, ordered pairs and so on. [Ranta, 1994] points out that for non-mathematical predicates the objects could be events as conceived by [Davidson, 1980]. Thus $\text{hug}(a, b)$ can be considered to be an event or a situation type. In some versions of situation theory [Barwise, 1989; Seligman and Moss, 1997], objects (called \textit{infons}) constructed from a relation and its arguments was considered to be one kind of situation type. Thus one view would be that these kinds of types are playing a similar role in type theory to the role that infons play in situation theory.

These types play a role in the “propositions as types” dictum which comes from type theory. If $\text{hug}(a, b)$ is the type of events where $a$ hugs $b$ then the sentence “$a$ hugs $b$” will be true just in case this type is non-empty, that is, just in case there is an event where $a$ hugs $b$. The type can function as the theoretical object corresponding to the informal notion of proposition. It is “true” just in case it is non-empty.

An important aspect of human cognition is that we seem to be able to treat the types themselves as if they were objects. This becomes apparent when we consider attitude predicates like ‘believe’. In classical model theoretic semantics we think of \textit{believe} as corresponding to a relation between individuals and propositions. In our type theory, however, we are subscribing to the “propositions as types” view. It then follows that the second argument to the predicate ‘believe’ should be a type. That is, we should be able to construct the type $\text{believe}(c, \text{hug}(a, b))$ corresponding to $c$ \textit{believes that $a$ hugs $b$}. We thus create intensional type systems where types themselves can be treated as objects and belong to types. Care has to be taken in constructing such systems in order to avoid paradoxes. We use here a standard technique known as stratification [Turner, 2005]. We start with a basic type system and then add higher order levels of types. Each higher order includes the types of the order immediately below as objects. In each of these higher orders $n$ there will be a type of all types of the order $n-1$ but there is no ultimate “type of all types” — such a type would have to have itself as an object.

We will argue below that it is very important that the complex types we introduce are structured which enables them to be compared and modified. This is what makes it possible to account for how agents exploit and adapt the resources they have as they create new language during the course of interaction. It is not
quite enough, however, simply to have objects with components. We also need
a systematic way of accessing these components, a system of labelling which will
provide us with handles for the various pieces. This is where the record types of
TTR come in. There is a large literature on type theories with records in computer
science, for example, [Tasistro, 1997; Betarte, 1998; Betarte and Tasistro, 1998;
Coquand et al., 2004]. Our notion of record type is closely related to those dis-
cussed in this literature, though (like the rest of TTR) couched in rather different
terms. For us a record type is a set of fields where each field is an ordered pair
of a label and a type (or a pair consisting of a dependent type and a sequence of
path names corresponding to what the type is to depend on). A record belonging
to such a type is a set of fields which includes fields with the same labels as those
occurring in the type. Each field in the record with a label matching one in the
type must contain an object belonging to the type of the corresponding field in
the type.

It is an important aspect of human cognition that we not only appear to con-
struct complex cognitive objects out of smaller ones as their components but that
we also have ways of accessing the components and performing operations like sub-
stitutions, deletions and additions. Cognitive processing also appears to depend
on similarity metrics which require us to compare components. Thus labelling
or the provision of handles pointing to the components of complex objects is an
important part of a formal theory of human cognition and in TTR it is the records
and record types which do this work for us.

The importance of labelling has been reflected in the use of features in linguistic
theorizing ranging from the early Prague school [Trubetzkoy, 1939] to modern
feature based grammar [Sag et al., 2003]. It appears in somewhat different form
in the use of discourse referents in the treatment of discourse anaphora in formal
semantics [Kamp and Reyle, 1993]. In [Cooper, 2005b] we argue that record types
can be used both to model the feature structures of feature based grammar and
the discourse representation structures of discourse representation theory. This
is part of a general programme for developing TTR to be a general type theory
which underlies all our linguistic cognitive processing. In fact, what we would like
to see in the future is a single type theory which underlies all of human cognitive
processing.

In contrast to the statement of TTR in [Cooper, 2005a] we attempt here to
present interacting modules which are not that complex in themselves. Never-
theless, knowing exactly what you have got when you put everything together
is not an entirely trivial matter. It would be nice from a logical point of view
if human cognition presented itself to us in neat separate boxes which we could
study independently. But this is not the case. We are after all involved in the
study of a biological system and there is no reason in principle why our cognitive
anatomy should be any simpler than our physical anatomy with its multiplicity
of objects such as organs, nerves, muscles and arteries and complex dependen-
cies between them, though all built up on the basis of general principles of cell
structure and DNA. Compared with what we know about physical anatomy, TTR
seems quite modest in the number of different kinds of objects it proposes and the interrelationships between them.

*2.3 Basic types

The simplest type system we will introduce has no complex types. All the types are atoms (that is they are objects which are not constructed from other objects in the system) and the of-type relation is determined by a function which assigns sets of objects to types. We will call this a system of basic types.

A system of basic types is a pair:

\[ \text{TYPE}_B = \langle \text{Type}, A \rangle \]

where:

1. \text{Type} is a non-empty set
2. \( A \) is a function whose domain is \text{Type}
3. for any \( T \in \text{Type} \), \( A(T) \) is a set disjoint from \text{Type}
4. for any \( T \in \text{Type} \), \( a : \text{TYPE}_B T \) iff \( a \in A(T) \)

Central to type theory is the notion of judgements that an object \( a \) is of a type \( T \) (in symbols \( a : T \)). We see this as being fundamentally related to perception. When we perceive objects in the world, we perceive them as belonging to a particular type (or perhaps several types). There is no perception without some kind of judgement with respect to types of the perceived object. When we say that we do not know what an object is, this normally means that we do not have a type for the object which is narrow enough for the purposes at hand. I trip over something in the dark, exclaiming “What’s that?”, but my painful physical interaction with it through my big toe tells me at least that it is a physical object, sufficiently hard and heavy to offer resistance to my toe. The act of perceiving an object is perceiving it as something. That “something” is a type.

The notion of type judgements yields a type theory with two domains: one domain for the objects and another domain for the types to which these objects belong. Thus we see types as theoretical entities in their own right, not, for example, as collections of objects. Diagrammatically we can represent this as in Figure 1 where object \( a \) is of type \( T_1 \).

*2.4 Complex types

We start by introducing the notion of a predicate signature.

A predicate signature is a triple

\[ \langle \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle \]

where:
1. \textbf{Pred} is a set (of predicates)

2. \textbf{ArgIndices} is a set (of indices for predicate arguments, normally types)

3. \textit{Arity} is a function with domain \textbf{Pred} and range included in the set of finite sequences of members of \textbf{ArgIndices}.

A \textit{polymorphic predicate signature} is a triple

\[ \langle \text{Pred}, \text{ArgIndices}, \text{Arity}\rangle \]

where:

1. \textbf{Pred} is a set (of predicates)

2. \textbf{ArgIndices} is a set (of indices for predicate arguments, normally types)

3. \textit{Arity} is a function with domain \textbf{Pred} and range included in the powerset of the set of finite sequences of members of \textbf{ArgIndices}.

A \textit{system of complex types} is a quadruple:

\[ \text{TYPE}_C = \langle \text{Type}, \text{BType}, \langle \text{PType, Pred, ArgIndices, Arity}, \langle A, F\rangle \rangle \]
where:

1. \( (\text{BType}, A) \) is a system of basic types

2. \( \text{BType} \subseteq \text{Type} \)

3. for any \( T \in \text{Type} \), if \( a : (\text{BType}, A) T \) then \( a : \text{TYPE}_C T \)

4. \( (\text{Pred}, \text{ArgIndices}, \text{Arity}) \) is a (polymorphic) predicate signature

5. If \( P \in \text{Pred} \), \( T_1 \in \text{Type}, \ldots, T_n \in \text{Type} \), \( \text{Arity}(P) = \langle T_1, \ldots, T_n \rangle \) and \( a_1 : \text{TYPE}_C T_1, \ldots, a_n : \text{TYPE}_C T_n \) then \( P(a_1, \ldots, a_n) \in \text{PType} \)

6. \( \text{PType} \subseteq \text{Type} \)

7. for any \( T \in \text{PType} \), \( F(T) \) is a set disjoint from \( \text{Type} \)

8. for any \( T \in \text{PType} \), \( a : \text{TYPE}_C T \) iff \( a \in F(T) \)

*2.5 Complex types in record types

If we look back at the record type in (2) now we notice that there is something odd about the type constructed with the predicate \( \text{temp at in} \), namely that the arguments to the predicate appear to be the labels ‘e-time’, ‘e-location’ and ‘x’ rather than objects that might occur under these labels in a record of this type. It is objects that are appropriate arguments to a predicate not the labels. (2) is actually a convenient abbreviatory notation for (3).

\[
\begin{bmatrix}
\text{x} & : & \text{Ind} \\
\text{e-time} & : & \text{Time} \\
\text{e-location} & : & \text{Loc} \\
\text{c_temp_at_in} & : & \langle \lambda v_1 : \text{Time}( \\
& & \quad \lambda v_2 : \text{Loc}( \\
& & \quad \quad \lambda v_3 : \text{Ind}( \\
& & \quad \quad \quad \text{temp_at_in}(v_1,v_2,v_3))), \\
& & \quad \quad \langle \text{e-time}, \text{e-location}, x \rangle) \rangle \\
\end{bmatrix}
\]

Here what occurs in the \( c_{\text{temp at in}} \)-field is a pair whose first member is a function and whose second member is a list of labels indicating the fields in a record where the objects which are to be the arguments to the function are to be found. When applied to these objects the function will return a type constructed from the predicate and the objects.

For many simple cases such as (2) the abbreviatory notation is adequate and much easier to read as long as we keep in mind how it is to be interpreted. Care has to be taken, however, when record types are arguments to predicates. Consider a putative representation of a type corresponding to a reading of *some man appears to own a donkey* where *appear* is treated as corresponding to a one place predicate taking a record type as argument:
Technically, this notation is incorrect since ‘x’ occurring within the argument to ‘appear’ picks up a path outside of the record type in which it occurs. The full and correct notation for this type would be:

\[
\begin{cases}
  x : \text{Ind} \\
  c_1 : \text{man}(x) \\
  c_3 : \text{appear}( \begin{cases}
    y : \text{Ind} \\
    c_2 : \text{donkey}(y) \\
    c_4 : \text{own}(x,y) 
  \end{cases} )
\end{cases}
\]

When labels are unique there is no harm in using the imprecise notation.

The full treatment of types constructed with predicates which depend on the values introduced in other fields as in these examples requires us to add functions and function types to our type theory. Furthermore, since the function returns a type, we will need a type of types since we want to be able to say that the function takes objects of types Time, Loc and Ind and returns a type, that is an object of type Type. Once we have done this we will be ready to give an explicit definition of record types.

### 2.6 Function types

A system of complex types \( \text{TYPE}_C = \langle \text{Type}, \text{BType}, \langle \text{PType}, \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle, \langle A, F \rangle \rangle \) has function types if

1. for any \( T_1, T_2 \in \text{Type} \), \( (T_1 \rightarrow T_2) \in \text{Type} \)

2. for any \( T_1, T_2 \in \text{Type} \), \( f : \text{TYPE}_C \) \( (T_1 \rightarrow T_2) \) if \( f \) is a function whose domain is \( \{ a \mid a : \text{TYPE}_C T_1 \} \) and whose range is included in \( \{ a \mid a : \text{TYPE}_C T_2 \} \)

### 2.7 The type Type and stratification

An intensional type system is one in which the types themselves become objects of a type. We introduce a distinguished type Type to which all the members of the set Type belong. Things are a little more complicated than this, though, since we want Type itself to be a type and therefore it should belong to the set Type. This would mean that Type belongs to itself, i.e. Type:Type. Allowing types to belong to themselves puts us in danger of creating a situation in which Russell’s paradox arises. If some members of Type belong to themselves then we should be able to talk of the set of types which do not belong to themselves, \( \{ T \in \text{Type} \mid T \not\in T \} \).

Suppose that some model assigns this set to \( T' \). Then the question arises whether
In order to avoid this problem we will stratify (or ramify) our type system by introducing types of different orders. A type system of order 0 will be a system of complex types in the way we have defined it. The set of types, $\text{Type}^1$, of a type system of order 1 based on this system will contain in addition to everything in the original type system a type, $\text{Type}^1$, to which all the types of order 0, members of the set $\text{Type}^0$, belong. In general for all the natural numbers $n$, $\text{Type}^{n+1}$ will be a type to which all the types in $\text{Type}^n$ belong. But there may be more additional types included in the higher sets of types. Suppose, for example, that we want to introduce a predicate $P$ expressing a relationship between individuals and types. (This will be our basic strategy for the treatment of attitude predicates such as believe and know.) Then $\text{Arity}(P)$ might be $\langle \text{Ind}, \text{Type}^n \rangle$. In systems of any order less than $n$, $P$ will not be able to be used to construct a type because clause 4 in our definition of systems of complex types requires that the types assigned to the arguments be types in the system. However, in systems of order $n$ or greater the required type will be present and the predicate will form a type.

This avoids the risk of running into Russell’s paradox but it introduces another problem which it is best we deal with straight away. We will illustrate the problem by creating a small example. Suppose that we have a system of complex types which includes the type $\text{Ind}$ (“individuals”) to which the objects $a$, $b$ and $c$ belong. Suppose further that we have three predicates $\text{run}$, $\text{know}$ and $\text{believe}$ and that $\text{Arity}(\text{run})=\langle \text{Ind} \rangle$ and $\text{Arity}(\text{know})=\text{Arity}(\text{believe})=\langle \text{Ind}, \text{Type}^1 \rangle$. The set $\text{Type}^0$ will contain the types $\text{run}(a)$, $\text{run}(b)$ and $\text{run}(c)$ but no types constructed with $\text{know}$ and $\text{believe}$. The set $\text{Type}^1$ will contain types such as $\text{believe}(a, \text{run}(a))$ and $\text{know}(c, \text{run}(b))$ in addition, since $\text{run}(a)$, $\text{run}(b)$ and $\text{run}(c)$, being members of $\text{Type}^0$ will belong to the type $\text{Type}^1$. The set $\text{Type}^2$ will not get any additional types constructed with predicates since the arity of the predicates restricts the second argument to be of $\text{Type}^1$. But suppose we want to express that $a$ believes that $b$ knows that $c$ runs, that is we want to construct the type $\text{believe}(a, \text{know}(b, \text{run}(c)))$. Perhaps we could solve this by saying that the arity of $\text{know}$ and $\text{believe}$ is $\langle \text{Ind}, \text{Type}^2 \rangle$. But now $\text{Type}^1$ will not contain any types constructed with these predicates and $\text{Type}^2$ will again only contain types such as $\text{know}(c, \text{run}(b))$.

In order to solve this problem we need to introduce a limited amount of polymorphism into our arities and assign these predicates the arity $\langle \text{Ind}, \text{Type}^n \rangle_{n>0}$ (that is, the set of sequences $\langle \text{Ind}, \text{Type}^n \rangle$ where $n$ is a natural number greater than 0). Predicates with this arity will be able to take arguments of any type $\text{Type}^n$ where $n>0$. We will say that the predicates $\text{know}$ and $\text{believe}$ have this arity. Now it will be the case that $\text{run}(c):\text{Type}^1$, $\text{know}(b, \text{run}(c)):\text{Type}^2$, $\text{believe}(a, \text{know}(b, \text{run}(c))):\text{Type}^3$ and so on.

An intensional system of complex types is a family of quadruples indexed by the natural numbers:

$$\text{TYPE}_{IC} = \langle \text{Type}^n, \text{BType}, \langle \text{PType}^n, \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle, \ldots \rangle.$$
\[ \langle A, F^n \rangle \subseteq \text{Nat} \]

where (using \( \text{TYPE}_{IC_n} \) to refer to the quadruple indexed by \( n \)):

1. for each \( n \), \( \langle \text{Type}^n, \text{BType}, (\text{PType}^n, \text{Pred}, \text{ArgIndices}, \text{Arity}), \langle A, F^n \rangle \rangle \)
   is a system of complex types

2. for each \( n \), \( \text{Type}^n \subseteq \text{Type}^{n+1} \) and \( \text{PType}^n \subseteq \text{PType}^{n+1} \)

3. for each \( n \), if \( T \in \text{PType}^n \) and \( p \in F^n(T) \) then \( p \in F^{n+1}(T) \)

4. for each \( n > 0 \), \( \text{Type}^n \in \text{Type}^n \)

5. for each \( n > 0 \), \( T : \text{TYPE}_{IC_n} \) \( \text{Type}^n \) iff \( T \in \text{Type}^{n-1} \)

We can represent a stratified intensional system of types diagrammatically as Figure 2 where we represent just the first three levels of an infinite stratification.

![Intensional system of types with stratification](image)

Figure 2. Intensional system of types with stratification

An intensional system of complex types \( \text{TYPE}_{IC} \),
\( \text{TYPE}_{IC} = \langle \text{Type}^n, B\text{Type}, \langle \text{PType}^n, \text{Pred, ArgIndices, Arity} \rangle, \langle A, F^n \rangle \rangle_{n \in \text{Nat}} \)

has dependent function types if

1. for any \( n > 0 \), \( T \in \text{Type}^n \) and \( F : \text{TYPE}_{IC, n} (T \rightarrow \text{Type}^n), ((a : T) \rightarrow F(a)) \in \text{Type}^n \)

2. for each \( n > 0 \), \( f : \text{TYPE}_{IC, n} ((a : T) \rightarrow F(a)) \) iff \( f \) is a function whose domain is \( \{ a \mid a : \text{TYPE}_{IC, n} T \} \) and such that for any \( a \) in the domain of \( f \), \( f(a) : \text{TYPE}_{IC, n} F(a) \).

We might say that on this view dependent function types are “semi-intensional” in that they depend on there being a type of types for their definition but they do not introduce types as arguments to predicates and do not involve the definition of orders of types in terms of the types of the next lower order.

*2.8 Record types*

In this section we will define what it means for a system of complex types to have record types. The objects of record types, that is, records, are themselves structured mathematical objects of a particular kind and we will start by characterizing them.

A record is a finite set of ordered pairs (called fields) which is the graph of a function. If \( r \) is a record and \( (\ell, v) \) is a field in \( r \) we call \( \ell \) a label and \( v \) a value in \( r \) and we use \( r.\ell \) to denote \( v. \) \( r.\ell \) is called a path in \( r \).

We will use a tabular format to represent records. A record \( \{ (\ell_1, v_1), \ldots, (\ell_n, v_n) \} \) is displayed as

\[
\begin{bmatrix}
\ell_1 &= v_1 \\
\cdots \\
\ell_n &= v_n
\end{bmatrix}
\]

A value may itself be a record and paths may extend into embedded records. A record which contains records as values is called a complex record and otherwise a record is simple. Values which are not records are called leaves. Consider a record \( r \)

\[
\begin{bmatrix}
f &= \begin{bmatrix}
f &= \begin{bmatrix}
ff &= a \\
gg &= b
\end{bmatrix} \\
g &= c
\end{bmatrix} \\
h &= \begin{bmatrix}
g &= a \\
h &= d
\end{bmatrix}
\end{bmatrix}
\]

Among the paths in \( r \) are \( r.f, r.g.h \) and \( r.f.f.ff \) which denote, respectively,

\[
\begin{bmatrix}
f &= \begin{bmatrix}
ff &= a \\
gg &= b
\end{bmatrix} \\
g &= c
\end{bmatrix}
\]
and \( a \). We will make a distinction between *absolute paths*, such as those we have already mentioned, which consist of a record followed by a series of labels connected by dots and *relative paths* which are just a series of labels connected by dots, e.g. \( g.h \). Relative paths are useful when we wish to refer to similar paths in different records. We will use *path* to refer to either absolute or relative paths when it is clear from the context which is meant. The set of leaves of \( r \), also known as its *extension* (those objects other than labels which it contains), is \( \{ a, b, c, d \} \). The bag (or multiset) of leaves of \( r \), also known as its *multiset extension*, is \( \{ a, a, b, c, d \} \). A record may be regarded as a way of labelling and structuring its extension. Two records are *(multiset) extensionally equivalent* if they have the same (multiset) extension. Two important, though trivial, facts about records are:

*Flattening.* For any record \( r \), there is a multiset extensionally equivalent simple record. We can define an operation of flattening on records which will always produce an equivalent simple record. In the case of our example, the result of flattening is

\[
\begin{bmatrix}
  f & f.f.f &= a \\
  f & f.g &= c \\
  g & g &= a \\
  g & g.h &= d \\
\end{bmatrix}
\]

assuming the flattening operation uses paths from the original record in a rather obvious way to create unique labels for the new record.

*Relabelling.* For any record \( r \), if \( \pi_1.\ell.\pi_2 \) is a path \( \pi \) in \( r \), and \( \pi_1.\ell'.\pi_2' \) is not a path in \( r \) (for any \( \pi_2' \)), then substituting \( \ell' \) for the occurrence of \( \ell \) in \( \pi \) results in a record which is multiset equivalent to \( r \). We could, for example, substitute \( k \) for the second occurrence of \( g \) in the path \( g.h.g \) in our example record.

\[
\begin{bmatrix}
  f &= \begin{bmatrix}
  ff &= a \\
  gg &= b \\
\end{bmatrix} \\
  g &= \begin{bmatrix}
  c \\
\end{bmatrix} \\
  h &= \begin{bmatrix}
  k &= a \\
  h &= d \\
\end{bmatrix}
\end{bmatrix}
\]

A record *type* is a record in the general sense defined above where the values in its fields are types or, in some cases, certain kinds of mathematical objects which can be used to construct types.

A record \( r \) is *well-typed* with respect to a system of types \( \text{TYPE} \) with set of types \( \text{Type} \) and a set of labels \( L \) iff for each field \( \langle \ell, a \rangle \in r \), \( \ell \in L \) and either \( a : \text{TYPE} T \) for some \( T \in \text{Type} \) or \( a \) is itself a record which is well-typed with respect to \( \text{TYPE} \) and \( L \).
A system of complex types \( \text{TYPE}_C = \langle \text{Type}, \text{BType}, \langle \text{PType}, \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle, \langle A, F \rangle \rangle \) has record types based on \( \langle L, \text{RType} \rangle \), where \( L \) is a countably infinite set (of labels) and \( \text{RType} \subseteq \text{Type} \), where \( \text{RType} \) is defined by:

1. \( \text{Rec} \in \text{RType} \)
2. \( r : \text{TYPE}_C \text{Rec} \) iff \( r \) is a well-typed record with respect to \( \text{TYPE}_C \) and \( L \).
3. if \( \ell \in L \) and \( T \in \text{Type} \), then \( \{ \langle \ell, T \rangle \} \in \text{RType} \).
4. if \( R \in \text{RType} \), \( \ell \in L \), \( \ell \) does not occur as a label in \( R \) (i.e. there is no field \( \langle \ell', T' \rangle \) in \( R \) such that \( \ell' = \ell \)), then \( R \cup \{ \langle \ell, T \rangle \} \in \text{RType} \).

This gives us non-dependent record types in a system of complex types. We can extend this to intensional systems of complex types (with stratification).

An intensional system of complex types \( \text{TYPE}_{IC} = \langle \text{Type}^n, \text{BType}, \langle \text{PType}^n, \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle, \langle A, F^n \rangle \rangle \) has record types based on \( \langle L, \text{RType}^n \rangle \) if for each \( n \), \( \langle \text{Type}^n, \text{BType}, \langle \text{PType}^n, \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle, \langle A, F^n \rangle \rangle \) has record types based on \( \langle L, \text{RType}^n \rangle \) and

1. for each \( n \), \( \text{RType}^n \subseteq \text{RType}^{n+1} \)
2. for each \( n > 0 \), \( \text{RecType}^n \in \text{RType}^n \)
3. for each \( n > 0 \), \( T : \text{TYPE}_{IC} \text{RecType}^n \) iff \( T \in \text{RType}^{n-1} \)

Intensional type systems may in addition contain dependent record types.

An intensional system of complex types \( \text{TYPE}_{IC} = \langle \text{Type}^n, \text{BType}, \langle \text{PType}^n, \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle, \langle A, F^n \rangle \rangle \) has dependent record types based on \( \langle L, \text{RType}^n \rangle \) if it has records types based on \( \langle L, \text{RType}^n \rangle \) and for each \( n > 0 \)

1. if \( R \) is a member of \( \text{RType}^n \), \( \ell \in L \) not occurring as a label in \( R \), \( T_1, \ldots, T_m \in \text{Type}^n \), \( R.\pi_1, \ldots, R.\pi_m \) are paths in \( R \) and \( F \) is a function of type \( (\langle a_1 : T_1 \rangle \to \ldots \to \langle a_m : T_m \rangle) \to \text{Type}^n \)\), then \( R \cup \{ \langle \ell, \langle F, \langle \pi_1, \ldots, \pi_m \rangle \rangle \} \in \text{RType}^n \).
2. \( r : \text{TYPE}_{IC} \text{R} \cup \{ \langle \ell, \langle F, \langle \pi_1, \ldots, \pi_m \rangle \rangle \} \) iff \( r : \text{TYPE}_{IC} \text{R} \), \( \langle \ell, a \rangle \) is a field in \( r \), \( r.\pi_1 : \text{TYPE}_{IC} T_1, \ldots, r.\pi_m : \text{TYPE}_{IC} T_m \) and \( a : \text{TYPE}_{IC} F(r.\pi_1, \ldots, r.\pi_m) \).

We represent a record type \( \{ \langle \ell_1, T_1 \rangle, \ldots, \langle \ell_n, T_n \rangle \} \) graphically as

\[
\begin{bmatrix}
\ell_1 & : & T_1 \\
\vdots & & \\
\ell_n & : & T_n
\end{bmatrix}
\]
In the case of dependent record types we sometimes use a convenient notation representing e.g.

$$\langle \lambda u \lambda v \text{love}(u, v), \langle \pi_1, \pi_2 \rangle \rangle$$

as

$$\text{love}(\pi_1, \pi_2)$$

Our systems now allow both function types and dependent record types and allow dependent record types to be arguments to functions. We have to be careful when considering what the result of applying a function to a dependent record type should be. Consider the following simple example:

$$\lambda v_0 : \text{RecType}(\{ c_0 : v_0 \})$$

What should be the result of applying this function to the record type

$$\left[ \begin{array}{l}
  x : \text{Ind} \\
  c_1 : (\lambda v_1 : \text{Ind}(\text{dog}(v_1)), \langle x \rangle)
\end{array} \right]$$

Given normal assumptions about function application the result would be

$$\left[ \begin{array}{l}
  c_0 : \left[ \begin{array}{l}
    x : \text{Ind} \\
    c_1 : (\lambda v_1 : \text{Ind}(\text{dog}(v_1)), \langle x \rangle)
  \end{array} \right]
\end{array} \right]$$

but this would be incorrect. In fact it is not a well-formed record type since $x$ is not a path in it. Instead the result should be

$$\left[ \begin{array}{l}
  c_0 : \left[ \begin{array}{l}
    x : \text{Ind} \\
    c_1 : (\lambda v_1 : \text{Ind}(\text{dog}(v_1)), \langle c_0.x \rangle)
  \end{array} \right]
\end{array} \right]$$

where the path from the top of the record type is specified. Note that this adjustment is only required when a record type is being substituted into a position that lies on a path within a resulting record type. It will not, for example, apply in a case where a record type is to be substituted for an argument to a predicate such as when applying the function

$$\lambda v_0 : \text{RecType}(\{ c_0 : \text{appear}(v_0) \})$$

to

$$\left[ \begin{array}{l}
  x : \text{Ind} \\
  c_1 : (\lambda v : \text{Ind}(\text{dog}(v)), \langle x \rangle) \\
  c_2 : (\lambda v : \text{Ind}(\text{approach}(v)), \langle x \rangle)
\end{array} \right]$$

where the position of $v_0$ is in an “intensional context”, that is, as the argument to a predicate and there is no path to this position in the record type resulting from applying the function. Here the result of the application is
with no adjustment necessary to the paths representing the dependencies.\(^3\) (Note that ‘c\(_0\).x’ is not a path in this record type.)

These matters arise as a result of our choice of using paths to represent dependencies in record types (rather than, for example, introducing additional unique identifiers to keep track of the positions within a record type as has been suggested by Thierry Coquand). It seems like a matter of implementation rather than a matter of substance and it is straightforward to define a path-aware notion of substitution which can be used in the definition of what it means to apply a TTR function to an argument. If \(f\) is a function represented by \(\lambda v : T(\phi)\) and \(\alpha\) is the representation of an object of type \(T\), then the result of applying \(f\) to \(\alpha\), \(f(\alpha)\), is represented by \(\text{Subst}(\alpha,v,\phi,\emptyset)\), that is, the result of substituting \(\alpha\) for \(v\) in \(\phi\) with respect to the empty path where for arbitrary \(\alpha,v,\phi,\pi\), \(\text{Subst}(\alpha,v,\phi,\pi)\) is defined as

1. \(\text{extend-paths}(\alpha,\pi)\), if \(\phi\) is \(v\)
2. \(\phi\), if \(\phi\) is of the form \(\lambda v : T(\zeta)\), for some \(T\) and \(\zeta\) (i.e. don’t do any substitution if \(v\) is bound within \(\phi\))
3. \(\lambda u : T(\text{Subst}(\alpha,v,\zeta,\pi))\), if \(\phi\) is of the form \(\lambda u : T(\zeta)\) and \(u\) is not \(v\).
4. \[
\begin{bmatrix}
\ell_1 & : & \text{Subst}(\alpha,v,T_1,\pi,\ell_1) \\
\ldots & & \\
\ell_n & : & \text{Subst}(\alpha,v,T_n,\pi,\ell_n)
\end{bmatrix},\text{ if } \phi \text{ is } \\
\begin{bmatrix}
\ell_1 & : & T_1 \\
\ldots & & \\
\ell_n & : & T_n
\end{bmatrix}
\]
5. \(P(\text{Subst}(\alpha,v,\beta_1,\pi),\ldots,\text{Subst}(\alpha,v,\beta_n,\pi))\), if \(\alpha\) is \(P(\beta_1,\ldots,\beta_n)\) for some predicate \(P\)
6. \(\phi\) otherwise

\(\text{extend-paths}(\alpha,\pi)\) is

1. \(\langle f, \langle \pi,\pi_1,\ldots,\pi,\pi_n \rangle \rangle\), if \(\alpha\) is \(\langle f, \langle \pi_1,\ldots,\pi_n \rangle \rangle\)
2. \[
\begin{bmatrix}
\ell_1 & : & \text{extend-paths}(T_1,\pi) \\
\ldots & & \\
\ell_n & : & \text{extend-paths}(T_n,\pi)
\end{bmatrix},\text{ if } \alpha \text{ is } \\
\begin{bmatrix}
\ell_1 & : & T_1 \\
\ldots & & \\
\ell_n & : & T_n
\end{bmatrix}
\]
3. \(P(\text{extend-paths}(\beta_1,\pi),\ldots,\text{extend-paths}(\beta_n,\pi))\), if \(\alpha\) is \(P(\beta_1,\ldots,\beta_n)\) for some predicate \(P\)
4. \(\alpha\), otherwise

\(^3\)This record corresponds to the interpretation of it appears that a dog is approaching.
3 FRAMES IN THE COMPOSITIONAL SEMANTICS OF VERBS

3.1 Verbs as functions from frames to frame types

Consider an intransitive verb such as *run*. Basically, this corresponds to a predicate of individuals. Thus (4) would represent the type of events or situations where the individual Sam (‘sam’) runs.

(4) \text{run}(\text{sam})

On FrameNet\(^4\) *run* on one of its readings is associated with the frame \text{Self motion}. Like many other frames in FrameNet this has a frame element \text{Time} which in this frame is explained in this case as “The time when the motion occurs”. This is what Reichenbach [Reichenbach, 1947] called more generally event time and we will use the label ‘e-time’. We will add an additional argument for a time to the predicate and create a frame-type (5).\(^5\)

\[
\begin{array}{c}
\text{e-time} : \text{TimeInt} \\
\text{c}_{\text{run}} : \text{run(sam, e-time)}
\end{array}
\]

For the type (5) to be non-empty it is required that there be some time interval at which Sam runs. We use \text{TimeInt} as an abbreviation for the type of time intervals, (6).

\[
\begin{array}{c}
\text{start} : \text{Time} \\
\text{end} : \text{Time} \\
\text{c} : \text{start < end}
\end{array}
\]

In (5) there are no constraints on the time interval apart from the requirement that Sam runs at that time. A record will be of this type just in case in provides some time interval at which Sam runs with the appropriate labels. Thus this frame type corresponds to a “tenseless proposition”, something that is not available in the Priorean setup [Prior, 1957; Prior, 1967] that Montague employs where logical formulae without a tense operator correspond to a present tense interpretation. In order to be able to add tense to this we need to relate the event time to another time interval, normally the time which Reichenbach calls the speech time.\(^6\) A past tense type anchored to a time interval \(\iota\) is represented in (7).

\[
\begin{array}{c}
\text{e-time} : \text{TimeInt} \\
\text{c}_{\text{tns}} : \text{e-time.end < } \iota.\text{start}
\end{array}
\]

---

\(^4\)Accessed 1st April, 2010.

\(^5\)Of course, we are ignoring many other frame elements which occur in FrameNet’s \text{Self motion} which could be added to obtain a more detailed semantic analysis.

\(^6\)Uses of historic present tense provide examples where the tense is anchored to a time other than the speech time.
This requires that the end of the event time interval has to precede the start of the speech time interval. In order for a past-tense sentence *Sam ran* to be true we would need to find an object of both types (5) and (7). This is equivalent to requiring that there is an object in the result of merging the two types given in (8). (We make the notion of merge precise in section *3.2.*

\[
\begin{align*}
(8) \quad & \begin{cases}
    \text{e-time} : & \text{TimeInt} \\
    c_{\text{tsns}} : & \text{e-time.end} < \iota.\text{start} \\
    c_{\text{run}} : & \text{run(sam,e-time)}
\end{cases}
\end{align*}
\]

Suppose that we have an utterance \( u \), that is, a speech event of type (9).

\[
\begin{align*}
(9) \quad & \begin{cases}
    \text{phon} : & \text{“sam”“ran”} \\
    \text{s-time} : & \text{TimeInt} \\
    c_{\text{utt}} : & \text{uttered(phon,s-time)}
\end{cases}
\end{align*}
\]

where “sam”“ran” is the type of strings of an utterance of *Sam* concatenated with an utterance of *ran*. (See section *3.4 for a discussion of string types.) Then we can say that the speech time interval \( \iota \) in (8) is \( u.\text{s-time} \). That is, the past tense constraint requires that the event happened before the start of the speech event.

(8) is a type which is the content of an utterance of the sentence *Sam ran*. In order to obtain the content of the verb *ran* we need to create a function which abstracts over the first argument of the predicate. Because frames will play an important role as arguments to predicates below we will not abstract over individuals but rather over frames containing individuals. The content of the verb *ran* will be (10).

\[
\begin{align*}
(10) \quad & \lambda r : \begin{cases}
    x : \text{Ind} \\
    \text{e-time} : & \text{TimeInt} \\
    c_{\text{tsns}} : & \text{e-time.end} < \iota.\text{start} \\
    c_{\text{run}} : & \text{run}(r.x,e-time)
\end{cases}
\end{align*}
\]

We show how this content can be utilized in a toy grammar in section *3.5.*

*3.2 Meets and merges*

A system of complex types \( \text{TYPE}_C = \langle \text{Type}, \text{BType}, \langle \text{PType}, \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle, \langle A, F \rangle \rangle \) has meet types if

1. for any \( T_1, T_2 \in \text{Type} \), \( (T_1 \land T_2) \in \text{Type} \)

2. for any \( T_1, T_2 \in \text{Type} \), \( a : \text{TYPE}_C (T_1 \land T_2) \) iff \( a : \text{TYPE}_C T_1 \) and \( a : \text{TYPE}_C T_2 \)

This definition does not make precise exactly which mathematical object is denoted by \( T_1 \land T_2 \). Our intention is that it denote an object which contains the symbol ‘\( \land \)’ as a component, for example, the triple \( \langle \land, T_1, T_2 \rangle \). Note that if \( T_1 \) and
$T_2$ are record types as defined in section *2.8, then $T_1 \land T_2$ will not be a record type in the sense of this definition, since it is not a set of fields as required by the definition. This is true despite the fact that anything which is of the type $T_1 \land T_2$ where $T_1$ and $T_2$ are record types will be a record. There will, however, be a record type which is equivalent to the meet type.

There is a range of notions of equivalence which are available for types. For present purposes we will use a notion we call necessary equivalence which says that two types $T_1$ and $T_2$ are necessarily equivalent just in case $a : T_1$ iff $a : T_2$ on any assignment to basic types, $A$, and assignment to types constructed from a predicate and its arguments, $F$. This relates to the definition of a system of complex types in section *2.4, that is a system $\text{TYPE}_C = \langle \text{Type}, \text{BType}, \langle \text{PType}, \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle, \langle A, F \rangle \rangle \rangle$. The idea is that a notion of equivalence related to a single system of complex types $\text{TYPE}_C$ that would say that $T_1$ is equivalent to $T_2$ just in case $a : \text{TYPE}_C T_1$ iff $a : \text{TYPE}_C T_2$ would be a weaker notion of “material equivalence”. Necessary equivalence is a stronger notion that requires that $T_1$ and $T_2$ have the same extension no matter which functions $A$ and $F$ are chosen. We make this precise by introducing modal systems of complex types (section *3.3).

If $T_1$ and $T_2$ are record types then there will always be a record type (not a meet) $T_3$ which is necessarily equivalent to $T_1 \land T_2$. Let us consider some examples:

$$[f:T_1] \land [g:T_2] \approx [f:T_1 \mid g:T_2]$$
$$[f:T_1] \land [f:T_2] \approx [f:T_1 \land T_2]$$

Below is a more logically oriented definition of the simplification of meets of record types than that given in [Cooper, 2008]. We define a function $\mu$ which maps meets of record types to an equivalent record type, record types to equivalent types where meets in their values have been simplified by $\mu$ and any other types to themselves:

1. If for some $T_1$, $T_2$, $T = T_1 \land T_2$ then $\mu(T) = \mu(\mu(T_1) \land \mu(T_2))$.

2. If $T$ is a record type then $\mu(T)$ is $T'$ such that for any $\ell, v$, $(\ell, \mu(v)) \in T'$ iff $(\ell, v) \in T$.

3. Otherwise $\mu(T) = T$.

$\mu'(T_1 \land T_2)$ is defined by:

1. if $T_1$ and $T_2$ are record types, then $\mu'(T_1 \land T_2) = T_3$ such that

   (a) for any $\ell, v_1, v_2$, if $(\ell, v_1) \in T_1$ and $(\ell, v_2) \in T_2$, then

      i. if $v_1$ and $v_2$ are $\langle \lambda u_1 : T'_1 \ldots \lambda u_i : T'_i(\phi), (\pi_1, \ldots, \pi_i) \rangle$ and $\langle \lambda u'_1 : T''_1 \ldots \lambda u'_{k'} : T''_k(\psi), (\pi'_1, \ldots, \pi'_{k'}) \rangle$ respectively, then $\langle \lambda u_1 : T'_1 \ldots \lambda u_i : T'_i, \lambda u'_1 : T''_1 \ldots \lambda u'_{k'} : T''_k(\mu(\phi \land \psi)), (\pi_1, \ldots, \pi_i, \pi'_1, \ldots, \pi'_{k'}) \rangle \in T_3$

      ii. if $v_1$ is $\langle \lambda u_1 : T'_1 \ldots \lambda u_i : T'_i(\phi), (\pi_1, \ldots, \pi_i) \rangle$ and $v_2$ is a type (i.e. not of the form $(f, \Pi)$ for some function $f$ and sequence of paths $\Pi$), then $\langle \lambda u_1 : T'_1 \ldots \lambda u_i : T'_i(\mu(\phi \land v_2)), (\pi_1, \ldots, \pi_i) \rangle \in T_3$
iii. if \( v_2 \) is \( \lambda u_1':T''_1 \ldots \lambda u_k':T''_k(\psi), \langle \pi'_1 \ldots \pi'_k \rangle \) and \( v_1 \) is a type, then
\[
\langle \lambda u_1':T''_1 \ldots \lambda u_k':T''_k(\mu(v_1 \land \psi)), \langle \pi'_1 \ldots \pi'_k \rangle \rangle \in T_3
\]
iv. otherwise \( \langle \ell, \mu(v_1 \land v_2) \rangle \in T_3 \)

(b) for any \( \ell, v_1 \), if \( \langle \ell, v_1 \rangle \in T_1 \) and there is no \( v_2 \) such that \( \langle \ell, v_2 \rangle \in T_2 \),
then \( \langle \ell, v_1 \rangle \in T_3 \)

(c) for any \( \ell, v_2 \), if \( \langle \ell, v_2 \rangle \in T_2 \) and there is no \( v_1 \) such that \( \langle \ell, v_1 \rangle \in T_1 \),
then \( \langle \ell, v_2 \rangle \in T_3 \)

2. Otherwise \( \mu'(T_1 \land T_2) = T_1 \land T_2 \)

\( T_1 \land T_2 \) is used to represent \( \mu(T_1 \land T_2) \).

This definition of \( \mu \) differs from that given in [Cooper, 2008] in three respects. Firstly, it is not written in pseudocode and is therefore a better mathematical abstraction from the algorithm that has been implemented. Secondly, it includes the details of the treatment of dependencies within record types which were omitted from the previous definition. Finally, it excludes reference to a notion of subtype (\( \sqsubseteq \)) which was included in the previous definition. This could be changed by adding the following clauses at the beginning of the definition of \( \mu \) (after providing a characterization of the subtype relation, \( \sqsubseteq \)).

1. if for some \( T_1, T_2 \), \( T = T_1 \land T_2 \) and \( T_1 \sqsubseteq T_2 \) then \( \mu(T) = T_1 \)

2. if for some \( T_1, T_2 \), \( T = T_1 \land T_2 \) and \( T_2 \sqsubseteq T_1 \) then \( \mu(T) = T_2 \)

The current first clause would then hold in case neither of the conditions of these two clauses are met. The definition without these additional clauses only accounts for simplification of meets which have to do with merges of record types whereas the definition with the additional clauses would in addition have the effect, for example, that \( \mu(T \land T_0) = T_0 \) and \( \mu(T_1 \land (T_1 \lor T_2)) = T_1 \) (provided that we have an appropriate definition of \( \sqsubseteq \) ) whereas the current definition without the additional clauses means that \( \mu \) leaves these types unchanged.

**3.3 Models and modal systems of types**

Consider the definition of a system of complex types \( \text{TYPE}_C = \langle \text{Type}, \text{BType}, \langle \text{PType}, \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle, \langle A, F \rangle \rangle \) in section *2.4. We call the pair \( \langle A, F \rangle \) a model because of its similarity to first order models. A model for classical first order logic provides a domain \( A \) in which the logic is to be interpreted and an assignment \( F \) of values based on \( A \) to constants and predicates. That is: for any constant \( c \), \( F(c) \in A \); for a 1-place predicate \( P \), \( F(P) \subseteq A \); for a 2-place predicate \( R \), \( F(R) \subseteq A \times A \) and so on. Classical first order logic is not sorted, that is, \( A \) is just a simple set of objects in the domain. Sorted first order logic provides a family of sets of different sorts. We can think of \( A \) as a function which provides for each sort the objects which are of that sort. Predicates are then associated with an arity which tells us to which sort the arguments of the predicate should
belong. Our models are similar to these models for sorted first order logic with our basic types corresponding to the sorts. Models of first order logic provide a way of making arbitrary connections between the basic expressions of the logic and another domain. Intuitively, we can think of the domain as being a part of the “real world” consisting of objects like people and tables, particularly if we are interested in natural language semantics. But domains can also be mathematical objects like numbers or sets. The exact nature of the objects in the domain in first order models is not of concern to the logician. It could, for example, be defined as a collection of sensor readings which are available to a particular robot. The model provides an interface between the logical expressions and some domain of our choosing. In a similar way the models in our type theory provide an interface between the type theory and a system external to the type theory of our choosing: the “real world”, robot sensations or whatever.

The $F$ of our models behaves a little differently from that in first order models in that it assigns objects to types constructed from predicates rather than the predicates themselves. Suppose we have a type $P(a)$ constructed from a predicate $P$ and an object $a$ which is of an appropriate basic type as required by the arity of $P$. We could have made our models even closer to those of first order logic by having $F$ assign sets to predicates in the same way as in first order logic. We could mimic truth-values by introducing a distinguished basic type $Truth$ such that $A(Truth) = \{true\}$. We could then have said that $true : P(a)$ iff $a \in F(P)$ and no other object $b$ is such that $b : P(a)$. From the perspective of type theory, however, this seems like an odd thing to do, in part because the object $true$ seems like an odd thing to have to include in your domain, being an artificial object which has to belong to so many different types and in part because it is missing a fundamental type theoretical intuition that something of type $P(a)$ is whatever it is that counts as a proof object for the fact that $a$ falls under the predicate $P$. So instead of having $F$ assign a value to the predicate $P$ we have it assign a value to the type $P(a)$. The exact nature of the object which is obtained by applying $F$ to $P(a)$ depends on what kind of model you are working with. A basic intuition corresponding to the idea that models represent the “real world” is that it is a situation (i.e. a “bit of the real world”) which shows that $a$ falls under predicate $P$. But we could also define models in terms of robot sensations, four-dimensional space coordinates, databases, urls or whatever takes our fancy. The model is the place where we can connect our type theoretical system to some system external to the type theory. By moving from truth-values to this richer world of proof objects we have not lost the notion of truth. The “true” types are just those that are assigned a non-empty set of objects.

There is another important way in which our models are different from classical first order models. In first order logic the model relates two entirely separate domains: the syntactic expressions of first order logic and the model theoretic domain in which it is interpreted. The syntax of the logic is defined independently of the model. What counts as a well-formed expression does not depend in any way on what particular model we are using. This is not true of the models we
have used in our type theoretical systems. Suppose that we have a predicate \( P \) with arity \( \langle T_1 \rangle \). Suppose furthermore that our model is such that \( A(T_1) = \{ a \} \) and \( A(T_2) = \{ b \} \). Then it will be the case that \( P(a) \) is a type, but not \( P(b) \). If we had chosen a different model the set of types might have been different. This fact alone might lead some people to conclude that it is confusing and misleading to call \( \langle A, F \rangle \) a model in our type systems and certainly there is much of value in this point of view. The term ‘model’ is firmly entrenched in logic as that which provides the arbitrary parts of the interpretation of an independently defined syntactic language. However, we have chosen to persist in using the term here to emphasize the correspondence to models in logic and necessity of introducing an arbitrary link between a type theoretical system and an “external world” of some kind. The intuitive connection to models in logic is reinforced by our use of models in the discussion of modal systems below.

A modal system of complex types provides a collection of models, \( \mathcal{M} \), so that we can talk about properties of the whole collection of type assignments provided by the various models \( M \in \mathcal{M} \).

A modal system of complex types based on \( \mathcal{M} \) is a family of quadruples:

\[
\text{TYPE}_{MC} = \langle \text{Type}, \text{BType}, \langle \text{PType}, \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle, M \rangle_{M \in \mathcal{M}}
\]

where for each \( M \in \mathcal{M} \), \( \langle \text{Type}, \text{BType}, \langle \text{PType}, \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle, M \rangle \) is a system of complex types.

This enables us to define modal notions:

If \( \text{TYPE}_{MC} = \langle \text{Type}, \text{BType}, \langle \text{PType}, \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle, M \rangle_{M \in \mathcal{M}} \) is a modal system of complex types based on \( \mathcal{M} \), we shall use the notation \( \text{TYPE}_{MC} \) (where \( M \in \mathcal{M} \)) to refer to that system of complex types in \( \text{TYPE}_{MC} \) whose model is \( M \). Then:

1. for any \( T_1, T_2 \in \text{Type} \), \( T_1 \) is (necessarily) equivalent to \( T_2 \) in \( \text{TYPE}_{MC} \), \( T_1 \approx_{\text{TYPE}_{MC}} T_2 \), iff for all \( M \in \mathcal{M} \), \( \{ a \mid a : \text{TYPE}_{MC,M} T_1 \} = \{ a \mid a : \text{TYPE}_{MC,M} T_2 \} \)

2. for any \( T_1, T_2 \in \text{Type} \), \( T_1 \) is a subtype of \( T_2 \) in \( \text{TYPE}_{MC} \), \( T_1 \sqsubseteq_{\text{TYPE}_{MC}} T_2 \), iff for all \( M \in \mathcal{M} \), \( \{ a \mid a : \text{TYPE}_{MC,M} T_1 \} \subseteq \{ a \mid a : \text{TYPE}_{MC,M} T_2 \} \)

3. for any \( T \in \text{Type} \), \( T \) is necessary in \( \text{TYPE}_{MC} \) iff for all \( M \in \mathcal{M} \), \( \{ a \mid a : \text{TYPE}_{MC,M} T \} \neq \emptyset \)

4. for any \( T \in \text{Type} \), \( T \) is possible in \( \text{TYPE}_{MC} \) iff for some \( M \in \mathcal{M} \), \( \{ a \mid a : \text{TYPE}_{MC,M} T \} \neq \emptyset \)

*3.4 Strings and regular types

A string algebra over a set of objects \( O \) is a pair \( \langle S, \sqcup \rangle \) where:

1. \( S \) is the closure of \( O \cup \{ e \} \) (\( e \) is the empty string) under the binary operation \( \sqcup \) (“concatenation”)
2. for any $s$ in $S$, $e \rightarrow s = s \rightarrow e = s$

3. for any $s_1, s_2, s_3$ in $S$, $(s_1 \rightarrow s_2) \rightarrow s_3 = s_1 \rightarrow (s_2 \rightarrow s_3)$. For this reason we normally write $s_1 \rightarrow s_2 \rightarrow s_3$ or more simply $s_1 s_2 s_3$.

The objects in $S$ are called strings. Strings have length. $e$ has length 0, any object in $O$ has length 1. If $s$ is a string in $S$ with length $n$ and $a$ is an object in $O$ then $s \leftarrow a$ has length $n + 1$. We use $s[n]$ to represent the $n$th element of string $s$.

We can define types whose elements are strings. Such types correspond to regular expressions and we will call them regular types. Here we will define just two kinds of such types: concatenation types and Kleene-+ types.

A system of complex types $\text{TYPE}_C = \langle \text{Type}, \text{BType}, \langle \text{PType}, \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle, \langle \text{A,F} \rangle \rangle$ has concatenation types if

1. for any $T_1, T_2 \in \text{Type}$, $T_1 \rightarrow T_2 \in \text{Type}$

2. $a : T_1 \rightarrow T_2$ iff $a = x \leftarrow y$, $x : T_1$ and $y : T_2$

$\text{TYPE}_C$ has Kleene-+ types if

1. for any $T \in \text{Type}$, $T^+ \in \text{Type}$

2. $a : T^+$ iff $a = x_1 \rightarrow \ldots \rightarrow x_n$, $n > 0$ and for $i$, $1 \leq i \leq n$, $x_i : T$

Strings are used standardly in formal language theory where strings of symbols or strings of words are normally considered. Following important insights by Tim Fernando [Fernando, 2004; Fernando, 2006; Fernando, 2008; Fernando, 2009] we shall be concerned rather with strings of events. (We shall return to this in section 5.) We use informal notations like '“sam”' and '“ran”' to represent phonological types of speech events (utterances of $Sam$ and $ran$). Thus '“sam”$\rightarrow$“ran”' is the type of speech events which are concatenations of an utterance of $Sam$ and an utterance of $ran$.

*3.5 Grammar and compositional semantics

In order to illustrate how the content we have given for $ran$ in section 3.1 figures in a grammar and compositional semantics we shall define a toy grammar which covers the sentence $Sam$ ran.

We will present our grammar in terms of signs (in a similar sense to HPSG, see, for example, [Sag et al., 2003]). Our signs will be records of type $\text{Sign}$ which for present purposes we will take to be the type:

\[
\begin{bmatrix}
\text{s-event} & : & \text{SEvent} \\
\text{synsem} & : & \left[ \begin{array}{c}
\text{cat} & : & \text{Cat} \\
\text{cnt} & : & \text{Cnt} \\
\end{array} \right]
\end{bmatrix}
\]
We shall spell out the nature of the types \( SEvent \), \( Cat \) and \( Cnt \) below. A sign has two main components, one corresponding to the physical nature of the speech event (‘s-event’) and the other to its interpretation (syntax and semantics, ‘synsem’, using the label which is well-established in HPSG).

\( SEvent \) is the type

\[
\begin{align*}
\text{phon} & : \text{Phon} \\
\text{s-time} & : \begin{bmatrix}
\text{start} & : \text{Time} \\
\text{end} & : \text{Time}
\end{bmatrix} \\
\text{utt\_at} & : \langle \lambda v_1 : \text{Str} (\lambda v_2 : \text{Time} (\lambda v_3 : \text{Time} (\text{uttered\_at}(v_1,v_2,v_3)))), \\
& \langle \text{event\_phon}, \text{event\_s\_time\_start}, \text{event\_s\_time\_end} \rangle \rangle
\end{align*}
\]

In the s-event component the phon-field represents the phonology of an expression. Here we will take phonology as a string of word utterances although in a complete treatment of spoken language we would need phonological and phonetic attributes. That is, we take \( \text{Phon} \) to be \( \text{Wrd}^+ \) where \( \text{Wrd} \) (the type of word utterances) is defined in the lexicon. The s-time (“speech time”) field represents the starting and ending time for the utterance. We assume the existence of a predicate ‘uttered\_at’ with arity \( \langle \text{Phon}, \text{Time}, \text{Time} \rangle \). An object of type ‘uttered\_at(a,t_1,t_2)’ could be an event where \( a \) is uttered beginning at \( t_1 \) and ending at \( t_2 \) or a corresponding hypothesis produced by a speech recognizer with time-stamps, depending on the application of the theory. In a more complete treatment we would need additional information about the physical nature of the speech event, such as the identity of the speaker and where it took place.

In the synsem component the cat-field introduces a category for the phrase. For present purposes we will require that the following hold of the type \( \text{Cat} \):

\[
\text{s, np, vp, n\_prop, v}_i : \text{Cat}
\]

The objects of type \( \text{Cat} \) (s, np, vp etc.) are regarded as convenient abstract objects which are used to categorize classes of speech events.

The cnt-field represents the content or interpretation of the utterance. Since the content types become rather long we will introduce abbreviations to make them readable:

\[
\text{Ppty}, \text{“property” is to be} [x:\text{Ind}] \rightarrow \text{RecType} \\
\text{Quant}, \text{“quantifier” is to be} \text{Ppty} \rightarrow \text{RecType}
\]

We only use a small finite number of function types for content types and thus we are able to define the type \( \text{Cnt} \) for present purposes as

\[
\text{RecType} \vee (\text{Ppty} \vee \text{Quant})
\]

This makes use of join types which are defined in a similar way to meet types:

\[
\text{TYPE}_C = \langle \text{Type}, \text{BType}, \langle \text{PType}, \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle, \langle A, F \rangle \rangle \text{ has join types if}
\]

1. for any \( T_1, T_2 \in \text{Type} \), \( (T_1 \vee T_2) \in \text{Type} \)
2. for any \( T_1, T_2 \in \text{Type} \), \( a : \text{TYPE}_C (T_1 \lor T_2) \) iff \( a : \text{TYPE}_C T_1 \) or \( a : \text{TYPE}_C T_2 \)

We will present first the lexicon and then rules for combining phrases.

**Lexicon**

We will define lexical functions which tell us how to construct a type for a lexical item on the basis of a phonological type and either an object or a type corresponding to an observation of the world. The idea is that an agent which is constructing a grammar for use in a particular communicative situation will construct lexical types on the basis of a coordinated pair of observations: an observation of a speech event and an observation of an object or event with which the speech event is associated. This is related to the idea from situation semantics that meaning is a relation between an utterance situation and a described situation [Barwise and Perry, 1983]. The use of types here relates to the idea of type judgements as being involved in perception as discussed in section *2.3.

We shall use the following notation:

If \( W \) is a phonological type, then \( c_W \) is a distinguished label associated with \( W \), such that if \( W_1 \neq W_2 \) then \( c_{W_1} \neq c_{W_2} \).

We shall also make use of singleton types. \( \text{TYPE}_C = \langle \text{Type}, \text{BType}, \langle \text{PType}, \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle, \langle A, F \rangle \rangle \) has singleton types if

1. for any \( T \in \text{Type} \) and \( b : \text{TYPE}_C T, T_b \in \text{Type} \)
2. for any \( T \in \text{Type} \), \( a : \text{TYPE}_C T_b \) iff \( a : \text{TYPE}_C T \) and \( a = b \)

In the case of a singleton type \( T_x \) we allow a variant notation in records (corresponding to the manifest fields of Coquand et al., 2004) using

\[
[ \ell=x : T ]
\]

for

\[
[ \ell : T_x ]
\]

When we have a field

\[
[ \ell : \langle \lambda v_1 : T_1 \ldots \lambda v_n : T_n(T_x), \langle \pi_1 \ldots \pi_n \rangle \rangle ]
\]

we allow for convenience notations such as

\[
[ \ell=\langle \lambda v_1 : T_1 \ldots \lambda v_n : T_n\{x\}, \langle \pi_1 \ldots \pi_n \rangle \rangle : T ]
\]

\[
[ \ell=x : \langle \lambda v_1 : T_1 \ldots \lambda v_n : T_n(T), \langle \pi_1 \ldots \pi_n \rangle \rangle ]
\]

or

\[
[\ell=\langle \lambda v_1 : T_1 \ldots \lambda v_n : T_n\{x\}, \langle \pi_1 \ldots \pi_n \rangle; \lambda v_1 : T_1 \ldots \lambda v_n : T_n(T), \langle \pi_1 \ldots \pi_n \rangle\rangle]
\]
depending on how $T_x$ depends on $\pi_1 \ldots \pi_n$. We use \{ and \} to delimit $x$ since $x$ itself may be a function thus leading to ambiguity in the notation if we do not distinguish which $\lambda$’s represent dependency and which belong to the resulting object. Note that this ambiguity only arises in the notation we are adoptong for convenience.

**Proper names**

The most straightforward view of proper names is that they are based on pairings of proper noun utterances and individuals. While the full story about proper names may have to be more complex, this will suffice for our present purposes.

We define a function $\text{lex}_{\text{nprop}}$ which maps phonological types corresponding to proper names like *Sam* and individuals to record types, such that if $W$ is a phonological type such as “Sam” or “John” and $a:\text{Ind}$, $\text{lex}_{\text{nprop}}(W,a)$ is

$$
\text{Sign} \land
\begin{cases}
\text{s-event} : \\
\text{synsem} : \\
\end{cases}
\begin{align*}
\text{phon} & : W \\
\text{cnt} = & : \text{Cat} \\
\text{v} = & : \text{Quant}
\end{align*}
$$

The idea of this function is that an agent could have it as a resource to construct a lexical item for a local language on observing a pairing of a particular type of utterance (e.g. utterances of *Sam*) and a particular individual. If the language we are building is small enough there will be only one individual associated with a given phonological type such as “sam” but it is easy to imagine situations where there will be a need to have different individuals associated with the same name even within a local language, for example, if you need to talk about two people named *Sam* who write a book together. While this creates potential for misunderstanding there is nothing technically mysterious about having two lexical types which happen to share the same phonology. This is in contrast to the classical formal semantics view of proper names as related to logical constants where it seems unexpected that proper nouns should be able to refer to different individuals on different uses.

An example of a set of basic proper names which could be generated with these resources given two individuals $a$ and $b$ (that is, $a,b:\text{Ind}$) would be

$$
\{\text{lex}_{\text{nprop}}(\text{“Sam”},a), \\
\text{lex}_{\text{nprop}}(\text{“John”},b)\}
$$

**Intransitive verbs**

For intransitive verbs we will take the paired observations to involve a phonological type corresponding to an intransitive verb on the one hand and a predicate on the other. Philosophically, it may appear harder to explain what it means to observe a predicate compared to observing an individual, even though if you dig
deep enough even individuals are problematical. However, it seems that any reasonable theory of perception should account for the fact that we perceive the world in terms of various kinds of objects standing in relations to each other. Our predicates correspond to these relations and we would want to say that our cognitive apparatus is such that relations are reified in a way that they need to be in order to become associated with types of utterances. For a verb like run we will say that the predicate is one that holds between individuals and time intervals. We will argue in section 4 that for other verbs we need frames instead of individuals.

We define a function lex, which maps phonological types corresponding to intransitive verbs like run and predicates with arity \( \langle \text{Ind}, \text{TimeInt} \rangle \), such that if \( W \) is a phonological type like “run” or “walk” and \( p \) is a predicate with arity \( \langle \text{Ind}, \text{TimeInt} \rangle \), \( \text{lex}_V(W, p) \) is

\[
\text{Sign} \land \\
s\text{-event:} \left[ \begin{array}{c} \text{phon:} \ W \\ \text{cat=} v_i : \text{Cat} \end{array} \right] \\
\text{synsem:} \left[ \begin{array}{c} \text{cnt=} \lambda r : [x : \text{Ind}] \\ \text{e-time:} \langle \text{TimeInt} \\ c_W : \langle \lambda v: \text{TimeInt} (p(r.x,v), \langle \text{e-time} \rangle) \rangle \rangle : \text{Ppty} \end{array} \right]
\]

Similar remarks hold for this function as for the one we used for proper names. For different local languages different predicates may be associated with utterances of run and even within the same local language, confusing though it may be, we may need to associate different predicates with different occurrences of run. In this way verbs are like proper names and one can think of verbs as proper names of predicates.

However, this is not quite enough if we want to handle different forms of verbs such as infinitives, and present and past tenses. For purposes of simplification as our concern is not with the details of morphological types we will assume that all finite verb occurrences are third person singular and will not represent these features. In order to achieve this we need to define \( \text{lex}_V \) not in terms of a single phonological type but a paradigm of phonological types corresponding to different configurations of morphological features. For present purposes we will think of there just being one morphological feature of tense which can take the values: inf (“infinitive”), pres (“present tense”), past (“past tense”). We will think of paradigms as functions which map records of type \([\text{tns: Tns}]\) to phonological types. Here the type \( \text{Tns} \) has elements inf, pres and past. Let \( \text{run} \) be the paradigm for run. The function is defined by

\[
\begin{align*}
\text{run}(\langle \text{tns=} \text{inf} \rangle) &= \text{“run”} \\
\text{run}(\langle \text{tns=} \text{pres} \rangle) &= \text{“runs”} \\
\text{run}(\langle \text{tns=} \text{past} \rangle) &= \text{“ran”}
\end{align*}
\]

and for \( \text{walk} \) we have

\[
\begin{align*}
\text{walk}(\langle \text{tns=} \text{inf} \rangle) &= \text{“walk”} \\
\text{walk}(\langle \text{tns=} \text{pres} \rangle) &= \text{“walks”} \\
\text{walk}(\langle \text{tns=} \text{past} \rangle) &= \text{“walked”}
\end{align*}
\]
In order to obtain the interpretations of the tensed forms of the verb we will need the following functions for present and past tense.

**Pres** which is to be \( \lambda t: \text{TimeInt} (\text{e-time: TimeInt} (\lambda v: \text{TimeInt}(v = t), \langle \text{e-time} \rangle)) \)

**Past** which is to be \( \lambda t: \text{TimeInt} (\text{e-time: TimeInt} (\lambda v: \text{TimeInt}(v.end < t.start), \langle \text{e-time} \rangle)) \)

The present tense function expresses that the event time is identical with the interval to which it is being compared. This is normally the speech time as in the grammar defined here, though it could also be a different time interval, for example in the interpretation of historic presents. The past tense function expresses that the end of the event time interval has to be prior to the start of the interval (e.g. the speech time) with which it is being compared.

We need also to make the distinction between finite and non-finite verb utterances and we will do this by introducing a field labelled ‘fin’ which will take values in the type \( \text{Bool} \) ("boolean") whose members are 0 and 1.

Now we redefine \( \text{lex}_V \) to be a function which takes a paradigm \( W \) such as run or walk, a predicate \( p \) with arity \( \langle \text{Ind}, \text{TimeInt} \rangle \) and morphological record \( m \) of type \([\text{tns:Tns}]\) such that

1. if \( m \) is \([\text{tns}=\text{inf}]\), \( \text{lex}_V(W, p, m) \) is

\[
\begin{align*}
\text{Sign} \land \\
\text{s-event:} & [\text{phon:} W(m)] \\
\text{cat} & = v_1 : \text{Cat} \\
\text{fin} & = 0 : \text{Bool} \\
\text{cnt} & = \lambda r: [x: \text{Ind}] (\text{e-time: TimeInt} \\
& \langle c_{W(m)}: \lambda v: \text{TimeInt}(p(r.x,v)), \langle \text{e-time} \rangle \rangle) : \text{Ppty}
\end{align*}
\]

2. if \( m \) is \([\text{tns}=\text{pres}]\), \( \text{lex}_V(W, p, m) \) is

\[
\begin{align*}
\text{Sign} \land \\
\text{s-event:} & [\text{phon:} W(m)] \\
\text{s-time:} & \text{TimeInt} \\
\text{cat} & = v_1 : \text{Cat} \\
\text{fin} & = 1 : \text{Bool} \\
\text{cnt} & = (\lambda v_1: \text{Time}\{\lambda r: [x: \text{Ind}] \\
& \text{e-time: TimeInt} \\
& \langle c_{W(m)}: \lambda v_2: \text{TimeInt}(p(r.x,v_2)), \langle \text{e-time} \rangle \rangle \} \land \text{Pres}(v_1)) \\
\langle \text{s-event.s-time} \rangle & : \text{Ppty}
\end{align*}
\]

3. if \( m \) is \([\text{tns}=\text{past}]\), \( \text{lex}_V(W, p, m) \) is
An example of a set of intransitive verbs which could be generated with these resources given appropriate predicates ‘run’ and ‘walk’ is

\[ \bigcup_{\alpha \in \{\text{inf}, \text{pres}, \text{past}\}} \{ \text{lex}_V(\text{run}, \text{run}, [\text{tns} = \alpha]), \text{lex}_V(\text{walk}, \text{walk}, [\text{tns} = \alpha]) \} \]

**Syntactic and semantic composition**

We will think of composition rules as functions which take a string of utterances of various types and return a type for the whole string. That is, the basic form of our composition rules will be:

\[ \lambda s : T_1(T_2) \]

where \( T_1 \) is a type of strings of signs and \( T_2 \) is a type of signs. More specifically we can say that unary rules are functions of the form

\[ \lambda s : T_1(T_2), \text{ where } T_1, T_2 \subseteq \text{Sign} \]

and binary rules are of the form

\[ \lambda s : T_1 \overset{\text{temp}}{\prec} T_2(T_3), \text{ where } T_1, T_2, T_3 \subseteq \text{Sign} \]

\( \overset{\text{\textasciitilde}}{\text{\textasciitilde}} \) here denotes the subtype relation defined in section 3.3. (We are suppressing the subscript used there.) We can, of course, generalize these notions to \( n \)-ary rules but unary and binary will be sufficient for our present purposes.

Note that to say that there is a string of signs \( s_1 \overset{\text{temp}}{\prec} s_2 \) does not necessarily mean that the signs are temporally ordered in the sense that \( s_1.s\text{-event}.s\text{-time}.\text{end} < s_2.s\text{-event}.s\text{-time}.\text{start} \). There could be an advantage in this for the treatment of discontinuous constituents or free word order. But we can also define a special “temporal concatenation” type for concatenation of signs:

A system of complex types \( \text{TYPE}_C = \langle \text{Type}, \text{BType}, \langle \text{PType}, \text{Pred}, \text{ArgIndices}, \text{Arity} \rangle, \langle A, F \rangle \rangle \) has temporal concatenation types for the type Sign if

1. for any \( T_1, T_2 \subseteq \text{Sign} \), \( T_1 \overset{\text{temp}}{\prec} T_2 \in \text{Type} \)
2. \( s : T_1 \text{~temp~} T_2 \) iff \( s = s_1 \text{~} s_2, s_1 : T_1, s_2 : T_2 \) and \( s_1.s\text{-event}.s\text{-time.end} < s_2.s\text{-event}.s\text{-time.start} \).

We will factor our rules into component functions which we will then combine in order to make a complete rule. The components we will use here are:

**unary\_sign** which we define to be

\[ \lambda s: \text{Sign}(\text{Sign}) \]

This takes any sign and returns the type \( \text{Sign} \)

**binary\_sign** which we define to be

\[ \lambda s: \text{Sign}\text{~temp~}\text{Sign}(\text{Sign}) \]

This takes any temporal concatenation of two signs and returns the type \( \text{Sign} \)

**phon\_id** which we define to be

\[ \lambda s: [s\text{-event}:[\text{phon}:=\text{Phon}}][] ([s\text{-event}:[\text{phon}=s.s\text{-event}.\text{phon}:=\text{Phon}}][]) \]

This takes any record \( s \) of type \( [s\text{-event}:[\text{phon}:=\text{Phon}}][] \) and returns a type which is the same except that the phonology field is now required to be filled by the value of that field in \( s \).

**phon\_concat** which we define to be

\[ \lambda s: [s\text{-event}:[\text{phon}:=\text{Phon}}][] \text{~temp~} [s\text{-event}:[\text{phon}:=\text{Phon}}][] \]

\[ ([s\text{-event}:[\text{phon}=s[1].s\text{-event}.\text{phon}~\text{temp~} s[2].s\text{-event}.\text{phon}:=\text{Phon}}][]) \]

This takes a string of two records with phonology fields and returns the type of a single record with a phonology field whose value is required to be the concatenation of the values of the phonology fields in the first and second elements of the string.

**unary\_cat** which we define to be

\[ \lambda c_1: \text{Cat}(\lambda c_2: \text{Cat}(\lambda s: [\text{cat}=c_1:\text{Cat}])] ([\text{cat}=c_2:\text{Cat}])) \]

This takes two categories and returns a function which maps a record with a category field with value the first category to a type of records with a category field which is required to be filled by the second category.

**binary\_cat** which we define to be

\[ \lambda c_1: \text{Cat}(\lambda c_2: \text{Cat}(\lambda c_3: \text{Cat}(\lambda s: [\text{cat}=c_1:\text{Cat}]) [\text{cat}=c_2:\text{Cat}] [\text{cat}=c_3:\text{Cat}])) \]
This takes three categories and returns a function which maps a string of two records with a category field with values identical to the respective categories to a type of records with a category field which is required to be filled by the third category.

\textbf{cnt\_id} which we define to be

\[ \lambda s: [\text{synsem:}\{\text{cnt:Cnt}\}][\{\text{synsem:}\{\text{cnt}=s.\text{synsem.cnt:Cnt}\}\}]. \]

This takes any record \(s\) of type \([\text{synsem:}\{\text{cnt:Cnt}\}]\) and returns a type which is the same except that the content field is now required to be filled by the value of that field in \(s\).

\textbf{cnt\_forw\_app} which we define to be

\[ \lambda T_1: \text{Type}(\lambda T_2: \text{Type}(\lambda s: [\text{synsem:}\{\text{cnt:T}_1 \rightarrow T_2\}][\text{synsem:}\{\text{cnt:T}_1\}][\text{synsem:}\{\text{cnt}=s[1].\text{synsem.cnt(s[2].\text{synsem.cnt):T}_2\}]]). \]

This takes any binary string of records \(s\) such that the content of the first record is a function which takes arguments of a type to which the content of the second record belongs and returns a type whose content field is now required to be filled by the result of applying the content of the first record to the content of the second record.

\textbf{fin\_id} which we define to be

\[ \lambda s: [\text{fin:Bool}][\{\text{fin}=s.\text{fin:Bool}\}]. \]

This requires that the value of a ‘fin’-field will be copied into the new type (corresponding to feature percolation in a non-branching tree in a more traditional feature-based grammar).

\textbf{fin\_hd} which we define to be

\[ \lambda s: \text{Sign}^\sim [\text{fin}=1:\text{Bool}][\{\text{fin}=s.\text{fin:Bool}\}]. \]

This requires that the second sign in a string of two has a positive specification for finiteness and copies it into the new type.

We will use the notion of merge defined in section *3.2 in the characterization of how these component functions are to be combined in order to form rules. Since the combination of these functions is so closely connected to the merge operation we will use a related symbol ‘\(\wedge\)’ with two dots rather than one. In the following definition we will use \(T_i\) to represent types which are not string types and \(v\) to represent an arbitrary variable.

1. \(\lambda v:T_1(T_2) \wedge \lambda v:T_3(T_4)\) is to be \(\lambda v:T_1\wedge T_3(T_2\wedge T_4)\)

2. \(\lambda v:T_1\wedge T_2(T_3) \wedge \lambda v:T_4\wedge T_5(T_6)\) is to be \(\lambda v:(T_1\wedge T_4)^\sim(T_2\wedge T_5)(T_3\wedge T_6)\)
3. \( \lambda v: T_1 \neg \neg \text{temp} T_2 T_3 \wedge \lambda v: T_4 \neg \neg \text{temp} T_5 T_6 \) is to be \( \lambda v: (T_1 \wedge T_4) \neg \neg \text{temp} (T_2 \wedge T_5) (T_3 \wedge T_6) \)

Since \( \wedge \), like \( \wedge \), is associative we will write \( f \wedge g \wedge h \) instead of \( (f \wedge g) \wedge h \) or \( f \wedge (g \wedge h) \).

Now we can use the rule components we have defined to express the three rules we need for this small fragment.

\[
S \rightarrow NP \ VP \\
\text{binary_sign} \wedge \text{phon_concat} \wedge \text{binary_cat(np)(vp)(s)} \wedge \text{fin_hd} \\
\wedge \text{cnt_forw_app}(Ppty)(RecType)
\]

\[
NP \rightarrow N \\
\text{unary_sign} \wedge \text{phon_id} \wedge \text{unary_cat(n_prop)(np)} \wedge \text{cnt_id}
\]

\[
VP \rightarrow V_i \\
\text{unary_sign} \wedge \text{phon_id} \wedge \text{unary_cat(v_i)(vp)} \wedge \text{fin_id} \wedge \text{cnt_id}
\]

This gives us a concise way to express rather complex functions corresponding to simple rules. The point of this is, however, not merely to give us yet another formalism for expressing natural language phrase structure and its interpretation but to show how such rules can be broken down into abstract components which an agent learning the language could combine in order to create rules which it has not previously had available in its resources. Thus an agent (such as a child in the one-word stage) which does not have a rule \( S \rightarrow NP \ VP \) but who observes strings of linguistic events where NP’s are followed by VP’s may reason its way to a rule that combine NP-events followed by VP-events into a single event. While this concerns linguistic events it is closely related to the way we take strings of non-linguistic events to form single events, for example, a going-to-bed-event for a child might normally consist of a string of events having-hot-milk\( \neg \neg \)putting-on-pyjamas\( \neg \neg \)getting-into-bed\( \neg \neg \)listening-to-a-story. Our general ability to perceive events, that is, assign types to events and to combine these types into larger event types seems to be a large part of the basis for our linguistic ability. We will return to this in our discussion of Fernando’s string theory of events in section 5.

4 USING FRAMES TO SOLVE A CLASSICAL PUZZLE ABOUT TEMPERATURE AND PRICES

4.1 The Partee puzzle

Montague [1973] introduces a puzzle presented to him by Barbara Partee:

From the premises the temperature is ninety and the temperature rises, the conclusion ninety rises would appear to follow by normal principles of logic; yet there are occasions on which both premises are true, but none on which the conclusion is.
Exactly similar remarks can be made substituting *price* for *temperature*. Montague’s [1973] solution to this puzzle was to analyze *temperature, price* and *rise* not as predicates of individuals as one might expect but as predicates of individual concepts. For Montague individual concepts were modelled as functions from possible worlds and times to individuals. To say that *rise* holds of an individual concept does not entail that *rise* holds of the individual that the concepts finds at a given world and time. Our strategy is closely related to Montague’s. However, instead of using individual concepts we will use frames. By interpreting *rises* as a predicate of frames of type *AmbTemp* as given in (2) we obtain a solution to this puzzle.

\[
\lambda r: [x: \text{Ind}](\begin{array}{l}
e\text{-time} : \text{TimeInt} \\
c_{\text{tns}} : e\text{-time}=t \\
c_{\text{run}} : \text{rise}(r, e\text{-time})
\end{array})
\]

Note that a crucial difference between (10) and (11) is that the first argument to the predicate ‘rise’ is the complete frame \( r \) rather than the value of the x-field which is used for ‘run’. Thus it will not follow that the value of the x-field (i.e. 90 in Montague’s example) is rising. While there is a difference in the type of the argument to the predicates (a record as opposed to an individual), the type of the complete verb content is the same: \([x: \text{Ind}] \rightarrow \text{RecType}\), that is, a function from records of type \([x: \text{Ind}]\) to record types.

*4.2 Additions to the grammatical resources*

The aim of this section is to add to the resources described in section *3.5* so that we can analyze sentences such as *the temperature rises* and *the price rises*.

**Lexicon**

**Intransitive verbs**

The ability to use different types internally but still have the same overall type for the content of the word means that we can incorporate verbs that take frame arguments into the lexicon without having to change the rest of the grammar resources. We add a paradigm *rise*:

\[
\begin{align*}
\text{rise}([\text{tns}=\text{inf}]) &= \text{“rise”} \\
\text{rise}([\text{tns}=\text{pres}]) &= \text{“rises”} \\
\text{rise}([\text{tns}=\text{past}]) &= \text{“rose”}
\end{align*}
\]

We now introduce a lexical function \( \text{lex}_{\text{V}_{1-\text{fr}}} \) to be a function which takes a paradigm \( \mathcal{W} \) corresponding to a verb whose predicate takes a frame argument, such as *rise*, a predicate \( p \) with arity \( ([x: \text{Ind}], \text{TimeInt}) \) and morphological record \( m \) of type \([\text{tns}: \text{Tns}]\) such that
1. if \( m \) is \([tns=\text{inf}]\), \( \text{lex}_{V\rightarrow t}(W, p, m) \) is

\[
\begin{align*}
\text{Sign} \land \\
\begin{array}{l}
\text{s-event:} \quad \text{phon:} W(m) \\
\text{cat:} v_1: \text{Cat} \\
\text{fin:} 0: \text{Bool} \\
\text{synsem:} \\
\begin{array}{l}
\text{cnt:} \lambda r: \{ x: \text{Ind} \} \left( c_{W(m)}: \langle \lambda v: \text{TimeInt}(p(r, v)), \langle \text{e-time} \rangle \rangle \right); \text{Ppty}
\end{array}
\end{array}
\end{align*}
\]

2. if \( m \) is \([tns=\text{pres}]\), \( \text{lex}_{V\rightarrow t}(W, p, m) \) is

\[
\begin{align*}
\text{Sign} \land \\
\begin{array}{l}
\text{s-event:} \quad \text{phon:} W(m) \\
\text{s-time:} \text{TimeInt} \\
\text{cat:} v_1: \text{Cat} \\
\text{fin:} 1: \text{Bool} \\
\text{synsem:} \\
\begin{array}{l}
\text{cnt:} \lambda v_1: \text{Time}\{ \lambda r: \{ x: \text{Ind} \} \\
\left( c_{W(m)}: \langle \lambda v_2: \text{TimeInt}(p(r, v_2)), \langle \text{e-time} \rangle \rangle \right) \land \text{Pres}(v_1) \}, \\
\langle \text{s-event.s-time} \rangle: \text{Ppty}
\end{array}
\end{array}
\end{align*}
\]

3. if \( m \) is \([tns=\text{past}]\), \( \text{lex}_{V\rightarrow t}(W, p, m) \) is

\[
\begin{align*}
\text{Sign} \land \\
\begin{array}{l}
\text{s-event:} \quad \text{phon:} W(m) \\
\text{s-time:} \text{TimeInt} \\
\text{cat:} v_1: \text{Cat} \\
\text{fin:} 1: \text{Bool} \\
\text{synsem:} \\
\begin{array}{l}
\text{cnt:} \lambda v_1: \text{Time}\{ \lambda r: \{ x: \text{Ind} \} \\
\left( c_{W(m)}: \langle \lambda v_2: \text{TimeInt}(p(r, v_2)), \langle \text{e-time} \rangle \rangle \right) \land \text{Past}(v_1) \}, \\
\langle \text{s-event.s-time} \rangle: \text{Ppty}
\end{array}
\end{array}
\end{align*}
\]

An example of a set of lexical intransitive verb types which could now be generated with these resources given appropriate predicates 'run', 'walk' and 'rise' is

\[
\bigcup_{\alpha \in \{\text{inf, pres, past}\}} \{ \text{lex}_{V_i}(\text{run}, \text{run}, [\text{tns}=\alpha]), \\
\text{lex}_{V_i}(\text{walk}, \text{walk}, [\text{tns}=\alpha]), \\
\text{lex}_{V\rightarrow t}(\text{rise}, \text{rise}, [\text{tns}=\alpha]) \}
\]

**Common nouns**

In our treatment of common nouns we will make the same distinction that we made for intransitive verbs between predicates that take individual arguments and those that take frame arguments.

We define a function \( \text{lex}_N \) which maps phonological types corresponding to common nouns like \textit{dog} and predicates with arity \((\text{Ind, TimeInt})\), such that if \( W \)
is a phonological type like “dog” and \( p \) is a predicate with arity \( \langle \text{Ind}, \text{TimeInt} \rangle \), \( \text{lex}_N(W, p) \) is

\[
\text{Sign} \wedge \\
\begin{aligned}
\text{s-event:} & \ [\text{phon:}W] \\
\text{cat=}: & \ \text{Cat} \\
\text{synsem:} & \ [\text{cnt=}\lambda r: [x: \text{Ind}] \ (e:\text{time:} \text{TimeInt} (c_W: (\lambda v: \text{TimeInt}(p(r, x, v), (e:\text{time})))), (\text{e-time})))]: \text{Ppty}
\end{aligned}
\]

We define a function \( \text{lex}_{N-fr} \) which maps phonological types corresponding to common nouns like \( \text{temperature} \) and \( \text{price} \) and predicates with arity \( \langle [x: \text{Ind}], \text{TimeInt} \rangle \), such that if \( W \) is a phonological type like “temperature” or “price” and \( p \) is a predicate with arity \( \langle [x: \text{Ind}], \text{TimeInt} \rangle \), \( \text{lex}_{N-fr}(W, p) \) is

\[
\text{Sign} \wedge \\
\begin{aligned}
\text{s-event:} & \ [\text{phon:}W] \\
\text{cat=}: & \ \text{Cat} \\
\text{synsem:} & \ [\text{cnt=}\lambda r: [x: \text{Ind}] \ (e:\text{time:} \text{TimeInt} (c_W: (\lambda v: \text{TimeInt}(p(r, x, v), (e:\text{time})))), (\text{e-time})))]: \text{Ppty}
\end{aligned}
\]

An example of a set of lexical common noun types which could be generated given appropriate predicates ‘dog’, ‘temperature’ and ‘price’ is

\[
\{\text{lex}_N(“dog”, \text{dog}), \text{lex}_{N-fr}(“temperature”, \text{temperature}), \text{lex}_{N-fr}(“price”, \text{price})\}
\]

Determiners

We define a function \( \text{lex}_{\text{Det-ex}} \) for the indefinite article which maps a phonological type like \( a \) to a sign type such that if \( W \) is a phonological type \( \text{lex}_{\text{Det-ex}}(W) \) is

\[
\text{Sign} \wedge \\
\begin{aligned}
\text{s-event:} & \ [\text{phon:}W] \\
\text{cat=}: & \ \text{Det:Cat} \\
\text{cnt=} & \ \lambda v_1: \text{Ppty} \\
\text{synsem:} & \ (\lambda v_2: \text{Ppty} \\
& \ (\text{par:} [x: \text{Ind}] \\
& \ (\text{restr:} (\lambda v: [x: \text{Ind}] (v_1(v)), (\text{par}))) \\
& \ (\text{scope:} (\lambda v: [x: \text{Ind}] (v_2(v)), (\text{par}))) )]: \text{Ppty}\rightarrow \text{Quant}
\end{aligned}
\]

We define a function \( \text{lex}_{\text{Det-unl}} \) for the universal determiner \( \text{every} \) which maps a phonological type such as “every” to a sign type such that if \( W \) is a phonological type then \( \text{lex}_{\text{Det-unl}}(W) \) is
We define a function \( \text{lex}_{\text{Det-def}} \) which maps phonological types to a sign type for the definite article \( \text{the} \) such that if \( W \) is an appropriate phonological type then \( \text{lex}_{\text{Det-def}}(W) \) is

\[
\begin{align*}
\text{Sign} \wedge \\
\text{s-event:} & [\text{phon:} W] \\
\text{cnt:} & \lambda v_1 : \text{Ppty} \\
\text{cat:} & \text{Det:} \text{Cat} \\
(\lambda v_2 : \text{Ppty}) & \\
\text{synsem:} & \\
& \\
& \left( \begin{array}{c}
\text{par:} [x : \text{Ind}] \\
\text{rest:} \langle \lambda v : [x : \text{Ind}] (v_1(v)) \rangle \\
\text{scope:} \langle \text{par} \rangle \\
\end{array} \right) : \text{Ppty} \rightarrow \text{Quant}
\end{align*}
\]

An example of a set of lexical determiner types that could be generated with these resources is

\[
\{ \text{lex}_{\text{Det-ex}}(\text{"a"}), \text{lex}_{\text{Det-unl}}(\text{"every"}), \text{lex}_{\text{Det-def}}(\text{"the"}) \}
\]

This is a classical treatment of quantification which uses existential quantification similar to that used in classical DRT [Kamp and Reyle, 1993] where existential quantification introduces a discourse referent corresponding to our par(arameter)-field. The indefinite article introduces a discourse referent corresponding to the common noun following the determiner and a scope field representing the scope of the quantifier, the verb phrase if the noun phrase built with the determiner is in subject position. The result of applying the content of the determiner to two properties will be a type which requires there to be an individual which meets the conditions provided by both the restriction and the scope.

Also similar to classical DRT is the use of dependent functions (as defined on p. 283) for universal quantification. The type resulting from the application of
the determiner content to two properties requires that there be a function from
individuals meeting the restriction type to a proof that these individuals also meet
the restriction. The use of dependent function types for universal quantification is
a classical strategy in the application of type theory to natural language semantics.
[Sundholm, 1986; Ranta, 1994, are examples of discussion of this].

The definite article content combines the content of both the existential and
the universal quantifier contents in the kind of Russellian treatment of definite
descriptions that Montague proposed. Applying this content to two properties
will return a type which requires that there is some individual which has the
first property and that anything which has this property is identical with this
individual (i.e. there is exactly one individual with this property) and furthermore
the individual also has the second property.

This treatment of quantification (like Montague’s) does not use generalized
quantifier relations even though the determiner contents are functions which apply
to two properties. [Cooper, 2010b, contains discussion of this issue].

Syntactic and semantic composition

We need one additional rule to combine determiners and nouns into noun phrases.
This rule is similar to the rule combining noun phrases and verb phrases into
sentences except that it uses different categories and content types and lacks the
finite head requirement.

\[
\text{NP} \rightarrow \text{Det N}
\]

\[
\begin{align*}
\& \text{binary\_sign} \land \text{phon\_concat} \land \text{binary\_cat(det)(n)(np)} \\
\& \land \text{cnt\_forw\_app(Ppty)(Quant)}
\end{align*}
\]

5 LEXICAL SEMANTICS, FRAMES AND FERNANDO EVENT STRINGS

5.1 Frames that rise

In the previous section we proposed to solve the Partee puzzle by allowing pred-
icates such as ‘rise’ to take frames as arguments. But now the question arises:
what can it mean for a frame to rise?

In an important series of papers including [Fernando, 2004; Fernando, 2006;
Fernando, 2008; Fernando, 2009], Fernando introduces a finite state approach to
event analysis where events can be seen as strings of punctual observations corre-
sponding to the kind of sampling we are familiar with from audio technology and
digitization processing in speech recognition. When talking about the intuition
behind this analysis Fernando sometimes refers to strings of frames in a movie
(e.g. in [Fernando, 2008]). But in many cases what he is calling a movie frame
can also be seen as a frame in the sense of this paper as well. Thus an event of a
rise in temperature could be seen as a concatenation of two temperature frames,
that is, an object of type \( \text{AmbTemp} \upharpoonright \text{AmbTemp} \). (12) shows a type of event for a
rise in temperature using the temperature frame \( \text{AmbTemp} \) in (2).
We will call this type TempRise. Now we can say something more precise about the content of rise expressed in (11). Recall that this introduces a predicate ‘rise’ which takes a frame and a time interval as arguments. Combining ‘rise’ with two arguments creates a type of event in which the frame “rises” during the time interval. We suggest that a candidate for such an event is an object of type TempRise. We will express this by

\[
\text{if } r: \text{AmbTemp} \text{ and } i: \text{TimeInt} \text{ then } e: \text{rise}(r, i) \text{ iff } e: \text{TempRise}, e.\text{start}=r \\
\text{and } e.e.-\text{time}=i.
\]

This is at most a very partial account of what objects belong to the types which are constructed from the predicate ‘rise’. It is limited to ambient temperature frames. It does not tell us what it would mean for any other kind of frame to rise. This incompleteness is, we believe, an important part of a cognitive theory based on type theory.

Our idea is to exploit an important aspect of the formal development of type theory in a cognitive theory of concept acquisition. We want to say that concepts are modelled by types in type theory. In the formal treatment of type theory when we introduce a new type or a predicate which combines with arguments to form a type there are always two things that have to be done. Firstly the type or predicate itself has to be introduced and we have to say what the type is or how types can be constructed using the predicate. Secondly we have to say what objects belong to the type(s) we have introduced. In the cognitive theory we want to say that we are often (perhaps standardly) in the situation where we have a type or predicate in our cognitive resources (that is, we have performed the counterpart of the first part of type introduction) but we have only a partial idea of what it means to be of the type(s) introduced (that is, we have not been able to complete the second part of the introduction). In fact, we will argue below that at least in the case of concepts corresponding to word meaning we can never be sure that we have a complete account of what it means to belong to the corresponding types.

Thus suppose that we have an agent who has just observed an utterance of the sentence *the temperature rises* and that the utterance of the word *rises* was the first time that the agent had heard this word. From various pieces of evidence the agent may be able to figure out that this is an intransitive verb, for example from
the present tense morphology and its position in the sentence. This will provide the agent with enough information to construct a predicate ‘rise’ given the linguistic resources at the agent’s disposal and will enable the agent to conclude that the content of the verb is (11), possibly with a question mark over whether the first argument to the predicate is the whole frame or the x-component of the frame. It is perhaps safer to assume first the more general case where the argument is the whole frame unless there is evidence for the more specific case.

If the agent is at a loss for what was being communicated this might be as far as she can get. That is, she will know that there is a predicate signified by *rise* but she will not have any idea of what it means for an event to fall under a type which is constructed with this predicate. However, there is very often other evidence besides the speech event which will give clues as to what it might mean for a temperature to rise. The agent, herself, may have noticed that it is getting hotter. The speaker of the utterance may indicate by wiping their brow or fanning their face as they speak that this is what is meant by *rise*. (Just think of the kind of gesticulations that often accompany utterances made to non-native speakers of a language with only basic competence.) Thus the agent may come to the account we have presented of what rising involves for ambient temperature situations. A question is: will this account generalize to other kinds of situations?

### 5.2 Word meaning in flux

For all (12) is based on a very much simplified version of FrameNet’s Ambient_temperature, it represents a quite detailed account of the lexical meaning of *rise* in respect of ambient temperature — detailed enough, in fact, to make it inappropriate for *rise* with other kinds of subject arguments. Consider price. The type of a price rising event could be represented by (13).

<table>
<thead>
<tr>
<th>e-time: TimeInt</th>
</tr>
</thead>
<tbody>
<tr>
<td>x: Ind</td>
</tr>
<tr>
<td>e-time = e-time.start: Time</td>
</tr>
<tr>
<td>e-location: Loc</td>
</tr>
<tr>
<td>commodity: Ind</td>
</tr>
<tr>
<td>c-price_of_at_in: price_of_at_in(start.commodity, start.e-time, start.e-location, start.x)</td>
</tr>
</tbody>
</table>

(13) is similar to (12) but crucially different. A price rising event is, not surprisingly, a string of price frames rather than ambient temperature frames. The
type of price frames \((\text{Price})\) is given in (14).

\[
\begin{bmatrix}
x & : & \text{Ind} \\
\text{e-time} & : & \text{Time} \\
\text{e-location} & : & \text{Loc} \\
\text{commodity} & : & \text{Ind} \\
\text{c-price\_of\_at\_in} & : & \text{price\_of\_at\_in(\text{commodity, e-time, e-location, x})}
\end{bmatrix}
\]

If you look up the noun \(\text{price}\) in FrameNet\(^7\) you find that it belongs to the frame \text{Commerce\_scenario} which includes frame elements for goods (corresponding to our ‘commodity’) and money (corresponding to our ‘x’-field). If you compare the FrameNet frames \text{Ambient\_temperature} and \text{Commerce\_scenario}, they may not initially appear to have very much in common. However, extracting out just those frame elements or roles that are relevant for the analysis of the lexical meaning of \text{rise} shows a degree of correspondence. They are, nevertheless, not the same. Apart from the obvious difference that the predicate in the constraint field that relates the various roles involves temperature in the one and price in the other, price crucially involves the role for commodity since this has to be held constant across the start and end frames. We cannot claim that a price is rising if we check the price of tomatoes in the start frame and the price of oranges in the end frame.

The fact that we need a different meaning for \text{rise} depending on whether the subject is a temperature or a price corresponds to a situation which is familiar to us from work on the Generative Lexicon \cite{Pustejovsky, 1995; Pustejovsky, 2006} where the arguments to words representing functions influence the precise meaning of those words. For example, \text{fast} means something different in \text{fast car} and \text{fast road}, although, of course, the two meanings are related. There are two important questions that arise when we study this kind of data:

- is it possible to extract a single general meaning of words which covers all the particular meanings of the word in context?
- is it possible to determine once and for all the set of particular contextually determined meanings?

Our suspicion is that the answer to both these questions is “no”. How exactly should we fill out our analysis in order to get the correct meaning of \text{rise} for prices? Is it sufficient to check the price at just two points of time? If not, how do we determine how many points need to be checked? Should we place restrictions on the time-span between points which are checked and if so, how can we go about determining the kind of time-span involved? Do we need to check that the rising is monotonic, that is, that there is no point in the period we are checking that the price is lower than it was at an earlier time in the period? And then there is the matter of how space is involved in the meaning. If I say \text{The price of tomatoes is rising} do I mean the price of tomatoes in a particular shop, a particular city, region

\(^7\) accessed 8th April, 2010
or in general wherever tomatoes are sold? This seems like a pragmatic dependence on context. But suppose we have determined a region we are interested in. Does the price of tomatoes have to be rising in every shop selling tomatoes in that region or for every kind of tomato? If not, what percentage of the tomatoes in the region need to be going up in price in order for the sentence to be true? This is perhaps a matter having to do with vagueness or generic interpretations. Then there are more technical questions like: is the price rising if it is keeping pace with some recognized index of inflation? Well, it depends what you mean by *rise*. Can the price of tomatoes be said to be rising if it stays the same during a period of deflation?

It seems unlikely that we could tie down the answer to all of these questions once and for all and characterize the meaning of *rise*. The techniques we have for dealing with context dependence and vagueness may account for some of the apparent variability, but in the end surely we have to bite the bullet and start building theories that come to grips with the fact that we adjust the meanings of words to fit the purposes at hand.

It seems that we are able to create new meanings for words based on old meanings to suit the situation that we are currently trying to describe and that there is no obvious requirement that all these meanings be consistent with each other, making it difficult to extract a single general meaning. Here we are following the kind of theory proposed by Larsson and Cooper [Larsson and Cooper, 2009; Cooper and Larsson, 2009]. According to such a theory the traditional meaning question “What is the meaning of expression $E$?” should be replaced by the following two questions relating to the way in which agents coordinate meaning as they interact with each other in dialogue or, more indirectly, through the writing and reading of text:

**the coordination question** Given resources $R$, how can agent $A$ construct a meaning for a particular utterance $U$ of expression $E$?

**the resource update question** What effect will this have on $A$’s resources $R$?

Let us look at a few examples of uses of the verb *rise* which suggest that this is the kind of theory we should be looking at. Consider first that a fairly standard interpretation of *rise* concerns a change in location. (15) is part of the description of a video game.

(15) As they get to deck, they see the Inquisitor, calling out to a Titan in the seas. The giant Titan rises through the waves, shrieking at the Inquisitor.

The type of the rising event described here could be something like (16).

This relies on a frame type \textit{Position} given in (17).

\begin{align*}
\text{(17)}\quad & \begin{bmatrix}
\quad x & : & \text{Ind} \\
\quad \text{e-time} & : & \text{Time} \\
\quad \text{e-location} & : & \text{Loc} \\
\quad \text{c}_{\text{at}} & : & \text{at}(x,\text{e-location},\text{e-time})
\end{bmatrix}
\end{align*}

(17) is perhaps most closely related to FrameNet’s \textit{Locative relation}. (16) is structurally different from the examples we have seen previously. Here the content of the \textquote{\textasciitilde{}x\textasciitilde{}}-field, the focus of the frame, which in the case of the verb \textit{rise} will correspond to the subject of the sentence, is held constant in the string of frames in the event whereas in the case of rising temperatures and prices it was the focus that changed value. Here it is the height of the location which increases whereas in the previous examples it was important to hold the location constant.\footnote{We have used \textquote{\textasciitilde{}height(start/end.e-location)} in (16) to represent the height of the location since we have chosen to treat \textit{Loc}, the type of spatial location, as a basic type. However, in a more detailed treatment \textit{Loc} should itself be treated as a frame type with fields for three coordinates one of them being height, so we would be able to refer to the height of a location \textit{l} as \textit{l.height}.} This makes it difficult to see how we could give a single type which is general enough to include both varieties and still be specific enough to characterize \textquote{the meaning of \textit{rise}}. It appears more intuitive and informative to show how the variants relate to each other in the way that we have done.

The second question we had concerned whether there is a fixed set of possible meanings available to speakers of a language or whether speakers create appropriate meanings on the fly based on their previous experience. Consider the examples in (18).

\begin{align*}
\text{(18)}\quad & \begin{array}{ll}
\text{a.} & \text{Mastercard rises} \\
\text{b.} & \text{China rises}
\end{array}
\end{align*}

While speakers of English can get an idea of the content of the examples in (18) when stripped from their context, they can only guess at what the exact content...
might be. It feels like a pretty creative process. Seeing the examples in context as in (19) reveals a lot.¹⁰

(19) a. Visa Up on Q1 Beat, Forecast; Mastercard Rises in Sympathy
By Tiernan Ray
Shares of Visa (V) and Mastercard (MA) are both climbing in the aftermarket, reversing declines during the regular session, after Visa this afternoon reported fiscal Q1 sales and profit ahead of estimates and forecast 2010 sales growth ahead of estimates, raising enthusiasm for its cousin, Mastercard.

b. The rise of China will undoubtedly be one of the great dramas of the twenty-first century. China’s extraordinary economic growth and active diplomacy are already transforming East Asia, and future decades will see even greater increases in Chinese power and influence. But exactly how this drama will play out is an open question. Will China overthrow the existing order or become a part of it? And what, if anything, can the United States do to maintain its position as China rises?

It seems like the precise nature of the frames relevant for the interpretation of rises in these examples is being extracted from the surrounding text by a technique related to automated techniques of relation extraction in natural language processing. We know from (19a) that Mastercard rises means that the share price of Mastercard has been going up. We have, as far as I can see, no clear way of determining whether this involves an adjustment to the meaning of rise or of Mastercard. It seems that no harm would arise from both strategies being available to an agent. (19b) is interesting in that the text preceding China rises prepares the ground so that by the time we arrive at it we have no trouble figuring out that what is meant here by rise has to do with economic growth, active diplomacy, power and influence.

Consider (20) out of context.

(20) dog hairs rise

Unless you know this example from the British National Corpus it is unlikely that you would get at the appropriate meaning for rise which becomes obvious when we look at some of the context as in (21).

(21) Cherrilyn: Yeah I mean (pause) dog hairs rise any-
way so
Fiona: What do you mean, rise?
Cherrilyn: The hair (pause) it rises upstairs.

BNC file KBL, sentences 4201–4203

(21) is an example of a clarification request (as discussed recently in [Ginzburg, forthcoming], and much previous literature cited there). Given that the meaning of *rise* does not appear to be fixed, we might expect that a lot of such clarification requests would occur. This, however, does not appear to be the case. Out of 205 occurrences of *rise* as a verb (in any of its inflectional forms\(^{11}\)) in the dialogue subcorpus of BNC there is one occurrence of a clarification, namely (21). It seems then that there is no evidence that *rise* is particularly hard to understand, which certainly seems to accord with intuition. It does seem, however, that human speakers are particularly adept at adjusting meaning to suit the needs of the current situation.

6 TOWARDS A THEORY OF SEMANTIC COORDINATION

It seems that we are constantly in the process of creating and modifying language as we speak. This phenomenon has been studied and theorized about for at least 25 years by psychologists of language, for example [Clark and Wilkes-Gibbs, 1986; Garrod and Anderson, 1987; Brennan and Clark, 1996; Healey, 1997; Pickering and Garrod, 2004]. However, in semantics (both formal and empirical) there is a tradition of abstracting away from this fact for rather obvious reasons. In formal semantics there is an obvious advantage of assuming that natural languages have a fixed semantics like formal languages in order to get started building a theory of compositional semantics. In empirically based semantics like frame semantics based on corpus data the point is to make statistically based generalizations over whatever varieties of language might be contained within the corpus. However, recently a number of linguists have become interested in trying to account for the way in which language gets created or adjusted through dialogue interaction. For some examples see [Cooper and Kempson, 2008].

One of the important facts to emerge from the psychological research is that dialogue participants tend to coordinate (or align) their language. For example, if you use a particular word for a particular concept, I will tend to use that word for that concept unless there is good reason for me do otherwise (such as I believe that your way of saying it is incorrect or I wish to mark that I speak a different dialect from you). If I use a different word there will be a tendency for my interlocutor to assume that I am referring to a different concept all else being equal. [Clark,
1993] refers to this as the principle of contrast. Similarly, if you use a word with what is for me an innovative meaning, I need to find a way of constructing that meaning so that either we can continue the dialogue using that meaning or we can negotiate what the word should mean. That is, we need to engage in semantic coordination [Larsson, 2007a; Larsson, 2007b]. Suppose you use the word rise with a meaning that is new for me. How should I go about figuring out what that meaning should be? One thing that seems clear is that I do not start from scratch, considering the space of all possible intransitive verb meanings. Rather I start from meanings I have previously associated with utterances of rise and try to modify them to fit the current case. The structured approach to meaning presented by TTR becomes very important here. Suppose that I have a meaning for rise of the classical Montague semantics kind, that is, a function from possible worlds and times to characteristic functions of sets of objects. The kinds of modifications that are natural to such an object could, for example, involve adding or subtracting objects which are associated with a particular world and time. Such modifications are not particularly useful or intuitive in helping us to figure out the answers to the kinds of questions about the meaning of rise. In contrast TTR provides us with components that can be modified, parameters which can be used to characterize variation in meaning and serve as a basis for a similarity metric. Components can be modified in order to create new meanings from old.

Our idea is that this view of semantics should be embedded in the kind of view of agents that coordinate linguistic resources which is presented in [Cooper and Larsson, 2009]. We will review the ideas about agents presented there which are in turn based on Larsson’s earlier work.

We conceive the theory as being within the gameboard approach to dialogue developed by Ginzburg [Ginzburg, 1994; Ginzburg, forthcoming, and much other literature in between] and the computationally oriented approach based on Ginzburg’s work which has come to be known as the information state update approach [Larsson and Traum, 2001, and much other literature developing from these ideas]. Here dialogue participants have information states associated with the dialogue which are updated by dialogue moves which they perceive as having been carried out by their interlocutors or themselves. The kind of update which this literature has normally been concerned with have to do with both informational content and metalinguistic information about what was said. Informational content includes, for example, propositions which have been committed to in the dialogue and questions under discussion. Metalinguistic information includes information about phonology and syntax as well as what content is to be associated with various parts of the utterance. This provides us with a basis for dealing with part of a theory of semantic coordination. In addition we need to be able to talk about updates to linguistic resources available to the agent (grammar, lexicon, semantic interpretation rules etc. in the sense discussed in [Cooper and Ranta, 2008]) which can take place during the course of a dialogue. The view presented in [Cooper and Larsson, 2009] is that agents have generic resources which they modify to construct local resources for sublanguages for use in specific situations.
Thus an agent $A$ may associate a linguistic expression $e$ with a particular concept (or collection of concepts if $e$ is ambiguous) $[e]^A$ in its generic resource. In this paper, we will think of $[e]^A$ not as a collection of concepts but as one of the sign types that we introduced above. In a particular domain $\alpha$ $e$ may be associated with a modified version of $[e]^A$, $[e]^A_\alpha$ [Larsson, 2007a].

The motor for generating new local resources in an agent lies in coordinating resources with another agent in a particular communicative situation $s$. In a communicative situation $s$, an agent $A$ may be confronted with an innovative utterance $e$, that is, a speech event which contains uses of linguistic expressions not already present in $A$’s resources or linguistic expressions from $A$’s resources which in $s$ are associated with an interpretation distinct from that provided by $A$’s resources. In the theory we have presented above either of these cases will involve the construction of a new sign type which is specific to $s$, $[e]^A_s$, and which may be anchored to the specific objects under discussion in $s$ (using the technique of manifest fields).

Whereas in a standard view of formal grammar there will be one sign (or in our terms sign type) corresponding to a particular interpretation of an expression, we want to see $e$ as related to a whole hierarchy of sign types: $[e]^A_s$ for communicative situations $s$, $[e]^A_\alpha$ for domains $\alpha$ (where we imagine that the domains are collected into a complex hierarchy or more and less general domains) and ultimately a general linguistic resource which is domain independent, $[e]^A$. We think of the acquisition of a particular sign type as a progression from $[e]^A_s$ for some particular communicative situation $s$, through potentially a series of increasingly general domains $\alpha$ yielding resources $[e]^A_\alpha$. In [Cooper and Larsson, 2009] we regarded complete acquisition process as ultimately leading to a domain independent generic resource, $[e]^A$. However, the more one thinks in these terms the more likely it seems that there is no ultimate domain independent resource at all (except perhaps for “logical” words like determiners) but rather a large collection of resources associated with domains of varying generality.

There is no guarantee that any sign type will survive even beyond the particular communicative situation in which $A$ first encountered it. For example, the kind of ad hoc coinages described in [Garrod and Anderson, 1987] using words like leg to describe part of an oddly shaped maze in the maze game probably do not survive beyond the particular dialogue in which they occur. The factors involved in determining how a particular sign-type progresses we see as inherently stochastic with parameters including the degree to which $A$ regards their interlocutor as an expert, how many times the sign type has been exploited in other communicative situations and with different interlocutors, the utility of the interpretation in different communicative situations, and positive or negative feedback obtained when exploiting the sign type in a communicative situation. For example, a particular agent may only allow a sign type to progress when it has been observed in at least $n$ different communicative situations at least $m$ of which were with an interlocutor considered to be an expert, and so on.

On this view the kind of question we need to be addressing in a formal linguistic
theory is not so much “What is the meaning of rises (with respect to price)?” but rather “How will agent $A$ with access to a resource $[\text{rises}]_\alpha^A$ (for domain $\alpha$) exploit this resource in a given communicative situation $s$?”. Here we assume that exploiting a resource standardly involves modifying it so that it matches the purposes at hand. The tradition that we have inherited from logical semantics has given us the idea of truth conditions and determining whether a given sentence is true under the fixed interpretation provided by the language. Here we are also allowing for the option of modifying the interpretation of a sentence so that it would be true in the current state of affairs. If $A$ says $\phi$ to $B$ in respect of situation $s$ and is being honest and is not mistaken about the nature of $s$, then $A$ must be interpreting $\phi$ in such a way that it correctly describes $s$ and it is part of $B$’s task to figure out what this interpretation might be. This is the task of semantic coordination. The challenge for $B$ is to figure out whether there is a reasonable interpretation of $\phi$ (not too different from an interpretation that could be achieved by the resources already at $B$’s disposal based on previous experience) or whether $A$ is in fact mistaken about the nature of $s$ and is saying something false. It seems that much of the misunderstanding that occurs in dialogue can be related to this delicate balancing act that is required of dialogue participants.

We will try to be a little more concrete about what kind of modifications can be made to resources. For the sake of argument we could say that $[\text{rises}]_\alpha^A$ is the type which is produced by the grammar we have defined above. The main opportunities for modification here lie in determining what kind of Fernando event strings constitute a rising. If we are talking about price the question is more specifically what strings constitute a rising in price. Since according to this resource the relevant strings are strings of price frames, modifications here may also have consequences for the type of price frames provided by the resource. For example, an agent might soon notice that location is an important parameter for prices (the price might be rising in Europe but not in China, for example). This would mean that strings of price frames constituting a rising could now be required to contain the same location. This is what we have represented in our type for price rising events.

Suppose that at a given location new higher quality tomatoes become available on the market in addition to the tomatoes that were already available which are still available at the same price. Has the price of tomatoes risen? In one sense, yes, the average price of tomatoes has risen. In another sense, no, people who want to buy the tomatoes they were buying before can continue to do so for the same price. To get the first of these meanings we need to include information about averages in a more detailed price frame. For the second of these meanings we could introduce a parameter for quality into price frames and require that quality be held constant in rising events. These are additions which we may well not want to make part of any general language resource - they are rather ad hoc adjustments for the situation at hand. Suppose now that the cheap tomatoes disappear from the market but the more expensive tomatoes are still available for the same price. Again, if what you mean by price is the average price then the price has risen. But actually there are no tomatoes on the market such that they have gone up
in price. So you could argue (and people do) that prices have not gone up (using price frames with the quality parameter). However, people who need tomatoes for their pasta sauce and who used to buy the cheap tomatoes will now notice a rise in price greater than the average price rise. Whereas they used to get tomatoes for the cheap price they now have to pay the expensive price. For them, the price of tomatoes has risen. Here we seem to need price frames which accommodate a range of prices for a given commodity, for example a price record that specifies the highest and lowest prices, and a characterization of rising in terms of the lowest price. Again, this is an ad hoc modification which will be useful for some dialogues but not for others. Once you have figured it out it might be useful to keep in your collection of resources in case you need it again.

An important part of this discussion is that in order to figure out what is meant in a particular speech event we need to match potential interpretations against what we can observe about the world. We observe that A uses $\phi$ to describe situation $s$ and thereby draw conclusions about what A meant by this particular utterance of $\phi$ as well as gaining information about $s$. Perhaps one of the most straightforward examples of this connection is in language acquisition situations where one agent indicates particular objects for another agent saying things like *This is a...* The challenge for an agent in this learning situation is not so much to determine whether the utterance is true as trying to construct an appropriate meaning for the utterance to make it true and storing this as a resource for future use. Getting this coupling between language and perception is, for example, one of the first challenges in getting a robot to learn language through interaction (see for example the roadmap presented by the ITalk project, http://www.italkproject.org/).

## 7 CONCLUSION

One of the great advances in the study of language during the twentieth century was the application of the theory of formal languages to natural language. Challenges for the study of language in the twenty-first century are to extend the formal approach to the study of

1. interaction in dialogue

2. language coordination, including the creation or modification of linguistic resources during the course of a dialogue

3. the relationship of these processes to other cognitive processes such as perception

During the first decade of the century we have made some significant progress on (1), for example, [Ginzburg, forthcoming]. We have also made a start on (2), for example, [Larsson, 2007a; Larsson, 2007b; Cooper and Larsson, 2009]. TTR plays a significant role in this literature. TTR might also be useful in addressing (3) in
Type Theory and Semantics in Flux

that type theory is a theory about type judgements which from a cognitive point of view has to do with how we perceive objects.

It is important that we do not lose what we gained during the twentieth century when we are working with these new challenges and we believe that by using the tools provided by TTR it is plausible that we can keep and improve on the twentieth century canon.

Type theory is appealing for application to the new challenges because it makes the connection between perception and semantics and, with records, provides us with the kind of structure (like frames) that we seem to need for semantic coordination, giving us handles (in the forms of labelled fields) to items of knowledge in a structure rather than the monolithic functions of classical model theoretic semantics.

ACKNOWLEDGEMENTS

I am grateful to Raquel Fernández, Jonathan Ginzburg, Ruth Kempson, Staffan Larsson and Bengt Nordström for comments. I am particularly grateful to Tim Fernando for very detailed comments that led to a major rewriting. This work was supported in part by a grant from Vetenskapsrädet, Library-based Grammar Engineering (2005-4211), The Swedish Bank Tercentenary Foundation Project P2007/0717, Semantic Coordination in Dialogue and VR project 2009-1569, Semantic analysis of interaction and coordination in dialogue (SAICD).

BIBLIOGRAPHY

Robin Cooper


[Cooper and Ranta, 2008] Robin Cooper and Aarne Ranta. Natural Languages as Collections of Resources. In Cooper and Kempson [2008], pages 109–120.


Type Theory and Semantics in Flux


LANGUAGE, LINGUISTICS AND COGNITION

Giosuè Baggio, Michiel van Lambalgen, and Peter Hagoort

1 INTRODUCTION

Experimental research during the last few decades has provided evidence that language is embedded in a mosaic of cognitive functions. An account of how language interfaces with memory, perception, action and control is no longer beyond the scope of linguistics, and can now be seen as part of an explanation of linguistic structure itself. However, although our view of language has changed, linguistic methodology is lagging behind. This chapter is a sustained argument for a diversification of the kinds of evidence applicable to linguistic questions at different levels of theory, and a defense of the role of linguistics in experimental cognitive science.

1.1 Linguistic methodology and cognitive science

At least two conceptual issues are raised by current interactions between linguistics and cognitive science. One is whether the structures and rules described by linguists are cognitively real. There exist several opinions in this regard, that occupy different positions on the mentalism/anti-mentalism spectrum. At one extreme is cognitive linguistics [Croft and Cruse, 2004], endorsing both theoretical and methodological mentalism. The former is the idea that linguistic structures are related formally and causally to other mental entities. The latter calls for a revision of traditional linguistic methodology, and emphasizes the role of cognitive data in linguistics. At the opposite side of the spectrum lies formal semantics which, partly inspired by Frege’s anti-psychologistic stance on meaning and thought [Frege, 1980; Lewis, 1970; Burge, 2005], rejects both versions of mentalism. Somewhere between the two poles is Chomsky’s [Chomsky, 1965] theoretical mentalism, which sees linguistic rules as ultimately residing in the brain of speakers. However, his commitment to the cognitive reality of grammar does not imply a revision of linguistic methodology, which is maintained in its traditional form based on native speakers’ intuitions and the competence/performance distinction.

The second problem, in part dependent on the first, is whether experimental data on language acquisition, comprehension and production have any bearing on linguistic theory. On this point too, there is no consensus among linguists. The division between competence and performance has often been used to secure linguistics from experimental evidence of various sorts [Bunge, 1984], while
intuitive judgments of native speakers were regarded as the only type of data relevant for the theory [Chomsky, 1965]. However, some authors have granted that at least behavioral data should be allowed to inform competence theories, for instance if the linguist is studying a language which is not her own native language [Marantz, 2005]. Others have proposed frameworks in which competence can be connected to performance mechanisms [Jackendoff, 2002]. But while these models account for how competence constrains performance [Jackendoff, 2007], they seem to overlook the possibility that the reverse is also the case: aspects of linguistic structure may be the outcome of evolutionary processes leading to an adaptation of the brain to language use, that is to performance [Pinker and Jackendoff, 2005]. Generative grammar and formal semantics have regarded accounts of competence as shielded from data provided by experimental psychology and neuroscience. A more inclusive attitude has been adopted by psycholinguists and cognitive brain scientists, driven by an increasing demand of theories and models that would account for their data [Carminati et al., 2000; Featherston et al., 2000; Geurts and van der Slik, 2005; McKinnon and Osterhout, 1996; McMillan et al., 2005]. Despite these attempts, a methodological framework relating linguistics, language processing and low-level neural models is still missing.

1.2 Language, lower and higher cognition

Most theories in cognitive linguistics and neuroscience regard language as grounded in perception and action. For instance, cognitive semanticists have proposed that the meanings of concrete nouns stored in memory include stereotyped visual-geometric representations of the entities they refer to [Jackendoff, 1987]. Analogously, representations of action verbs might embrace aspects of the relevant motor programs [Hagoort, 1998]. It has also been suggested that the building blocks of semantics like the predicate-argument structure originate in the functional and anatomical organization of the visual and auditory systems [Hurford, 2003]. Experimental work in cognitive neuroscience indicates that language has ties with the sensory-motor systems, but methodology, specific data points and accounts of how exactly language connects to ‘lower’ cognition are still debated [Pulvermüller et al., 2001; Pulvermüller, 2005; Pulvermüller et al., 2005; Ferreira and Patson, 2007; Haslam et al., 2007; Hurford, 2007; Papafragou et al., 2008; Toni et al., 2008; Taylor et al., 2008]. The interested reader may want to follow further these references: in this chapter we will focus on language and ‘higher’ cognitive domains such as planning and reasoning. A motivation for this choice is that planning and reasoning shed light on the computation of complex linguistic structures, which is where language really comes into its own, whereas looking at the interactions between language and sensory-motor systems may give us more insights into representations and processes at the lexical level.

It has been proposed that the recursive organization of plans supplies a mechanism for combinatorial operations in grammar [Steedman, 2002], and the goal-directed nature of planned actions constrains cognitive constructions of time,
causality and events, with consequences for the semantics of tense, aspect and modality [Steedman, 2002; van Lambalgen and Hamm, 2004]. Planning might as well be implicated in the production and comprehension of discourse. Language processing requires adjusting the current discourse model incrementally given the input. If further information counters earlier commitments or expectations, a recomputation of the initial discourse model may be necessary to avoid inconsistencies. This process is best accounted for by the non-monotonic logic underlying planning, and more generally executive function, where the chosen action sequence may be readjusted if obstacles are encountered along the way.

On a par with planning, reasoning is of special interest in this chapter. Some have seen non domain-specific thought and reasoning as the most sophisticated among the cognitive skills subserved by language [Carruthers, 1996; Carruthers, 2002]. This suggestion is sometimes implicit in logical approaches to language since Boole [Boole, 1958, Ch. 2, p. 24] and bears some resemblance to the psycholinguistic notion that reasoning follows and builds upon interpretation [Johnson-Laird, 1980; Singer, 1994]. In this perspective, interpretation equals deriving logical (often classical) form from a sentence’s surface structure for subsequent elaboration involving inference. So there is a one-way dependency of reasoning from interpretation: interpretation supports reasoning, though not vice versa. Others have seen the relation between interpretation and inference as based on a two-way interaction [Stenning and van Lambalgen, 2008]: reasoning is involved in computing a model of what is said and in deriving conclusions from it. Human communication is thus regarded as the foremost skill enabled by language, and reasoning serves the purpose of coordinating different interpretations of an utterance or different situation models across speakers [Stenning, 2003].

2 LINGUISTICS AND COGNITIVE DATA

Let us address in more detail the issues anticipated in section 1.1. In what follows we will present Chomsky’s early views on linguistic methodology, introducing a paradox that we believe still lurks in current thinking about relations between linguistics and cognitive data. We will argue that the main problems with the competence/performance distinction are how ‘performance’ is defined, and what a theory of performance is supposed to include. We will show how this, and the use of intuitions as the only source of evidence in linguistic practice, constitutes an obstacle to a deeper integration of linguistics within cognitive science. These critical sections will be followed by a more constructive part (2.3-2.4), in which Marr’s three level scheme is proposed as a replacement of and, we will suggest, an improvement over competence/performance.

2.1 A methodological paradox

It is often said that the relations between cognitive science and linguistics began to be fully appreciated only after the publication of Chomsky’s early writings, and
in particular *Aspects of the Theory of Syntax* in 1965. This is certainly true if what is at stake is theoretical mentalism – the notion that linguistic theory deals ultimately with a system of representations and rules in the speaker’s mind/brain. However, although this particular form of theoretical mentalism encourages and to some extent requires some interaction between the two disciplines, the choice of regarding the study of competence as in principle indifferent to the results of experimental research had rather the opposite effect, that of separating theories of meaning and grammar from models of language processing. Many would agree that the contacts between linguistics and cognitive psychology have not been as deep and systematic as they could have been, had various obstacles to fruitful interaction been removed. What is more difficult to appreciate is the existence of a tension in the very foundation of generative grammar, and the inhibiting effect that tension had on the growth of linguistics within cognitive science. Before we move on, it may be worth recovering the terms of this ‘paradox’ directly from Chomsky’s text.¹

One side of the dilemma is represented by a number of remarks contained in §1 of the first chapter of *Aspects*, where Chomsky writes:

> We thus must make a fundamental distinction between *competence* (the speaker-hearer’s knowledge of his language) and *performance* (the actual use of language in concrete situations). Only under [...] idealization [...] is performance a direct reflection of competence. In actual fact, it obviously could not directly reflect competence. A record of natural speech will show numerous false starts, deviations from rules, changes of plan in mid-course, and so on. The problem for the linguist, as well as for the child learning the language, is to determine from the data of performance the underlying system of rules that has been mastered by the speaker-hearer and that he puts to use in actual performance. Hence, in the technical sense, linguistic theory is mentalistic, since it is concerned with discovering a mental reality underlying actual behavior. [Chomsky, 1965, p. 4]

The task of the linguist is that of providing an account of competence based on performance data, that is on normalized records of linguistic behavior. Chomsky grants that performance data are essential to linguistic theorizing. But the issue to be settled, which in fact lies at the heart of the paradox, is exactly what counts as linguistic behavior, or more precisely what kind of performance data can constitute the empirical basis of competence theories. Generative linguists would contend that it was never a tenet of their research program to admit data other than native speakers’ intuitions, but this is not what Chomsky’s remarks suggest. On the contrary, he seems to admit a variety of data types:

¹Over the years Chomsky has entertained different opinions on these issues. Here we choose to focus on those expressed in *Aspects of the Theory of Syntax* [Chomsky, 1965] because these have probably been the most influential. So we identify Chomsky with this particular text rather than with the actual linguist.
Mentalistic linguistics is simply theoretical linguistics that uses performance as data (along with other data, for example, the data provided by introspection) for the determination of competence, the latter being taken as the primary object of its investigations. [Chomsky, 1965, p. 193]

The evidential base of linguistics consists of introspective judgments and performance data, that Chomsky mentions here as if they were in an important sense different from intuitions. Moreover, intuitions are alluded to here as a subsidiary source of evidence, and as part of a larger class of data types. The question is precisely what should be considered performance data. Is elicited and experimentally controlled behavior allowed to exert some influence on accounts of competence? There are reasons to believe that Chomsky would have answered affirmatively, the most important of which has to do with his remarks on the limits of intuitions. In 1955, in *The Logical Structure of Linguistic Theory* [Chomsky, 1955] he wrote:

> If one of the basic undefined terms of linguistic theory is ‘intuition’, and if we define phonemes in this theory as elements which our intuition perceives in a language, then the notion of phoneme is as clear and precise as is ‘intuition’. [...] It should be clear, then, why the linguist interested in constructing a general theory of linguistic structure, in justifying given grammars or (to put the matter in its more usual form) in constructing procedures of analysis should try to avoid such notions as ‘intuition’. [Chomsky, 1955, pp. 86-87]

An even more explicit position was expressed in the 1957 book *Syntactic Structures*, where Chomsky suggests that hypotheses on properties of linguistic strings and their constituents should be evaluated on the basis of controlled operational tests. Relying on native speaker’s judgments or intuitions, he writes,

> amounts to asking the informant to do the linguist’s work; it replaces an operational test of behavior (such as the pair test) by an informant’s judgment about his behavior. The operational tests for linguistic notions may require the informant to respond, but not to express his opinion about his behavior, his judgment about synonymy, about phonemic distinctness, etc. The informant’s opinions may be based on all sorts of irrelevant factors. This is an important distinction that must be carefully observed if the operational basis for grammar is not to be trivialized. [Chomsky, 1957, pp. 8-9][2]

Controlled operational tests are thus necessary in order to overcome the difficulties arising from relying exclusively on native speakers’ intuitions. This implies that

---

[2] The circularity which Chomsky is alluding to here is also mentioned by Quine in his 1970 paper on linguistic methodology: “We are looking for a criterion of what to count as the real or proper grammar, as over against an extensionally equivalent counterfeit. [...] And now the test suggested is that we ask the native the very question we do not understand ourselves: the very question for which we ourselves are seeking a test. We are moving in an oddly warped circle.” [Quine, 1970, p. 392].
introspective data are dismissed as an inadequate source of evidence for linguistic theory. So here is one horn of the dilemma: mentalistic linguistics rejects speakers’ intuitions and requires performance data, including controlled behavioral tests, to constrain the theory of competence.

The other side of the paradox is represented by a series of remarks in §4 of chapter 1 of *Aspects*, where Chomsky questions the nature of the empirical basis of competence theories:

There is, first of all, the question of how one is to obtain information about the speaker-hearer’s competence, about his knowledge of the language. Like most facts of interest and importance, this is neither presented for direct observation nor extractable from data by inductive procedures of any known sort. Clearly, the actual data of linguistic performance will provide much evidence for determining the correctness of hypotheses about underlying linguistic structure, along with introspective reports (by the native speaker, or the linguist who has learned the language). [Chomsky, 1965, p. 18]

Experimental research based on controlled observation and statistical inference is seen as providing facts of no ‘interest and importance’, and rejected as ineffective for the purposes of the theory of competence. Interestingly, intuitions are treated as if they were on a par with performance data. Not for long, however, because Chomsky a few paragraphs later takes an important step away from psychology:

The critical problem for grammatical theory today is not a paucity of evidence but rather the inadequacy of present theories of language to account for masses of evidence that are hardly open to serious question. The problem for the grammarian is to construct a description and, where possible, an explanation for the enormous mass of unquestionable data concerning the linguistic intuition of the native speaker (often, himself); the problem for one concerned with operational procedures is to develop tests that give the correct results and make the relevant distinctions. […] We may hope that these efforts will converge, but they must obviously converge on the tacit knowledge of the native speaker if they are to be of any significance. [Chomsky, 1965, pp. 19-20]

The range of data that could affect the theory of competence has been narrowed down to intuitions, and more specifically to those of the linguist. The task of experimental research, Chomsky says, is to develop tests that would ultimately align with introspective data. The convergence of linguistics and psychology is thus projected forward in time as a desirable outcome not of the joining of efforts, but of their strict segregation. Not only are linguistics and psychology now regarded as separate enterprises, but psychology is also required – in order to meet a standard of explanatory adequacy – to provide results that are consistent with the theory of competence as based on the linguist’s intuitions. The second horn of the dilemma
is thus the following: linguistic theory is based primarily on the intuitions of native speakers, and does not require controlled experimentation to constrain accounts of competence.

2.2 The vagaries of intuition

For some linguists, in particular in generative grammar and formal semantics, the intuitions of native speakers constitute the empirical basis of the theory of competence. But the prominent place assigned to intuitions by modern linguistic methodology seems at odds with the principles of mentalism. If competence is a system of rules and structures realized in the speaker’s brain, and if behavior reflects the functioning of such system, then a linguist constructing a competence theory – and perhaps analogously a child learning a language – must solve an ‘inverse problem’, that of inferring the rules of competence from observable performance. In order to solve this problem, the linguist might need to take into account a broad range of data. So any reliable physiological or behavioral measure of performance should, at least in principle, be allowed to contribute to the theory of competence. The question is where should one draw a line between relevant (intuitions?) and irrelevant (neurophysiology?) data, and why. Until convincing answers are found, it would seem that the more comprehensive one’s methodological framework, the better. Here is why mentalism is to be preferred over traditional philosophies of language.

The conflict with mentalism is however not the only problem raised by introspective judgments. Another source of concern is Chomsky’s claim that intuitions are not only the starting point of linguistic theorizing, but also the standard to which any grammar should conform:

A grammar can be regarded as a theory of language; it is descriptively adequate to the extent that it correctly describes the intrinsic competence of the idealized native speaker. The structural descriptions assigned to sentences by the grammar, the distinctions that it makes between well-formed and deviant, and so on, must, for descriptive adequacy, correspond to the linguistic intuition of the native speaker (whether or not he may be immediately aware of this) in a substantial and significant class of crucial cases. [Chomsky, 1965, p. 24]

Supposing the tension with mentalism were relieved, allowing other data types to influence competence models, and introspective judgments were used only at the outset of linguistic inquiry, intuitions would still pose a number of serious methodological problems. It is not just the role of intuitions in linguistic theorizing that has to be put under scrutiny, but also the claim that intuitions offer a vantage point on tacit knowledge.
2.2.1 *Intuitions in linguistics*

If the system of linguistic rules in a speaker’s brain really is “deeply unconscious and largely unavailable to introspection” [Jackendoff, 2003, p. 652], one should see discrepancies between overt linguistic behavior, that reflects ‘unconscious’ competence rules, and the intuitions or beliefs that speakers have about these rules. This notion has been substantiated by Labov [Labov, 1996], who collected evidence on a wide variety of cases in regional American English. One example is the positive use of ‘anymore’ in various sections of the Philadelphia white community, meaning that a situation which was not true some time in the past is now true, roughly equivalent to ‘nowadays’:

(1) Do you know what’s a lousy show anymore? Johnny Carson.

Labov interviewed twelve speakers who used the adverb freely and consistently with its vernacular meaning exemplified in (1). He reported a majority of negative responses when they were asked whether a sentence like (1) is acceptable, and surprisingly weak intuitions on what the expression signifies in their own dialect, which contexts are appropriate for its use, and what inferences can be drawn from its occurrences.

Other arguments suggest that the use of intuitions in linguistics is problematic in many ways. For instance, [Marantz, 2005] has observed that grammaticality is a technical term defined within linguistic theories: a sound/meaning pair is grammatical or well-formed with respect to a grammar if and only if that grammar generates or assigns a structural description to the pair such that all relevant grammaticality or well-formedness constraints can be satisfied. In the quote from Aspects above, Chomsky takes for granted that structural descriptions assigned by some grammar to sentences can be checked for correspondence against native speakers’ judgments. However, native speakers of a language can hardly be said to have intuitions of grammaticality in the technical sense, nor can they grasp other properties of strings as these are defined within a formal grammar. Moreover, naïve language users might conflate into the notion of grammaticality different morphosyntactic, semantic and pragmatic criteria of acceptability, and they might do so in a way that is beyond control for the linguist. Similar observations would also apply to intuitive judgments of synonymy or truth-conditions, as opposed to formal definitions within a semantic theory.

As a way out, one might argue that a caveat only applies to naïve informants, and that the intuitions of linguists, immune to pre-theoretical notions of grammaticality, synonymy, and the like, are in fact reliable [Devitt, 2006]. Relevant to this issue, is an experiment by [Levelt, 1972] in which the intuitions of twenty-four trained linguists were investigated. Participants were presented with fourteen examples from their own field’s literature, among which:

(2) a. No American, who was wise, remained in the country.
   b. The giving of the lecture by the man who arrived yesterday assisted us.
None of the linguists rated correctly the ungrammatical sentence (2a), and sixteen judged the well-formed sentence (2b) as ungrammatical. Ungrammatical sentences had less chance of being judged ungrammatical than grammatical items. Levelt warns against taking these results too seriously, but he observes with some reason that “they are sufficiently disturbing to caution against present day uses of intuition” [Levelt, 1972, p. 25].

We could go on providing other examples of the problems that might arise with the use of introspective reports in the analysis of specific natural language sentences. However, we should now like to take a different approach, considering an argument targeted at the nature and scope of intuitions. The argument, introduced and discussed by Hintikka [Hintikka, 1999], starts with the observation that intuitions of grammaticality, synonymy etc. always relate to particular sentences (i.e. tokens), and not to entire classes of items, or to the common syntactic or semantic structure they share (i.e. types). Hintikka writes that

intuition, like sense perception, always deals with particular cases, however representative. […] But if so, intuition alone cannot yield the general truths: for instance, general theories for which a scientist and a philosopher is presumably searching. Some kind of generalizing process will be needed, be it inductive inference, abduction, or a lucky guess. The intuitions [Chomsky] recommended linguists to start from were intuitions concerning the grammaticality of particular strings of symbols, not concerning general rules of grammar. [Hintikka, 1999, p. 137-138]

Against Hintikka’s claim, one may argue that also paradigmatic variation is a proper object of intuition. The linguist would then be able to generalize over the properties of linguistic structures by constructing a paradigmatic set of sentences exhibiting those properties. This view however can be countered with the observation that the supposed ‘intuitions’ about paradigmatic cases are more similar to theory-laden hypotheses than to introspective judgments of naïve informants. The linguist, in order to construct such paradigmatic items, has to be able to control all irrelevant variables and systematically manipulate the factors of interest. This, in turn, requires that the linguist knows details of the grammar or the logical structure of the language which seem inaccessible to naïve speakers. It is this knowledge, which is often drawn from existing theories, that allows the linguist to have intuitions about linguistic structure. This leads us to Hintikka’s key statement, that competence theories are not equipped with built-in devices for deriving abstract grammatical or semantic forms from particular linguistic samples. That is,

reliance on generalization from particular cases is foreign to the methodological spirit of modern science, which originated by looking for dependencies of different factors in instructive particular cases (often in controlled experimental situations), and by studying these dependences by the same mathematical means as a mathematician uses in studying
the interdependencies of different ingredients of geometrical figures in analytic geometry. [...] transformational grammarians and other contemporary linguists would do a much better job if, instead of relying on our intuitions about isolated examples, they tried to vary systematically suitable ingredients in some sample sentence and observed how our `intuitions' change as a consequence. Now we can see why such systematic variation is a way of persuading our intuitions to yield general truths (dependence relations) rather than particular cases. [Hintikka, 1999, p. 135]

If intuitions are to serve as a reliable starting point in linguistic inquiry, they should be proved to have systematic properties. Observing patterns of covariation of introspective judgments and other factors – such as the structure, the context of occurrence of the sentence, the attitude of the speaker, and so on – would make the particular examples under consideration instructive and thus effective as part of the empirical basis of linguistic theories. The important consequence is that, in order to systematically alter the ingredients of sample sentences, the linguist should be able to control these factors in a manner similar to the manipulation of experimental variables in laboratory research. The solution offered by Hintikka to the problem of intuitions points in the direction of infusing linguistic practice with psychological experimentation. The linguist would as usual start from intuitions, but only the systematic aspects of these as revealed by experimentation, and if necessary statistical tests, would be preserved and transferred into the theory (see [Bunge, 1984, pp. 158-163] for a similar position).³ Hintikka offers an intriguing example, in which one tries to define the meaning of an expression in Montague grammar on the basis of systematic dependencies between subjects' intuitions and the contexts of occurrence of the expression of interest. In particular, he writes, if the notion of possible world is allowed in the theory,

then there is, in principle, no definite limit as to how far your experimentation (construction of ever new situations) can carry you in determining the class of scenarios in which the word does or does not apply. And such a determination will, at least for a Montagovian semanticist, determine the meaning of the word. Indeed, in Montague semantics, the meaning of a term is the function that maps possible worlds on references (extensions) of the appropriate logical type (category). And such functions can, in principle, be identified even more and more fully by systematic experimentation with the references that a person assigns to his terms in different actual or imagined scenarios.

³[Bunge, 1984, p. 168] pinpoints several methodological choices in generative linguistics which seem to diminish its relevance in empirical science, such as the “conduct of linguistic inquiry in total independence from neuroscience, social science, and even scientific psychology” and “a heavy reliance on intuition”. We too consider these as obstacles to understanding language, but we disagree with the judgment that Bunge formulates based on these remarks – that modern linguistics is (or has been) pseudo-scientific.
However, it may be a fair point in favor of introspective judgments in a broader sense to add that Hintikka considers thought experiments on a par with genuine experimentation [Hintikka, 1999, p. 146]. Thus, instead of eliciting overt responses from subjects in a number of conditions, the experimenter imagines herself in such situations. If the relevant variables are controlled with as much care as one would exercise in an experimental setting, introspection can reveal systematic aspects of language use, and thus contribute to theories of competence.

Hintikka’s argument can be made more explicit with reference to a number of studies investigating the role of the most important of his ‘suitable ingredients’ – context. Linguistic and psycholinguistic research has demonstrated that the context in which a sentence occurs can affect judgments of acceptability. [Bolinger, 1968] reported that sentences, which speakers judge as semantically implausible when presented in isolation, are regarded as acceptable when embedded in context. Consider the following examples:

(3) a. It wasn’t dark enough to see.
   b. I’m the soup.

These sentences are typically judged as semantically deviant, although for different reasons: (3a) because one normally needs light in order to see, and (3b) because the predicate ‘being a soup’ cannot be applied to a human being. Now, consider the same sentences embedded in a suitable discourse context, with (4b) being spoken at a cashier’s counter in a restaurant:

(4) a. I couldn’t tell whether Venus was above the horizon. It wasn’t dark enough to see.
   b. You’ve got us confused. You’re charging me for the noon special. The man in front of me was the noon special. I’m the soup.

Examples (3) in an appropriate context seem perfectly acceptable. Because context has such marked effects on intuitions, linguistic theory, if it has to rely on introspective judgments, should explicitly take into account this fact.

2.2.2 Intuitions in psycholinguistics

The appeal to intuitions was not an explicit choice of methodology in psycholinguistics and the cognitive neuroscience of language. In fact, the method of introspection was discarded in scientific psychology after its failures in the 19th century. However, it is adopted in language processing research as a means of establishing differences between sentence types to be used as stimuli in actual experiments. 

4Although we consider Hintikka’s an informative example of linguistic theorizing based on covariation patterns of contextual factors and intuitions, we must also add that there are serious problems with the notion of meaning (that is, Frege’s Sinn) in Montague semantics. For instance, since the theory allows for infinitely many possible worlds, it becomes unclear whether we can even approximate the meaning of an expression using Hintikka’s method.
The typical procedure for setting up a language processing study is to start with a few sample sentences differing with respect to some linguistic feature, the assessment of which is initially left to the intuitions of the experimenter. For instance, let us consider one of the first ERP studies on syntactic constraint violations, by [Osterhout and Holcomb, 1992]. Here the starting point is a pair – or a relatively small set of pairs – of sentences containing either an intransitive (5a) or a transitive (5b) verb, using a direct object construction:

\[(5)\]

a. The woman struggled to prepare the meal.
b. The woman persuaded to answer the door.

Up to this stage, the methodology is by and large the same as that of the linguist. However, while the latter would then proceed with, say, formalizing the requirements of intransitive and transitive verbs with respect to direct objects, the psycholinguist, to make sure there is sufficient statistical power to test her processing hypotheses in a dedicated experiment, would have to construct a large set of sentences with the same structure and properties of (5a-b). In the next step, the sentences would be presented to subjects while the dependent variables of interest are measured, which in the study of Osterhout and Holcomb were ERPs and grammaticality ratings. Grammatical sentences like (5a) were judged to be acceptable in 95% of the cases, and supposedly ungrammatical items like (5b) in 9% of the cases. One may argue, as an academic exercise towards an explanation of the 9% figure, that (5b) does have contexts in which it is both grammatical and semantically acceptable, for instance if it is interpreted as a reduced relative clause (‘The woman who was persuaded to answer the door’), and is uttered as an answer to a who question, as in the following dialogue:

\[(6)\]

A: Who stumbled on the carpet in the hallway?
B: The woman persuaded to answer the door.

We have already encountered this phenomenon discussing Bolinger’s examples above. In a context such as (6), Osterhout and Holcomb’s ungrammatical sentence becomes perfectly admissible. Acceptability judgments, therefore, depend on the range of uses (or contexts of use) readers are willing to consider. In this sense, subjects’ intuitions may differ from those of the experimenter. For example, a linguist would remind us that ‘The woman persuaded to answer the door’ is an NP, and not a sentence. But what prevents naïve language users from including well-formed NPs into their notion of ‘sentence’? Here the answer can only be: the linguist’s own notion of ‘sentence’. This also suggests that discrepancies between the intuitions of naïve informants and trained scientists may be more important than isolated linguists’ intuitions when it comes to fully explaining a data set.

2.3 Beyond competence and performance

Intuitions are but one of the many sources of concern for a thoroughly mentalistic approach to language. As [Jackendoff, 2002, p. 29] has pointed out, there is
Jackendoff suggests we adopt a ‘soft’ competence/performance distinction, adding a third component to the framework [Jackendoff, 2002]. The theory of competence is seen as the characterization of phonologic, syntactic and semantic data structures as they are processed and stored in the brain of speakers during language acquisition. The theory of performance is the description of the actual occurrence of such data structures in language comprehension and production. The theory of neural instantiation is an account in terms of brain structures and processes of competence and performance. Jackendoff provides an architecture in which competence components (phonology, syntax and semantics, plus interface rules) interact in a manner that is consistent with the incrementality and the ‘opportunism’ (his label for immediacy) of language processing [Jackendoff, 2007]. However, to solve the dilemma described above, it is not enough to show that competence determines the state-space available to users of a language during performance [Jackendoff, 2002, p. 56]. The issue is, rather, whether there is interplay between competence and performance, that is – turning Jackendoff’s tag line upside down – whether the logic of processing dictates the logic of competence, and to what extent.

As we saw above, in his early writings Chomsky claimed that theories of competence have nothing to learn from processing data [Chomsky, 1965]. Minimalists have suggested that syntax is adapted to the requirements holding at the interface with other cognitive modules, such as the sensory-motor and conceptual systems. However, they deny what functionalists on the contrary accept, namely
that syntax is well-designed for use [Chomsky et al., 2002; Hauser et al., 2002; Fitch et al., 2005]. Evidence against adaptation to performance is provided, according to minimalists, by memory limitations, constructions such as garden-path and center embedding sentences, which seem suboptimal in various respects. Here two remarks are in order. The first is that such phenomena do not constitute evidence against adaptation per se, but rather (if anything like that is possible) against ‘perfect’ adaptation. Minimalists seem to commit what optimality theorists have called the ‘fallacy of perfection’ [McCarthy and Prince, 1994], consisting in confusing optimal outcomes, which are the result of equilibria between different variables, with best possible outcomes for just a subset of the factors involved, for instance the absence of unstable or ambiguous constructions (see [Pinker and Jackendoff, 2005] for discussion). The second remark is that, even if we assume that competence is neither perfectly nor optimally adapted to use, it still seems conceivable that performance constraints shaped competence rules. Therefore, the problem is not whether language is an adaptation: that some traits of competence reflect the outcomes of adaptive evolutionary processes acting on actual brain systems, including adaptation to communication needs, seems to be a widely accepted view [Hauser et al., 2002; Pinker and Jackendoff, 2005; Fitch et al., 2005]. The problem is rather: (how) can we construct a methodological framework in which it is possible to determine what aspects of competence can be explained adaptively?

The reason why generative linguistics does not seem capable of addressing this issue is, in our opinion, to be attributed more to how performance is defined than to a rigid view of the competence/performance distinction. [Jackendoff, 2002, p. 30] rightly observes that in Chomsky’s original proposal a large and heterogeneous set of phenomena were collapsed into ‘performance’: errors, shifts of attention, memory limitations, processing mechanisms, and so on. Only a very superficial assessment of the factors involved in language use could justify the notion that a single, relatively compact theory of performance could account for all those phenomena. It seems more reasonable to assume that different theories, developed using different analytical approaches, are necessary to understand how language interacts with memory and attention, how errors of different type and origin are produced (for also language disorders give rise to performance failures), and so on. We agree with Jackendoff on the characterization of competence and neural implementation, but we believe a more appropriate intermediate level should be chosen.

2.4 Marr’s three-level scheme as applied to language

Jackendoff’s updated view of the competence/performance distinction as a soft methodological separation, plus a theory of neural realization, resembles Marr’s 1982 tripartite scheme for the analysis of cognitive systems [Spivey and Gonzalez-Marquez, 2003]. Marr suggested that cognitive processes should be modeled at three, nearly independent levels of analysis: a computational level (what is com-
puted?), an algorithmic level (how is computation carried out?), and a level of physical implementation (what are the properties of the real machines that can execute the algorithms?). From this perspective, Jackendoff’s trajectory away from Chomsky appears incomplete. There is a partial redefinition of competence/performance, and the addition of a third level, but it is doubtful whether this move leads to the kind of transitions and mutual constraints between levels of analysis afforded by Marr’s scheme. So it may be worth asking what would be the advantages of taking one step further, that is replacing the competence/performance distinction with Marr’s distinction between computational and algorithmic analyses.

An important consequence of this choice is that performance theory is now seen as an intermediate level of analysis at which the algorithms and memory mechanisms supporting specific linguistic computations are described. That might seem a rather abrupt move, as it restricts the scope of performance to algorithms, and thereby leaves aside a large number of phenomena which, some might suggest, cannot be adequately treated in algorithmic terms. For instance, conscious inner speech is an important function of language [Carruthers, 1996; Carruthers, 2002], and one in which there seems to be no definite input-output mapping involved. On the other hand, for those phenomena that are best treated as structured input-output processes, for example language production and comprehension, Marr’s framework allows competence theories, if properly construed, to be investigated as part of actual information processing systems. Applications of this idea to semantics will be shown below.

Does our appeal to Marr’s three-level scheme solve the problems associated with the competence/performance distinction? It seems it does, because the variety of phenomena that were collapsed into performance can now be understood in their distinctive features. For instance, working memory as involved in a given task can be examined at the level of algorithms. The algorithmic analysis may suggest a description of the memory architecture and the memory resources required by the system, and this constitutes a first step toward an explanation in neural terms. Conversely, memory organization constrains the classes of algorithms that can be computed by that machine, and the type of data structures that the computational theory can produce. An example of bottom-up adjustment is Yngve’s 1960 explanation of the dominance of right-branching over left-branching and center-embedding structures. Another example are ‘minimal models’ of discourse, as we shall see later.

In brief, one key feature of Marr’s scheme is that the way levels of analysis are defined makes it easier to see how formal theories of grammar, language processing and neural computation could be integrated and mutually constrained. It seems that, preserving the notion of ‘performance’, and a fortiori ‘competence’, such well-constrained formal routes between levels of analysis would become less accessible.

Below we apply this tentative methodological sketch to the semantics of tense, aspect and event structure. Our goal is to devise semantic analyses that are formally specified and cognitively motivated, that is, highlighting connections be-
between the meanings of temporal expressions, planning and reasoning. The semantic analyses should also be algorithmically explicit, such that processing predictions, or general constraints on processing architecture, can be formulated. We hope to show that our theory of tense, aspect, and event structure not only meets these requirements, but can also be used to provide alternative explanations of existing experimental data on language comprehension. The last part of this chapter puts our enterprise into a wider neuroscience-oriented perspective, introducing the ‘binding problem for semantics’.

3 PLANNING, REASONING, MEANING

3.1 The cognitive substrates of tense and aspect

We see it as the essential purpose of tense and aspect to facilitate the computation of event structure as described in a narrative. One consequence of this characterization is that, contrary to what generative and formal semantic approaches maintain, it is not very useful to study tense and aspect at the sentence level only. Tense, aspect and event structure really come into their own only at the discourse level [Comrie, 1976; Comrie, 1985]. Tense and aspect, however, cannot by themselves determine event structure, and must recruit world knowledge. Examples (7a-c) will make clear what we have in mind.

French has several past tenses (Passé Simple, Imparfait, Passé Composé), which differ in their aspectual contributions. The following mini-discourses in French all consist of one sentence in the Imparfait and one in the Passé Simple. However, the structure of the set of events differs in each case.

(7) a. Il faisait chaud. Jean ôta sa veste. (Imp, PS)  
*It was hot. Jean took off his sweater.*

b. Jean attrapa une contravention. Il roulait trop vite. (PS, Imp)  
*Jean got a ticket. He was driving too fast.*

c. Jean appuya sur l’interrupteur. La lumière l’éblouissait. (PS, Imp)  
*Jean pushed the button. The light blinded him.*

In the first example, the Imp-sentence describes the background against which the event described by the PS-sentence occurs. In the second example, the PS-event terminates the Imp-event, whereas in the third one the relation is rather one of initiation. These discourses indicate that world knowledge in the form of knowledge of causal relations is an essential ingredient in determining event structure. This knowledge is mostly applied automatically, but may also be consciously recruited if the automatic processing leaves the event structure undetermined. It is the task of cognitive science to determine what this algorithm looks like, and how

---

5 Taken from an unpublished manuscript on French tenses by [Kamp and Rohrer, 1985]. See also [Eberle and Kasper, 1989] and [Asher and Lascarides, 2003].
it is actually implemented. The prominent role of causal relationships in (7a-c) suggests a direction in which to look for that algorithm.\footnote{There is a body of literature on what are called 'situation models' (what we have called 'discourse models' or 'event structures') which contains suggestive evidence to show that these models not only represent objects and events introduced by the discourse, but also general and specific causal information about the world not explicitly mentioned in the discourse. Space constraints forbid extensive discussion of this line of research; we can only direct the reader to the survey article [Zwaan and Radvansky, 1998].}

3.2 Planning, causality and the ordering of events

Stated bluntly, our hypothesis is:

The ability to automatically derive the discourse model determined by a narrative is subserved by the ability to compute plans to achieve a goal.

At first this hypothesis may seem unintelligible: what do goals and plans have to do with discourse? But as we will see, it is possible, indeed advantageous, to represent tense and aspect formally as goals to be satisfied. A discourse then sets up a system of interlocking goals, which is at least formally similar to the complex goals that occur in, say, motor planning. The hypothesis then says that the formal similarity arises from the fact that the very same cognitive mechanism is responsible for dealing with goals in both domains, motor control and language processing.

Here we present several converging lines of evidence which lend some plausibility to this conjecture. Firstly, at a very general level one can say that a distinguishing feature of human cognition is that it is goal-oriented, with goals that range from very short-term (get a glass of water) to very long-term (having sufficient income after retirement). In each case, the goal is accompanied by a plan which produces an ordered collection of actions, be they motor actions or transfers of money to a special account. More precisely, planning consists in

the construction of a sequence\footnote{More complex plans are possible, involving overlapping actions.} of actions which will achieve a given goal, taking into account properties of the world and the agent, and also events that might occur in the world.

Given the ubiquity of goal-plan organisation in human cognition, it is not surprising that there have been numerous attempts to link the language capacity with the planning capacity. The setting is usually a discussion of the origins of language. Even if it is granted that some non-human primates have learned a primitive form of language, there is still a striking difference in language proficiency between chimpanzees and ourselves. It is still a matter of ongoing debate to determine exactly what this difference consists in. Some would say that the difference is in the syntax: human syntax is recursive, the chimpanzee’s syntax (if that is the word) is not. One may then point to an analogy between language and
planning. Language production can be characterized as transforming a semantic structure, to which the notion of linearity may not be applicable, into linear form, that is an utterance. Planning also involves linearization, and that is how the language-planning connection is drawn. An alternative strategy, not inconsistent with the first, is to show that the recursive structure of syntax is linked to the recursive structure (or hierarchical organization) of plans [Greenfield, 1991; Steedman, 2002]. Non-human primates engage in planning for time spans not exceeding 48 hours, as is known since Kohler’s 1925 observations. This has also been attested in squirrel monkeys in experiments by [McGonigle et al., 2003]. What these experiments jointly show is that on the one hand planning provides a link between humans and non-human primates, but that on the other hand complex planning sets humans apart from non-human primates. As such the planning capacity can be a starting point for discussions of the origins of language because it may account for both continuity (capacity for planning shared between humans and their ancestors) and change (increased capacity for planning leading to human linguistic abilities).

A more direct connection between language and planning, and one focussing on semantics rather than syntax, was investigated experimentally by [Trabasso and Stein, 1994] in a paper whose title sums up their program: *Using goal-plan knowledge to merge the past with the present and the future in narrating events on-line*. Trabasso and Stein argue that “the plan unites the past (a desired state) with the present (an attempt) and the future (the attainment of that state)” [Trabasso and Stein, 1994, p. 322], “[c]ausality and planning provide the medium through which the past is glued to the present and future” [Trabasso and Stein, 1994, p. 347]. They present the results of a study in which children and adults were asked to narrate a sequence of 24 scenes in a picture storybook called *Frog, where are you?*, in which a boy tries to find his pet frog which has escaped from its jar. The drawings depict various failed attempts, until the boy finds his frog by accident. The aim of the study is to determine what linguistic devices, in particular temporal expressions, children use to narrate the story as a function of age. The authors provide some protocols which show a child of age 3 narrating the story in a tenseless fashion, describing a sequence of objects and actions without relating them to other objects and actions. None of the encoded actions is relevant to the boy’s ultimate goal. Temporal sequencing comes at age 4, and now some of the encoded actions are relevant to the goal. Explicit awareness that a particular action is instrumental towards the goal shows up at age 5. At age 9, action-goal relationships are marked increasingly, and (normal) adults structure the narrative completely as a series of failed or successful attempts to reach the goal. Thus we see that part of what is involved in language learning is acquiring the ability to produce discourse in such a way that a goal-plan structure is induced in the hearer.

---

8This is a classic experimental paradigm for investigating the acquisition of temporal notions in children. See Berman and Slobin [Berman and Slobin, 1994] for methods, results and, last but not least, the frog pictures themselves. We will come back to this paradigm when discussing use of verb tense in children with ADHD in section 3.2.3.
The authors’ claim is that such discourse models are never mere enumerations of events, but that our very mental representation of time privileges discourse models in which events can be viewed as part of a plan.

Our proposal is that also when viewed computationally, discourse models are best treated as plans, i.e., as output of the planning mechanism. Indeed, it is of some interest to observe that the ingredients that jointly enable planning have a prominent role to play in the construction of a discourse model. Take for instance causality, shown to be involved in the interpretation of (7a-c). Planning essentially requires knowledge of the causal effects of actions as well as of the causal effects of possible events in the world. Accordingly, the planning capacity must have devised ways of retrieving such knowledge from memory. Planning also essentially involves ordering actions with respect to each other and to events occurring in the world which are not dependent upon the agent. Furthermore, the resulting structure must be held in memory at least until the desired goal is attained. The reader can easily envisage this by considering the planning steps that lead to a pile of pancakes. For instance, causal knowledge dictates that one has to pour oil in the frying-pan before putting in the batter, and this knowledge has to remain active as long as one is not finished.

While the preceding considerations point to some data structures common to both plans and discourse models, the fundamental logical connection between discourse processing and planning is that both are non-monotonic. When we plan, deliberately or automatically, we do so in virtue of our best guess about the world in which we have to execute our plan. We may plan for what to do if we miss the bus, but we don’t plan for what to do if the bus doesn’t come because the gravitational constant changes, even though that is a logical possibility. Similarly, the computation of a discourse structure may be non-monotonic. For instance, the reader who sees (8a) is likely to infer (that is, to read off from the discourse model) that Bill is no longer a member, but that implicature can easily be canceled, as in (8b):

(8)  a. Bill used to be a member of a subversive organization.
    b. Bill used to be a member of a subversive organization, and he still is.

The discourse model belonging to (8b) is not simply an extension of the one for (8a), although (8b) is an extension of (8a); but the temporal interpretation of the main clause must be recomputed in going from (8a) to (8b). We will see more examples of this phenomenon when investigating the relation between verb tenses and planning.

We propose that the link between planning and temporal semantics is provided by the notion of goal. In both comprehension and production, the goal interpreting the tensed VP is to introduce the event corresponding to the tensed VP into the already existing event structure. This goal always has two components:

1. location of event in time;
2. meshing it with other events.
The role of planning is to establish definite temporal relationships between the events involved.

How planning can do this is best illustrated by means of an example. Consider what goes on in comprehending

(9) Max fell. John pushed him.

On one prominent reading, the event described in the second sentence precedes, indeed causes, that described in the first sentence. The relevant goals are in this case:

update discourse with past event $e_1 = \text{fall}(m)$ and fit $e_1$ in context
update discourse with past event $e_2 = \text{push}(j,m)$ and fit $e_2$ in context

Planning must determine the order of $e_1$ and $e_2$, and to do so the planning system recruits causal knowledge as well as the principle that causes precede effects. We sketch an informal argument here; the next subsection gives more formal details, necessary to understand the connection between discourse processing viewed as non-monotonic computation, and traces of this computation in neural responses. The informal argument runs like this. There is (assumed to be) no context yet in which $e_1$ must be processed, so $e_1$ is simply located in the past. When it comes to interpreting $e_2$, we have a context ($e_1$). The planning mechanism now retrieves possible relationships involving both $e_1$ and $e_2$, and one of these is that a push initiates falling. Since the cause comes before its effect, this yields that $e_2$ precedes $e_1$. Observe that this is a default inference only; as we will see, it is possible to unify $e_1$ and $e_2$ with other material such that their temporal order becomes reversed.

3.2.1 Computations on event structures

To give the reader a detailed picture of what goes on in such computations, we have to introduce some notation, borrowed from the Event Calculus [van Lambalgen and Hamm, 2004], which will also be useful for our discussion of the ‘binding problem’ later in this chapter. We make a distinction between events (denoted $e, e', ..., e_0, ...$) and processes or fluents (denoted $f, f', ..., f_0, ...$). We say that events occur or happen at a particular time, and represent this by the expression $\text{Happens}(e, t)$. By contrast, processes do not occur, but are going on at a particular time, and for this we use the predicate $\text{HoldsAt}(f, t)$. Events and processes can stand in causal relations. For instance, an event may kick off a process: $\text{Initiates}(e, f, t)$; or it may end one: $\text{Terminates}(e, f, t)$. We will use these predicates to mean the causal relation only. It is not implied that $e$ actually occurs. Finally, a useful predicate is $\text{Clipped}(s, f, t)$, which says that between times $s$ and $t$ an event occurs which terminates the process $f$. The predicates just introduced are related by axioms, of which we shall see a glimpse below. With this notation, and using $\varphi(x)$ succeeds

---

9This section can be skipped by readers who have never seen the Event Calculus before.
to abbreviate: ‘make it the case in our discourse model that \( \varphi(x) \)'\(^{10}\), we can write the two update instructions involved in comprehending the discourse as:

\[ (10) \ f_succeeds^{\text{Happens}}(e_1, t), t < \text{now}, \ f_succeeds^{\text{Happens}}(e', t') \]

\[ (11) \ f_succeeds^{\text{Happens}}(e_2, s), s < \text{now}, \ f_succeeds^{\text{Happens}}(e'', t'') \]

Here \( e' \) and \( e'' \) are variables for event types in the discourse context which have to be found by substitution or, more precisely, unification. These two update instructions have to be executed so that \( e'' = e_1 \) and \( s < t'' \). If ‘Max fell’ is the first sentence of the discourse, we may disregard \( e' \).\(^{11}\) In order to formulate the causal knowledge relevant to the execution of these instructions, we introduce a process \( f \) (falling) corresponding to the event \( e_1 = \text{fall}(m) \), where \( f, e_1 \) and \( e_2 \) are related by the following statements:

\[ (12) \ \text{HoldsAt}(f, t) \rightarrow \text{Happens}(e_1, t) \]

\[ (13) \ \text{Initiates}(e_2, f, s) \]

The system processing the discourse will first satisfy the update request corresponding to ‘Max fell’ by locating the event \( e_1 \) in the past of the moment of speech. The second sentence, ‘John pushed him’, is represented by the request (11) which contains the variable \( e'' \). The system will try to satisfy the goal by reducing it using relevant causal knowledge. Applying (12) and unifying\(^{12}\) \( e'' = e_1 = \text{fall}(m) \), the second request (11) is reduced to:

\[ (14) \ f_succeeds^{\text{Happens}}(e_2, s), s < \text{now}, \ f_succeeds^{\text{Happens}}(e_1, t''), \ \text{HoldsAt}(f, t'') \]

The system now applies a general causal principle, known as inertia, which says that, if an event \( e \) has kicked off a process \( f \) at time \( t \), and nothing happens to terminate the process between \( t \) and \( t' \), then \( f \) is still going on at \( t' \). This principle rules out spontaneous changes, that is changes which are not caused by occurrences of events. Inertia can be formulated as the following axiom:

\[ (15) \ \text{Happens}(e, t) \land \text{Initiates}(e, f, t) \land t < t' \land \neg \text{Clipped}(t, f, t') \rightarrow \text{HoldsAt}(f, t') \]

Using this axiom, the request (14) is further reduced to:

\[ (16) \ f_succeeds^{\text{Happens}}(e_2, s), s < \text{now}, \ f_succeeds^{\text{Happens}}(e_1, t''), \ \text{Happens}(e, t), \ \text{Initiates}(e, f, t), t < t'', \ 
\neg \text{Clipped}(t, f, t'') \]

\(^{10}\)This notation derives from logic programming. By itself, \( ?\varphi(x) \) denotes a goal or query, a request for a value \( a \) of \( x \) such that \( \varphi(a) \) is true. The answer may be negative, if the database against which \( \varphi(x) \) is checked contains no such individual. By \( ?\varphi(x) \) succeeds we mean that in such cases the database must be updated with an \( a \) making \( \varphi \) true. These instructions or requests for updates are also known as integrity constraints.

\(^{11}\)Here we regard context as provided by the preceding discourse, but one may conceive of ‘forward-looking’ notions of context as well.

\(^{12}\)This form of unification will be important in our discussion of the ‘binding problem’ for language.
Using (13) and unifying $e = e_2 = \text{push}(j, m)$ and $s = t$ we reduce this request to:

\[
?\text{Happens}(e_2, s), s < \text{now}, \text{Happens}(e_1, t''), s < t'', \neg \text{Clipped}(s, f, t'') \text{ succeeds}
\]

This is a definite update request which almost says that \textit{push} precedes \textit{fall}, except for the formula $\neg \text{Clipped}(s, f, t'')$, which expresses that $f$ has not been terminated between $s$ and $t''$. If $f$ were terminated between $s$ and $t''$, we would have a situation as in:

\[
\text{(18) Max fell. John pushed him a second time and Max fell all the way to the bottom of the pit.}
\]

Since we have no positive information to this effect, we may assume $\neg \text{Clipped}(s, f, t'')$. This form of argument is also known as \textit{closed world reasoning}: ‘assume all those propositions to be false which you have no reason to assume to be true’. Closed world reasoning is essential to planning, and to discourse comprehension, as it allows one to discount events which are logically possible but in practice irrelevant. The final update request is thus:

\[
?\text{Happens}(e_2, s), s < \text{now}, \text{Happens}(e_1, t''), s < t'' \text{ succeeds}
\]

which is the instruction to update the discourse model with the past events $e_1$ and $e_2$ such that $e_2$ precedes $e_1$.

Just as plans may have to be revised in mid-execution (for instance, if it turns out there is not sufficient oil to produce the projected number of pancakes), discourse models may have to be recomputed when additional information is provided. Suppose the discourse does not stop after ‘John pushed him’ but, instead, continues:

\[
\text{(20) Max fell. John pushed him, or rather what was left of him, over the edge.}
\]

One obvious interpretation is that now $e_2 = \text{push}(j, m)$ comes after $e_1 = \text{fall}(m)$. This is the result of a recomputation, since after the first ‘him’ the hearer may have inferred that $e_2$ precedes $e_1$. Let us give a brief, informal sketch of this recomputation. The phrase ‘or rather what was left of him’ suggests Max is now \textit{dead}, therefore the update request corresponding to the second sentence is something like:

\[
?\text{Happens}(e_2, s), s < \text{now}, \text{HoldsAt}(\text{dead}(m), s), \text{Happens}(e'', t'') \text{ succeeds}
\]

perhaps together with a requirement to the effect that the entire pushing event occurs while \text{dead}(m) obtains. It now seems reasonable to assume that, at the start of \textit{falling} (the process denoted by $f$), Max is still \textit{alive}. Unifying $e'' = e_1$ and applying property (12), the request reduces to finding instants $s, t''$ such that:

\[
?\text{Happens}(e_2, s), s < \text{now}, \text{HoldsAt}(\text{dead}(m), s), \text{HoldsAt}(\text{alive}(m), t''), \\
\text{Happens}(e_1, t'') \text{ succeeds}
\]
can be satisfied. Since *alive* always precedes *dead* and not conversely, it follows that we must have that $e_1 = \text{fall}$ precedes $e_2 = \text{push}$.

In summary, what we have outlined here is a computational mechanism for determining event structure from discourse, based on planning. Temporal expressions are hypothesized to determine requests to be satisfied by an update of the current discourse model. Processing these requests involves unification, search through semantic memory, as well as setting up temporary structures in working memory.

3.2.2 *Computing event structures for (PS, Imp) combinations*

Similar arguments apply to the French examples with which we started this section:

(7) a. Il faisait chaud. Jean ôta sa veste. (Imp, PS)

*It was hot. Jean took off his sweater.*

Intuitively, this narrative determines an event structure in which *hot* acts as a background which is true all the time, and the foreground event (taking off one’s sweater) is located within this background. One arrives at this structure by means of the following argument. World knowledge contains no causal link to the effect that taking off one’s sweater changes the temperature. The goal corresponding to the first sentence dictates that it is hot at some time $t$ before *now*. By the principle of inertia, the state *hot* must either hold initially (at the beginning of the narrative) or have been initiated. The latter requires the occurrence of an initiating event, which is however not given by the discourse. Therefore, *hot* holds initially. Similarly, no terminating event is mentioned, so *hot* extends indefinitely, and it follows that the event described by the second sentence must be positioned inside *hot*.

The second example dates back to the bygone days when speeding cars were stopped by the police instead of being photographed:

(7) b. Jean attrapa une contravention. Il roulait trop vite. (PS, Imp)

*Jean got a ticket. He was driving too fast.*

It is given that the event of getting a ticket occurred sometime in the past, and it is also given that the fluent *speeding* was true sometime in the past. Hence, it holds initially or has been initiated. We have to determine the relative position of event and fluent. World knowledge yields that getting a *ticket* terminates, but does not initiate *speeding*. Because this is the only event mentioned, *speeding* holds from the beginning of discourse, and is not reinitiated once it has been terminated.

In the third example, the same order of the tenses yields a different event order, guided by the application of causal knowledge:

(7) c. Jean appuya sur l’interrupteur. La lumière l’éblouissait. (PS, Imp)

*Jean pushed the button. The light blinded him.*
One (occurrence of an) action is mentioned, pushing the light button, which has the causal effect of initiating the light being on when its current state is off. No terminating event is mentioned, therefore the light remains on. It also follows that the light must be off for some time prior to being switched on, and that it must be off at the beginning of discourse. The definite article in ‘La lumière’ leads to a search for an antecedently introduced light, which successfully terminates after unification with the light introduced in the first sentence. As a consequence, it is this light which is too bright.

3.2.3 Deviant verb tenses and ADHD

In cognitive terms, planning is part of ‘executive function’, an umbrella term for processes responsible for higher-level action control which are necessary for maintaining a goal and achieving it in possibly adverse circumstances. Executive function comprises maintaining a goal, planning, inhibition, coordination and control of action sequences. Since we have postulated that tense processing involves planning toward goal, we see that several components of executive function are involved in comprehension and production of tense and aspect. A corollary is that failures of executive function can show up in deviant use of tense and aspect and in impairments in processing temporal discourse, for instance in ASD (Autistic Spectrum Disorder), ADHD (Attention Deficit Hyperactivity Disorder), and schizophrenia. Of particular interest here will be children with ADHD, a disorder that is characterised by persistent and developmentally inappropriate levels of inattention, impulsivity and hyperactivity. About 2% of children (mainly boys) are severely affected; 3–6% suffer from less severe symptoms. It has been hypothesized to be an executive function disorder, and indeed children with ADHD score significantly lower on a number of standard tests measuring components of executive function, such as planning, inhibition, and self-monitoring. The precise pattern of executive deficits in ADHD is not yet known, and it is not yet determined whether there is a single executive deficit that explains most of the symptoms. Below we will investigate linguistic consequences of the hypothesis that goal maintenance is affected in ADHD, evidence for which can be found in [Shue and Douglas, 1992; Pennington and Ozonoff, 1996]. For instance, it is known that children with ADHD have trouble with retelling a story, a task that involves presenting information so that it is organized, (temporally) coherent, and adjusted to the needs of the listener. The ability to attend to these requirements presupposes that one is able to retain goals in working memory while planning the necessary steps and monitoring their execution. This ability requires executive function as defined above [Purvis and Tannock, 1997], and is known to be compromised in ADHD. On difficulties with in maintaining goals in working memory, see [Geurts, 2003].

Given that goal maintenance in working memory is compromised in children with ADHD, together with the proposal that such maintenance is necessary to allow computation of event structures (i.e. tense processing) we are led to the

\[13\] These are figures for the Netherlands, supplied by the Gezondheidsraad.
following suggestion [van Lambalgen et al., 2008].

Recall that update requests, that is the goals to be satisfied, corresponding to a VP’s tense and aspect, consist of two components:

1. location of an event in time;
2. meshing the event with other events.

If a child has trouble maintaining a goal in working memory, this may lead to a simplified representation of that goal. In the case of verb tenses, the most probable simplification is of ‘location of event in time’ (never mind the meshing with other events), since this involves the least processing (search through semantic memory and unification). This simplification affects both comprehension and production, the case of interest here. Indeed, in producing speech which is attuned to the needs of the listener, the speaker may construct a discourse model of his own utterances, to determine whether it is sufficiently unambiguous. Slightly more formally, our main hypothesis is:

A speaker with ADHD simplifies the goals corresponding to tenses at the expense of the hearer.

We list here a number of ways in which these goals can be simplified. An extreme form of complexity reduction is not to use tensed forms at all. For example, in frog-story experiment on ADHD narration, we saw discourses like this

En hij is vroeg op. En wat ziet die daar? Kikker verdwenen.
And he is up early. And what does he see there? Frog gone. [7yrs, ADHD]

The difference between control and ADHD children was quite striking: only 2.9% of controls used tenseless utterances in their narratives, whereas 19.2% of the ADHD children did so.

A second way in which the child with ADHD can ease the burden on himself while increasing that of the hearer, is using reported speech (‘quotes’, ‘direct speech’ only). Here’s an example of the contrast between the two groups: two ways of narrating the same scene, that in which the boy climbs into a tree and looks into a hole, whereupon an owl appears in the hole, and knocks the boy out.

a. En die jongen ging zoeken in de boom. En toen zag die een uil. En toen valt ’ie van de boom.
And that boy started looking in the tree. And then he saw an owl. And then he falls from the tree. [8yrs, CG] 14

b. ‘Oh nee, ik val!’ ‘Hellup!’ ‘Ga weg, stomme uil, ga weg!’
‘Oh no, I’m falling!’ ‘Help!’ ‘Go away, stupid owl, go away!’ [9yrs, ADHD]

14 The child makes a mistake in combining present tense ‘valt’, which could be interpreted as a narrative present heightening the tension, with the adverbial ‘en toen’, which needs a past tense.
There are several other ways the child with ADHD can reduce the complexity of its goals, e.g., reducing the number of context setting elements, or avoiding the perfects (which are computationally intensive). We may now take a more global view, and look at the children who apply one or more complexity-reducing strategies. For example, a child may use up all his computational resources by avoiding direct speech, thereby producing, say, more erratic shifts in the perfect. Both in case of excessive use of direct speech and of erratic tense shifts the hearer must work hard to construct a coherent story, even though he may not understand why he has to work so hard. Thus, taking the point of view of the hearer, what is necessary is a general definition of complexity-reducing strategy, incorporating the more specific strategies discussed above. Motivated by the analyses given above, we define the overall complexity-reducing strategy of a child as consisting of three components: tenseless utterances, direct speech, avoidance of the perfect. For the precise definition of ‘strategy’ we refer the reader to [van Lambalgen et al., 2008], here we state only the main result, that children with ADHD use a strategy with the aim of reducing tense complexity significantly more often. This may explain the sense of unease a hearer fails when listening to such narrative.

Before we close this section, one last word on methodology. The predictions concerning the use (or non-use) of verb tenses in ADHD were derived from a formal model [van Lambalgen and Hamm, 2004] of tense production and comprehension involving satisfaction of complex goals, together with neuropsychological evidence indicating difficulties with goal maintenance and/or planning toward that goal. The formal model is responsible for the specificity of the predictions. Without the formal model, but equipped only with, say, Trabasso and Stein’s general characterisation of narrative as governed by a hierarchy of goals [Trabasso and Stein, 1994], one expects some breakdown in the coherence of story-telling in ADHD, as was indeed found by Purvis and Tannock [Purvis and Tannock, 1997]. The formal model allows one to be more specific about the computational cost of the devices used to ensure discourse coherence. The model thus acts as a searchlight that allows one to see phenomena one would not have thought of otherwise.

4 THE BINDING PROBLEM FOR SEMANTICS

The goal of a theory of language is to deliver analyses at each of Marr’s levels, and to bridge them in a perspicuous manner. One way of achieving this is to define a notion that acts as a ‘wormhole’ [Hurford, 2003] connecting linguistic structures, algorithms, and neurobiological events. A candidate notion is that of ‘unification’, which has been applied on several occasions in this chapter. Below we provide a broad, neuroscience-oriented framework for the concept of unification.

An influential statement of the ‘binding problem’ for cognitive representations is due to [von der Malsburg, 1981], who regarded the binding approach to brain function as a response to the difficulties encountered by classical connectionist networks. Von der Malsburg 1999 refers to a well-known example by [Rosenblatt, 1962] to illustrate the issue. Consider a network for visual recognition constituted
by four output neurons. Two neurons fire when a specific shape (either a triangle or a square) is presented and the other two fire depending on the shape’s position (top or bottom of a rectangular frame). So, if there is a square at the top, the output will be [square, top]. If there is a triangle at the bottom, the output will read [triangle, bottom]. However, if a triangle and a square are presented simultaneously, say, the triangle at the top and the square at the bottom, the output would be [triangle, square, top, bottom], which is also obtained when the triangle is at the bottom and the square at the top. This is an instance of the ‘binding problem’. Malsburg writes:

The neural data structure does not provide for a means of binding the proposition top to the proposition triangle, or bottom to square, if that is the correct description. In a typographical system, this could easily be done by rearranging symbols and adding brackets: [(triangle, top),(square, bottom)]. The problem with the code of classical neural networks is that it provides neither for the equivalent of brackets nor for the rearrangement of symbols. This is a fundamental problem with the classical neural network code: it has no flexible means of constructing higher-level symbols by combining more elementary symbols. The difficulty is that simply coactivating the elementary symbols leads to binding ambiguity when more than one composite symbol is to be expressed. [von der Malsburg, 1981, p. 96]

Examples of the binding problem are bistable figures such as Necker’s cube and Jastrow’s duck-rabbit, where the exact same visual features of the stimulus lead to two incompatible representations, depending on how these features are bound together. Since the availability of different representations essentially depends upon the geometric properties of the figure, rather than upon the constitution of perceptual systems as would be the case, for example, for after images [Marr, 1982, pp. 25-26], bistability requires an explanation at Marr’s computational level, where properties of stimuli are described and related to information processing goals. Without a characterization of the geometric properties of the figure, and of the mappings between the figure and the two different entities which it can stand for, there would be no basis upon which to claim that the two representations are mutually exclusive.

There exist analogous cases of structural ambiguity in language:

(23) a. The woman saw the man with the binoculars.
   b. Respect remains.

Example (23a) has two alternative syntactic representations, one in which the phrase ‘with the binoculars’ is a PP attached to the NP ‘the man’ (the man that was seen by the woman had binoculars), and another in which it modifies the VP

---

15Different solutions to Rosenblatt’s problem are possible. See [von der Malsburg, 1999] for a proposal in line with the binding hypothesis and [Riesenhuber and Poggio, 1999] for an alternative approach.
(the woman used binoculars to see the man). Here too the features of the stimulus lead to two interpretations, depending on which attachment option is eventually pursued. These sentences typically result in specific neurophysiological responses, suggesting that syntactic binding is a genuine information processing problem for the brain. Sentence (23b) also has two possible parses, and this has consequences for its meaning: it can either be used as a directive speech act, if ‘respect’ is the verb and ‘remains’ the object noun; or it can be used as an assertion, if ‘respect’ is the object noun and ‘remains’ the verb.

There are some similarities between perceptual bistability in the visual and linguistic domains, such as the fact that in both cases we seem to ‘flip’ between the two incompatible representations. But there is also a deeper analogy between the two: structural ambiguity is defined at the topmost level of analysis in both cases, as [Marr, 1982, pp. 25-26] pointed out. Without an independent characterization it remains unclear why such representations are mutually exclusive in the first place. Extending Marr’s line of argument, we emphasize that the binding problem for semantics is best formulated at the computational level, although attempted solutions are bound to require significant contributions at all levels of analysis, including – perhaps most interestingly – the level of neural implementation [Hagoort, 2005; Hagoort, 2006].

ACKNOWLEDGMENTS

We wish to thank Oliver Bott, Travis Choma, Bart Geurts, Fritz Hamm, Karl Magnus Petersson, Keith Stenning, Martin Stokhof, Julia Uddén, Theo Vosse, Roel Willems and an anonymous reviewer for their useful comments on earlier versions of this chapter. We are grateful to the Netherlands Organization for Scientific Research for support under grant 051.04.040.

BIBLIOGRAPHY


1 POSITING REPRESENTATIONS

In the analysis of natural language phenomena, linguistic theories typically have recourse to representations of one form or another. Different types of representation are often posited as a means of generalising over aspects of form or interpretation as displayed in natural language constructions, and these are frequently invested with considerable theoretical significance. There are proposed representations of structure at all levels of linguistic systems: sounds, words, sentence strings, as well as representations of the meanings of words, sentences in the abstract and uttered sentences, and even representations of other people’s intentions. Such a representationalist stance was firmly set in place by Chomsky [1965] as part of, indeed the central core of, cognitive science, with language defined as a system of principles for correlating phonological representations (on some abstraction from phonetics) with some representation of interpretation (on some abstraction from denotational contents), via mappings from a central syntactic system. In such an approach, more than one level of representation may be posited as interacting in different ways with other types of representation; for example the deep structure and surface structure levels of syntax of Chomsky [1965] were taken to interact in different ways with other types of representation, in particular semantics and phonology.

Chomsky’s move towards the explicit representation of linguistic properties as part of human cognition came to be assumed almost universally within theoretical linguistic frameworks, whether formally characterised or not. But all details remain controversial, as there are no a priori constraints on the number of levels or types of representations that may be posited to account for natural language phenomena, nor on the modes of interaction between them. A multiplicity of types and levels of representation, however, threatens to result in a characterisation of

---

1 This chapter in part reports work to which many people have contributed. We thank in particular, Eleni Gregoromichelaki, Wilfried Meyer-Viol, Matthew Purver for essential input to these developments. We also thank Andrew Gargett, Stergios Chatzikyriakidis, Peter Sutton, Graham White and many others for comments both over the years and in the preparation of this chapter. Research for this paper includes support from the Leverhulme Trust to the first author (MRF F00158BF), and from the ESRC to the second author (RES-062-23-0962).
language as merely a composite of different factors for which no integrated systematic underpinning is offered. Different types of representation (phonological, morphological, syntactic, semantic and so on) are typically associated with different types of data structure, with mapping relations needing to be defined between the different representation systems or between different levels within such systems and levels of representation in other systems. Such accounts may be descriptively successful but, their potential complexity aside, hardly provide an explanation of language as a system, for distinct types of representation may be posited \textit{a priori} for all languages, suggesting that some central organisational principle is somehow being missed.

With this worry in mind, the question raised by the postulation of representations of one type or another is what such talk of representations amounts to. A representationalist account, according to linguists’ use of that term, is one that involves essential attribution of structure intrinsic to the phenomenon under investigation. To say that the characterisation of the human capacity for language requires representations is not to establish just which systems of representation should be posited. The issue is then what the attribution of structure consists in — what underpins it and what it entails — and how many distinct types of structure have to be invoked.

First, and relatively uncontentiously, the characterisation of the sound system of a language as expressed through a phonological framework demands a system of representations that is distinct from that warranted by structural properties of language as expressed in their syntax/semantics (though issues of representation arise in phonology too: see Carr this volume). Phonology apart, the primary debate over the types of representation to be invoked takes place against two background assumptions. The first, derived principally from the structuralist practices of the early twentieth century (although with roots in the Graeco-Roman grammatical tradition), is that natural language strings exhibit syntactic structure as defined over words and the expressions that they make up within some string. This assumption is adopted almost without caveats in the major current linguistic formalisms: Categorial Grammar (e.g. [Steedman, 1996; 2000; Morrill, 1994; 2010]) Head-driven Phrase Structure Grammar (e.g. [Sag et al., 2003]), Minimalist versions of Transformational Grammar (e.g. [Chomsky, 1995, Hornstein et al., 2005]), Lexical Functional Grammar (e.g. [Bresnan, 2001]). The second is that such syntactic structures have to be defined independently of the characterisation of meaning for natural language. This second assumption is far from universally adopted (in particular not within categorial grammar), but the presumed independence of syntax and semantics is nonetheless widespread. Part of the debate about representations in linguistics is thus about whether that semantic characterisation itself has intrinsic structural properties that warrant a level of representation independent of syntactic structure.\footnote{We note in passing that this is not the only characterisation of representationalism. There are inferentialist approaches to natural language understanding, hence semantics, according to which what is taken as basic are the inferential processes associated with language use. The}
This aspect of the representationalism debate emerges out of an apparent incommensurability between natural languages and the formal languages used to articulate logics (the language of predicate logic in particular), whose grammatical properties constitute the starting point for defining the representations that are used to characterise the properties of natural languages. The problem lies in the fact that natural languages are endemically context-dependent in their construal. Each natural language has a stock of words and expressions which would seem to provide some stable input to the process of construal, but which nevertheless in all usages allow a vast array of interpretations, determined in some sense from the context in which an utterance occurs. This phenomenon does not arise with formal languages, an asymmetry which was initially analysed as somewhat peripheral, a matter to be stipulated but not of central importance. However, increasingly, theoreticians have been addressing the issues of context-dependence (see sections 4-5 below); and formal models of language are now seeking to address this core phenomenon of natural languages directly. As we shall see, it is in characterising this aspect of natural-language construal that structural aspects of interpretation seem to be needed that are distinct from what is made available by the conventional concepts of syntax.

This tension between the context-dependence of natural language construal and the context-independent interpretation of formal languages, is in principle logically distinct from any general debate about the need for representations of linguistic phenomena, but has nevertheless become intrinsically linked to that debate. Firstly, as noted above, this is because the general methodology has been to use some type of formal system to express the properties of syntactic relations which are interpreted through some decontextualised, denotational semantic machinery [Montague, 1974] or through some equally decontextualised ‘Conceptual/Intentional’ interface [Chomsky, 1986; 1995; Jackendoff, 2002]. Secondly, it results from the move instigated in Chomsky [1965] that knowledge of language (‘competence’) should be isolated from all applications of this knowledge to parsing or production (‘performance’). This separation imposes a methodology in which the primary data of linguistics are not those of language as used in context, but judgements of the grammaticality of sentences without any reference to context or the dynamics of real-time language activities.³ This methodology of taking representational aspect of language, in the sense that expressions represent their attributed denotational content, is claimed to be wholly derivative on such practices [Brandom, 1994, 2008]. On this view, the term representationalism applies more broadly than just to those linguistic models which invoke a semantic level of representation over and above syntactic structure; for Brandom, the term applies to all linguistically based forms of explanation in so far as in all of these a primitive notion of representation is induced as an intrinsic property of natural language strings. On the Brandom view any such assumption of representations is seen as reducible to patterns of usage established through inferential practices involving assertion, justification, etc. Since the debate between Brandom and his adversaries takes place wholly within philosophy of language, we do not adopt this terminology here.

³So-called ‘intuitions’ of native speakers, whose precise nature is unclear and controversial, are problematic particularly as apparent categorical judgements of (un)grammaticality, the bedrock of this sort of linguistic investigation, may be manipulated in context to acceptability (see, for
sentences rather than utterances as the core language data, with grammars constructed that bear no relation to performance considerations, has been adopted by linguists across different theoretical persuasions.\(^4\)

In this chapter, we seek first to establish in a pre-theoretic way why representations of some sort have to be invoked to characterise the properties of natural languages, and, more particularly, why representations of meaning attributed to an uttered string have to be invoked. We then turn to the formal language concept of grammar in which syntax is by definition a set of principles for inducing all and only the set of sentence-strings of the language. In applying this methodology to natural language grammar construction, grammars are defined to be systems for inducing structure for sentence-strings, so that the natural-language grammar too could be seen as analogously inducing all and only the set of grammatical strings of the language. The competence-performance distinction is essential here, to isolate a concept of linguistic knowledge commensurate with the formal language concept. Indeed, it has been assumed, broadly following the formal-language methodology, that grammars for languages comprise a set of principles that pair strings with structures inhabited by those strings, and that from those structures mappings are definable onto interpretations, thus capturing all and only the grammatical sentences of the language, paired appropriately with sentence-meanings. This, it is almost universally agreed, is the syntactic heart of the human capacity for language as expressed in a grammar, with performance factors totally excluded.

Once we have sketched this formal-language background, we will set out the arguments as to whether a distinct type of representation is warranted in addition to string-structure pairings, namely representations of such meanings as are expressible by that string. As we shall see, it is increasingly central to explanations of the meanings of natural language expressions to characterise the way in which the interpretation of a string relates to the context in which it is uttered. This has proved particularly important in attempts to articulate integrated explanations of anaphora and ellipsis, and it is here that representations of structure specific to interpretation are usually invoked. But in this exploration of context-dependence, the sentence remit of the grammar and the total exclusion of the dynamics of natural-language performance will be seen to be problematic.

Finally, we shall turn to claims that a characterisation of the dynamics of how utterance understanding is incrementally built up in context is not merely essential to the explication of natural-language interpretation, but turns out to be a possible grounding of syntactic explanations of natural language strings, reversing the generally presumed order of dependence. Moreover, according to these arguments, it is only by making this shift to a more dynamic perspective for natural-language grammars that an integrated account of context-dependent phenomena such as anaphora and ellipsis becomes possible. On this view, it is the example, [Cann et al., 2005] for a discussion of resumption in English relative clauses). See [Baggio et al., this volume] for detailed discussion and critical evaluation.

\(^4\)It is, as we shall see, a primary cause of the apparent need to posit multiple types of representation.
concept of syntactic structure independent of semantic characterisation that turns out to be epiphenomenal. So finally we shall come to the conclusion that only a minimum duo of types of representations is required after all — a pairing of structured representations of sound with structured representations of meaning, as made available by a system of growth of such representations. But this conclusion is made available only by a shift in perspective to one in which the dynamics of language processing is directly reflected in the grammar itself. In this respect, the conventional notion of competence, and its associated methodology, has to be replaced by one in which the articulation of linguistic knowledge is much more closely connected to the uses to which it is put.

2 REPRESENTATIONALISM IN LINGUISTIC THEORY

Capacity for language is seen as the definitive mental activity of human beings. Humans manipulate sequences of sounds of a language that they understand with the clear presumption of these being interpretable relative to some rule-governed system, such that one agent can apparently succeed in passing on information to another, request clarification of another, induce activity on the part of another, and so on. This is a capacity that we all share, and one that children naturally acquire very early on in their development. The question at the core of the linguistic enterprise is: how should this ability be explained? It cannot be reduced merely to the observables of the speech signal and some induced behaviour, for there is no relation between any one speech signal and some correlated behaviour. But nor can it be seen as just knowledge about sentences in some abstract use-removed sense, since there is no deterministic relation between any one speech signal, simple or complex, and some fixed interpretation. Interpretations are highly dependent on context in what seem to be multifarious ways. This is very well-known with regard to anaphoric expressions, as witness the interpretation of the pronoun *it* in (1-4):

(1) It was rock-hard.
(2) The cake had been cooked for so long it was rock-hard.
(3) Each cake had been cooked so long that it was rock-hard
(4) Each cake was inedible. It had been cooked for so long that it was rock-hard.

(1) involves a so-called indexical pronoun, with the pronoun construed as referring directly to some individual in context. (2) involves a so-called coreferential use of the pronoun, referring to some individual by virtue of the previous use of an expression that is taken to pick that individual out. (3) is a bound-variable use of a pronoun, its interpretation controlled by the quantifying expression, *each cake*, that “binds” it. In these cases, as in all others, it is the combination of an
available antecedent and containing structure that, in some sense, determine the
way that the pronoun gets to be understood.

Here problems immediately arise, for an interpretation may be decidable within
the boundary of the uttered sentence, as in (2), but it may not be, as (4) shows:
this example appears to involve bound-variable uses of a pronoun across a sentence
boundary, unless the pronouns in (4) are given a wholly different explanation from
that of (3). Then with (5), there appear to be two formulations of what is conveyed.
Either there can be said to be a bound-variable use of a pronoun across a sentence
boundary, or the sequence can be analysed as having a pronoun that picks out
some arbitrary witness of what makes the first sentence true, viz. whatever cake
was depicted to be inedible:

(5) One cake was inedible. It had been cooked so long that it was rock-hard.

In uses of this type, the pronoun establishes its construal by a process of composing
the full meaning of the first sentence and using that composed result as the basis
for interpreting the pronoun in the subsequent sentence. These are the so-called
E-type pronouns [Evans, 1980]. Given the assumption that quantifiers are a type
of expression whose interpretation is not available across sentence boundaries,
these E-type pronoun uses have been seen as constituting a distinct way in which
expressions can display dependence on the context in which they are understood.
So we appear to have at least four different types of interpretation for pronouns.

There is yet a further way in which pronouns can vary: a use in which they
apparently do not pick up on the interpretation assigned to the antecedent expres-
sion, nor on its containing environment, but on some weaker concept of sameness
of interpretation, the pronoun being interpreted in the same MANNER as its an-
tecedent. In (6), for example, the pronoun it is interpreted as though it somehow
contains a genitive pronoun, just like its antecedent (his in his bread), and fur-
thermore this “covert pronoun”, is likewise interpreted relative to its local, newly
presented subject:

(6) Sandro gets his bread to rise using a bread-machine; but Gianni gets it to
rise in his airing cupboard.

So the resulting interpretation of the second conjunct (in most contexts) is that
‘Gianni gets his own bread to rise (not Sandro’s)’. But if this is an appropriate
characterisation of this form of anaphoric interpretation, then the nature of the
antecedent provided has to be that of some representation, since attributes of this
replicated representation have to be interpreted relative to the new clausal context.

This conclusion is controversial, and, as we shall see, it remains a live is-
ssue whether representations of content are essential to capturing the nature of
anaphoric dependence at an appropriate level of generality. In the meantime,
what is certainly perplexing about the overall phenomenon of anaphora is that,
despite the robust folk intuition that anaphoric expressions are transparently de-
pendent on context for their construal, the detailed specifications of the apparently
required meanings constitute little more than a heterogeneous list of types of interpretation, at least if meanings are to be given in denotational terms, following the formal language methodology. If we are to address the challenge of modelling the phenomenon of context-dependence and the twinned flexibility of natural-language expressions themselves that makes this possible, then this heterogeneity has somehow to be explained as something other than a mere list.

Tense and aspect give further indications of the context-dependent construal of language: their uses in placing described events in time involve many complexities that depend on specifics of the eventualities being described ([Kamp, 1980], and many others since). The issue is not simply that tense is a deictic category whose interpretation depends on the time of utterance; it also displays anaphoric properties, so one tensed clause may depend on another to identify the time at which it holds. For example, in a sequence of such clauses, it may be the case that the time at which one eventuality occurs is the same as that of some other eventuality (contrary to any naive assumption that discourse is made up of eventualities simply narrated in sequence):

(7) When John came into the room, he was smiling.

Furthermore, the lexical semantics of a verb may determine how eventualities relate to each other temporally. So states and events differ with respect to how eventuality times and reference times relate, and other contextually determined subtleties of interpretation abound (see, for example, [Kamp and Reyle, 1993]).

The appearance of complex anaphoric phenomena against the background of a more general and intuitive sense of context-dependence is redolent of pronoun construal (and the pronoun/tense parallel has long been noted; [Partee, 1973]). Indeed, construal of tense has also been used in the debate over whether representations of content are essential to the analysis of anaphora [Kamp, 1980; Hamm et al., 2006; Baggio et al., this volume].

In ellipsis, another context-dependent phenomenon, the need for structural reconstruction is, in some pretheoretical sense, not in question. Indeed the motivation for positing structural representations of meaning for at least some cases of ellipsis is so clear that ellipsis has been taken by many not to be a matter of context-dependence at all, but a grammar-internal matter, with the syntactic characterisation of the string said to have some of its elements simply unpronounced:

(8) John was over-wrought about his results though Bill wasn’t.

This analysis of ellipsis as invisible syntactic structure, preserving some concept of sentencehood for the elliptical fragment, might superficially seem to be the appropriate form of explanation. However, as with pronouns, such context-dependent effects cannot be reduced to mere reconstruction of the words themselves. We know this because of the many challenges which ellipsis construal poses for the analyst [Fiengo and May, 1994]. Among these is the way in which ellipsis construal can be ambiguous, even given a single interpretation of its antecedent phrase. For example, restricting the first predicate of (8) to the interpretation that John was
over-wrought about his own results, the elliptical form wasn’t in the second con-
junct can be understood in two ways. Either Bill was not over-wrought about
John’s results (matching the predicate content of the antecedent clause, the so-
called “strict” interpretation), or Bill was not overwrought about his own results
(the so-called “sloppy” interpretation).

Moreover, despite the superficial distinctiveness of anaphora and ellipsis, this
ambiguity echoes that of anaphoric dependence. On the “strict” interpretation,
the interpretation of the predicate is exactly that of its antecedent, just as in
instances of pronominal coreference. On the sloppy interpretation, the pronoun
within the reconstructed predicate has to get reinterpreted relative to the new
subject parallelling the MODE of interpretation assigned to its antecedent, rather
than picking up on the denotational content of that antecedent. This more indirect
form of interpretation parallels the mode of interpretation needed for (6) (itself
labelled with the term “lazy pronoun” usage: [Karttunen, 1976]). In both cases, in
some sense to be made precise, it is the way in which interpretation is built up in
the antecedent that is replicated in the construal of the ellipsis-site/lazy-pronoun.
So there are clear parallelisms between types of construal in anaphora and ellipsis.

Like anaphora and tense construal, the strict/sloppy ambiguity associated with
VP ellipsis may span more than one sentence:

(9) John was over-wrought about his results. Bill wasn’t.

(9) is ambiguous in exactly the same way as (8). This phenomenon is not restricted
to particular types of expression, as there is nominal as well as predicate ellipsis;
for example, in (10) most is construed as ‘most of my students’, again proving to
be both intra-sentential and super-sentential:

(10) I wrote round to all my students, though most were away.

(11) I wrote round to all my students. Most were away, but a few answered.

However, the issues are broader than some single sub-type of ellipsis, indeed
broader than the objective of explaining variability in anaphora and ellipsis. The
challenge for linguistic explanation posed by context-dependence in general is that
of formulating what it is about expressions of language that enables them to make
some single systematic contribution to the way information is conveyed while nev-
ertheless giving rise to multiple interpretations. In the case of anaphora and,
even more strikingly, ellipsis, this would seem to necessitate some form of rep-
resentation other than that inhabited just by the linguistic string of words. As
the “lazy”/“sloppy” forms of pronoun/ellipsis construal show, the input to con-
strual may involve adopting some particular interpretation of an antecedent string
and building up structural properties in parallel with that to yield some novel
interpretation. In addition, it is far from obvious that the remit of characteris-
ing sentence-internal phenomena provides a natural basis for characterising such
context-dependence, since all of these various context-dependence phenomena can
be associated with an antecedent either sentence-internally or sentence-externally.
Indeed, if we stick to identification of sentence-internal dependencies as the exclusive remit of grammar-internal characterisations, leaving sentence-external dependencies on one side, all context-dependent phenomena will fail to receive complete characterisation.

Further problems arise when the remit of data to be considered is extended to conversational dialogue, where ellipsis is rampant. Both speakers and hearers have to be able to manipulate decisions relative to the context in which the communication is taking place; indeed, the planning and linearization of some intended thought may make essential reference to the way choices made are relative to the context. In conversational dialogue, these decisions, made severally by speaker and hearer, appear to come together, giving rise to what seems to be yet a further type of ellipsis. Utterances can be split across more than one person, with the speaker who is interrupting her interlocutor transparently building upon the point which that interlocutor has reached:

(12) Father: We’re going to Granny’s.
    Mother: to help her clean out her cupboards.
    Child: Can I stay at home?
    Mother: By yourself? You wouldn’t like that.

In (12), the mother seamlessly extends the sentence initiated by the father by providing a non-sentential addition, continuing the narrative for the child in ways that are dependent on what is previously uttered (here understanding the subject of the nonfinite *help* as the mother and father, and the pronoun *her* as Granny). The child, having heard both, then interacts with his parents by asking his own question, which the mother responds to with another fragment, *by yourself*, as an extension of his utterance. This is done in the full expectation that the child will understand her use of the demonstrative *that* as referring to the proposal that he should stay at home by himself, even though this proposal exists in the context only as a result of the extension of the child’s suggestion by the mother’s utterance; it has not been expressed by any one participant in the conversation. Such an interpretation essentially depends on the coordinated accumulation of information by both parties, each building on the structure to date, whether as speaker or as hearer.

Seen as a phenomenon of ellipsis, this is no more than an illustration of the context-relativity of the construal of elliptical fragments; but in these cases such relativity strikingly includes dependence on partial structure provided by the context, which is essential to determine the wellformedness of the whole as built up from the severally uttered parts. In this connection, the last exchange between mother and child has special significance. It shows that the concept of structure presumed upon by both participants in the building up of interpretation is not that of the string each utters, but is rather some representation of the content that their strings give rise to. What the two participants — mother and child — utter is *Can I stay at home all by yourself?*. It is clear that neither child nor mother takes what they have jointly been building to be a structure decorated
by these words: as a sentence, this sequence is plainly ungrammatical, but what 
the interlocutors understand each other to have said is taken by both to be fully 
wellformed (as evidenced by the unproblematic subsequent use of anaphoric that 
discussed above). Looked at in more abstract linguistic terms, the issue here is 
what kind of structure crucial notions like ‘locality’ are defined over. Reflexive 
pronouns like yourself are conventionally defined as requiring identification of an 
antecedent determining their interpretation within a locally identifiable structure. 
This example provides evidence that the structure over which such locality has 
to be checked is not that of the words themselves. If it were, the result of the 
exchange between child and mother would not be wellformed: there is mismatch 
in person specification between the reflexive pronoun and the form of wording in 
which its antecedent has been presented. Since this kind of exchange is wholly 
acceptable, it must instead be the case that the relevant notion of locality is de-
defined over some representation of interpretation successfully constructed by the 
conversational participants. Hence the necessity for structures representing the 
meanings established by the use of linguistic expressions.

Note that this is a general phenomenon: such speaker-hearer role switches can 
take place across any syntactic dependency whatsoever:

(13) A: Has John B: read any books on the reading list? [negative polarity dependency]

(14) A: I need the.. B: mattock. [determiner-noun dependency]

(15) A: I think she needs B: to see a doctor. [PRO control dependency]

As these examples illustrate, it is characteristic of conversational dialogue that 
the construal of fragments may involve both recovery of content and recovery of 
structure, each participant adding a subpart to what becomes a shared whole. 
This phenomenon is so widespread that it is arguably diagnostic of human con-
versational dialogue. The question is: what does all this amount to in the search 
for what constitutes the human capacity for language, or knowledge of language?

Until very recently, such dialogue data have been totally ignored by linguists as 
no more than a demonstration of dysfluencies observable in language use. Being 
designated a performance matter, such data lie quite outside the normal canon of 
data familiar to linguistic argumentation (though see [Purver et al., 2006; Cann 
et al., 2007; Gargett et al., 2009; Kempson et al., 2009]). Yet this phenomenon 
is wholly systematic, forming a large proportion of the sole type of data to which 
young children are exposed in learning language. If we take seriously the challenge 
that grammars of natural language must be definable as licensing systematicity 
of structure in the data the language-acquiring child is exposed to, then in these 
dialogue data we are brought up against evidence that models of linguistic knowl-
edge must include representations of content as part of the vehicle of explanation. 
For cases such as (12)–(15) show that these, not representations inhabited by the
string of words, are the representations over which syntactic restrictions have to be defined. So the conclusion here buttresses that reached through more conventionally recognised forms of ellipsis: structural representations of content and context are essential to the characterisation of such fragments as part of a general explanation of the context-dependence that natural languages invariably display.

Of course, there is always the option of ignoring these data, simply deeming them to be a dysfluency of performance, and in principle not within the remit of a conventional competence-based form of explanation. In particular, the phenomenon of split utterances is at odds with a grammar formalism in which the concept of sentence is the central notion. Standardly, all dependencies to be characterised within the grammar are defined relative to sentence-sized units independently of any application of them, and without any concepts of underspecification and growth.

It is important to recognise that such a decision comes with a cost. The effect will be that no single linguistic phenomenon will receive a complete characterisation. On the one hand, accounts of context-dependent phenomena such as anaphora and ellipsis will be incomplete, bifurcated into those phenomena which can be characterised sentence-internally, and those which cannot. On the other hand, the structural dependencies themselves will suffer the same fate: not a single one of them will be fully characterised by the grammar, for such a characterisation will not include the use of fragments in discourse in which participants shift roles mid-sentence. By the same token, we cannot look to performance mechanisms to unify such phenomena: as long as we stick to the sentence-based view of competence, the grammar will explain what it can, and stop short of the rest. Note that, on conventional assumptions, fragmentary utterances like those in (12)–(15) do not even provide a grammatical trigger for performance mechanisms. Not being wellformed sentences, these fragments simply will not form part of the output of the grammar. There is therefore no possibility of even a degree of continuity between such examples and the more widely recognised kinds of ellipsis that they resemble in significant ways.

This failure of the grammar to provide any coherent basis for an account of the general phenomenon of context-dependence is a serious defect, since it constitutes a failure to address the very phenomenon which critically distinguishes natural languages from formal languages. It is thus this conception of the competence-performance distinction, and the divide it imposes on the data to be characterised by taking the remit of grammar to be sentence-internal phenomena, to which we shall ultimately have to return in seeking to include some explanation of the phenomenon of context-dependence as an essential objective for any model of natural language. But this is to leap ahead. To see how and why certain representationalist views of language emerged as a result of articulating formal grammar specifications of natural language, we have to go back to the point of departure for core notions in syntax and semantics — the grammars of formal languages.
3 SYNTACTIC AND SEMANTIC REPRESENTATIONS

3.1 Formal languages and the form of grammar

The classical formal languages of propositional and predicate logic were defined not for the study of language but for the formal study of mathematical inference, though predicate logic incorporated a partial reflection of natural-language structure in its articulation of subpropositional predicate-argument structure. Logic is the formal modelling of inference, which involves truth-dependence between well-formed formulae of the defined language. The objective in defining such a formal language is to capture all and only the valid inferences expressible in that language, via some concept of truth with respect to a model. The task is to posit a minimal number of appropriate units and structure-inducing processes that together yield all and only the appropriate outputs of the grammar, viz. the infinite set of wellformed strings, over which the inferences under study can be defined. The objective is to derive the infinite complexity of valid inference patterns from the interaction of a small number of primitive notions.

The perspective of mathematical inference imposes an important restrictive capacity: it is global and static. There is no modelling of context external to the structure being defined — mathematical truths are by definition independent of context. There is no modelling of growth of information or of its corollaries, underspecification of content and the concept of update. In fact, the flow of information is in exactly the opposite direction: rather than building information, inference involves only what follows from some information that is already given, the premises. There are therefore fundamental reasons to doubt that the methodology of describing these formal languages could ever be sufficient for modelling natural languages. If the interpretation of expressions of natural language necessarily involves the building up of information relative to context, then a formal explication of this process is required. Models based in mathematical inference will not provide this, even though insights as to the nature of inference undoubtedly form a sub-part of a full characterisation of natural language interpretation.

Despite its restrictiveness, the methodology for articulating formal languages has transformed the landscape within which formal frameworks for natural language and natural language grammars have developed; and the assumption of a truth-conditional basis to semantics for natural language is very widely adopted. In predicate logic, the grammar defines a system for inducing an infinite set of propositional strings which are taken to be truth-value denoting; and sentence-sized units are defined as having predicate-argument structure made up of predicate forms and individual constants, with naming and quantificational devices. Syntactic rules involve mappings from (sub)-formulae to (sub)-formulae making essential reference to structural properties: these rules constitute a finite set of principles inducing an infinite set of strings. Semantics is the provision of an algorithm for computing denotations for arbitrarily complex strings: the result is a pairing of strings and the objects they represent. This pairing is determined on
the basis of some notion of content assigned to elementary parts, plus rules that
determine how such contents are to be composed, through stepwise correspondence
with syntax, yielding the values true and false as output.

The pattern provided by such formal languages was famously extended to nat-
ural language semantics by Montague [1974], who argued that natural languages
could be characterised as formal languages, with semantics defined in terms of
reference/denotation/truth with respect to a model. To achieve this, Montague
developed a program of logical types and formulations of content for expressions of
the language which are defined and articulated in the lambda calculus. These were
defined to enable the preservation of predicate logic insights into the meanings to
be assigned to these expressions even while sustaining the view that composition of
such meanings is determined from some conventional analysis of the syntax of the
language. Consequently, the natural language grammar, like a formal language, is
conceived of as a system that induces an infinite set of strings paired with deno-
tations, where these denotations are determined by semantic rules which directly
match the combinatorial operations that produce the strings themselves. For ex-
ample, a functional abstract can be defined using a functional operator, the λ
operator, which binds an open variable in some propositional formula to yield a
function from individuals to propositional contents, as in λx.x smiled. If we take
this to be the contribution of the intransitive verb smiled, and we take a constant of
type ⟨ε⟩, john, to be the contribution of the noun phrase John, then it is clear that
this allows semantic composition to mirror the combination of the words yielding
the string John smiled. At the same time, a further functional abstract can be
defined in which a predicate variable is bound in some open propositional formula
to yield a function from properties to propositional contents, as in λx.x smiled. If we take
this to be the contribution of the intransitive verb smiled, and we take a constant of
type ⟨ε⟩, john, to be the contribution of the noun phrase John, then it is clear that
this allows semantic composition to mirror the combination of the words yielding
the string John smiled. At the same time, a further functional abstract can be
defined in which a predicate variable is bound in some open propositional formula
to yield a function from properties to propositional contents λP.P(john) (semant-
ically, the set of properties true of John, or the set of classes that include John of
type ⟨⟨e,t⟩,t⟩). This is equally able to combine with the predicate λx.x smiled to
yield the proposition expressible by John smiled, only in this case the contribution
of the subject is the functor and that of the verb is the argument.

As even these simple examples show, there are in principle a number of differ-
ent modes of combination available, involving different functor/argument relations
— with potentially many ways of deriving the same string-meaning pair as more
complex sentences are considered. If this approach is applied with no indepen-
dent constraints on syntactic analysis, the syntactic structure assigned to strings
is effectively epiphenomenal, being no more than a vehicle for the semantic opera-
tions: this account is notably espoused in categorial grammar formalisms [Moort-
gat, 1988; Morrill, 1994; 2010, this volume; Steedman, 2000]. These grammars are
non-representationalist in that, on the one hand, the semantics of strings is defined
in denotational terms (in terms of individuals, sets of individuals, and functions
on those which ultimately map on to concepts of truth and inference); and, on
the other, the rules of syntax constitute nothing more than mappings from strings
onto denotationally interpreted strings (that is to say, mappings from strings to
strings suitably paired with mappings from denotational contents to denotational
contents). Any invocation of structure is then no more than a convenient way
of talking about such pairwise mappings of strings and assignable denotational contents.

Even without adopting this strict variant of the Montague claim about natural languages as formal languages, the influence of the formal-language methodology holds sway very generally. On a more pluralistic model of natural-language grammar — influenced by Montague solely with respect to semantics — a natural language grammar is a finite set of rules which assigns structure to sentences, and it is these syntactic structures to which denotational interpretations are assigned, defined in terms of truth with respect to a model. Semantic operations are thus defined in tandem with syntactic ones, most transparently applied in the so-called ‘Rule-to-Rule Hypothesis’ whereby each structure-defining syntactic rule is paired with an appropriate semantic one. This is the dominant model in work within theoretical linguistics that labels its object of study ‘the syntax-semantics interface’ (e.g. [Heim and Kratzer, 1998]) — which is to say most work that purports to be formally explicit about both syntactic structure and its interpretation. In terms of representationalism as a whole, this constitutes a mixed approach. The view of syntax is representationalist, in that there are assumed to be fixed structures defined over strings of words. But the semantics is not representational, at least if conceived of in terms of conventional logic formalisms, because the semantic characterisation assigned to each syntactic structure is given in terms of denotation with respect to a model (or some equivalent).

It remains a matter of controversy in linguistic theory whether syntactic and semantic operations can be directly matched in this way. While for some analysts the direct assignment of content to syntactic structures remains an ideal worth striving for, others work on the basis that this is demonstrably impossible for natural languages. Broadly, speaking, there are two common kinds of claim for the necessary divergence of syntactic and semantic structures, necessitating multiple levels of representation. One is that the interpretation of natural languages requires representations of meaning that are not directly interpretable in terms of denotations. This relates to issues of context-dependence, and we return to it in section 4. The other kind of claim is that natural language syntax has distinctive properties that are neither reducible to, nor reflected in, its semantics; this is our next topic of discussion.

3.2 The syntax-semantics split

As we saw in a preliminary way in section 1, the phenomena of anaphora and ellipsis and their interaction in conversational exchanges provide informal evidence that in order to interpret natural language expressions, structure may have to be built up over and above that presented by the expressions themselves. The minimal stance, as we set it out there, was that (contrary to the Montagovian position outlined in section 3.1) representations of content must indeed be posited, in addition to representations of the words themselves and their phonological properties. We return to this claim in more detail in section 4. Assuming for now that it
is correct, an obvious question arises with respect to the economy of the overall system (though it is rarely asked): do sequences of words (i.e. sentences) need to be assigned representations of structure over and above whatever representations of content might be attributable to them?

At the level of the observable data, the answer is very widely taken to be yes. This assumption is generally held to follow from the incontrovertible fact that there is no one-to-one correspondence between a string and an assignable interpretation. In the one direction, there are clearcut cases where expressions have more than one interpretation, and these are taken to warrant the invocation of discrete tokens; i.e. so-called structural ambiguities. More tellingly, in the other direction, there are strings which systematically differ in some structural properties but have identical interpretations, at least at the level of denotational content. These are the cases of so-called discontinuity made famous by Chomsky [1957], which feature a pair of sentences that express the same semantic content, but are asymmetrically related to each other in a particular way: one appears to have structure that can be mapped directly onto the associated semantic content, while the other seems to map onto this content only by making reference to the structure of the first sentence (at least so Chomsky [1957, 1965] and others argue). There are both local and non-local variants of this phenomenon. The first of these is displayed by so-called expletive pronouns:

(16) That Eliot will bring food for the party is likely.

(17) It is likely that Eliot will bring food for the party.

(18) A man is singing outside in the garden.

(19) There is a man singing outside in the garden.

These examples show relatively local discontinuity between the expression in question, here the expletive pronoun, and the linguistic material providing its interpretation (in (17), for example, the interpretation assigned to the pronoun is provided by the end-placed clause *that Eliot will bring food for the party*).

Non-local cases, like (20) and (22), have famously been said to involve movement from one structural position to another. Here, the discontinuity between a certain expression and the site at which it is assigned an interpretation may span an indefinitely long intervening sequence: *what* in (20) and *the new book by Sue* in (22) must each be interpreted as the internal argument of the predicate given by the verb *read*, and so, by hypothesis, must be related to the normal, postverbal position of the syntactic object of *read* (which is shown in the position of these expressions in (21) and (23)):

(20) What did John say that we should all read by tomorrow?

(21) John said that we should all read what by tomorrow?

(22) The new book by Sue, John said we should all read by tomorrow.
(23) John said we should all read the new book by Sue by tomorrow.

Arguably, both (20) and (21) express a question about what John said should be read by tomorrow, yet in (20) this involves front-placement of the expression what. In like manner, (22) and (23) do not essentially differ in interpretation. There thus appear to be structural properties of strings that have no necessary counterpart in their semantic characterisation, where this is being presumed to be a characterisation of conditions for the truth of such strings.

Moreover, such ‘displacement’ effects have other properties that have been claimed to necessitate separate representations for syntax and semantics. Since the work of Ross [1967], linguists have investigated apparent constraints on the locations of certain expressions, relative to the positions in which they seem to be interpreted (often conceived of as constraints on syntactic movement). These constraints appear to be inexpressible in semantic terms, and so are taken to warrant an independent level of representation for syntax. For example, according to one of the most well known of these constraints, the Complex NP Constraint, expressions cannot be construed across a relative clause boundary. Though (20) is well-formed, (24) is not:

(24) *What did John introduce us to the American lecturer who said that we should all read by tomorrow?

To understand why this is taken to demand the separation of syntactic and semantic representations, it is important to note that the conventional tools of formal semantic analysis fail to predict the illformedness of (24). Standard methods definable using the lambda calculus allow the material following what to be treated as a form of predicate abstract, with the lambda operator binding a variable (the internal argument of read) at an arbitrary depth of embedding. This means that what can be related to the object of read in (24), just as in (20). More specifically, it should be possible to treat what as a functor of type $\langle\langle e, t \rangle, t \rangle$, which will apply the relevant lambda-abstract to yield an overall interpretation in which what queries the object of read. There is nothing on the semantic side that blocks this in the case of (24). In essence, the problem is that the lambda calculus is blind to syntactic details such as the presence or absence of a relative clause boundary. It would seem to follow that semantics and syntax require distinct forms of generalisation, expressed in different vocabularies. Indeed the sensitivity of some phenomenon to the Complex NP Constraint is commonly taken to be diagnostic of its syntactic underpinnings (see e.g. [Partee, 1976; Merchant, 2009; Stainton, 2006]).

We have thus arrived at a point where phonology, syntax and semantics are widely accepted to require discrete forms of generalisation.\footnote{It is not standardly argued in phonology that structures in phonology should be reducible to syntax.} As such, the overall tendency towards representationalism is solidly established. Given this, it has seemed to many linguists a relatively small step to posit additional types of representation in a grammar, any one of which by definition requires a different form
Representationalism and Linguistic Knowledge

of vocabulary. This is manifested in a framework like Lexical Functional Grammar (LFG, [Bresnan, 2001]), which posits separate levels of representation for a range of putatively distinct types grammatical information, each of which is expressed with wholly distinct vocabulary, and furthermore with transition rules in yet a further vocabulary mapping representations at one level onto those at another [Dalrymple, 1999]. So we have c(onstituent)-structure representations with enriched ‘functional equations’ to derive the mapping from this type of representation onto f(unctional)-structure, which encodes the grammatical functions played by the various constituents and is expressed in an entirely distinct vocabulary. Further rules relate f-structures to semantic structures and other types of representation have been variously posited as necessary to correctly characterise the grammatical properties of natural languages. LFG is by no means unique in this tendency, however, nor the most extreme at positing distinct modes of grammatical representation (cf. for example Role and Reference Grammar, [van Valin and LaPolla, 1997]), but it illustrates one end of a spectrum of representationalist approaches.

There is a distinction at this point which is important to bear in mind. The issue of how many levels a grammar defines is independent of the issue of types of representation advocated. In one direction, discrete levels may be posited relative to a single type of representation, as in the transformational grammar of the nineteen-seventies with its deep and surface structure levels, both of which were expressed as syntactic forms of representation, articulated in a common vocabulary. In the inverse direction, a grammar may be mono-level (or ‘mono-stratal’), in which a single format of presentation is adopted. In Head-driven Phrase Structure Grammar (HPSG, [Pollard and Sag, 1994]), for example, all grammatical information is encoded in the Attribute Value Matrix (AVM) notation and linguistic signs are considered to be fully defined by such matrices, but this apparent single level of representation nonetheless masks the fact that wholly discrete vocabularies are used within various parts of an AVM, with different modes of unifying information as the signs for individual words and phrases are combined. These different vocabularies correspond to disjoint categories, with distinct forms of generalisation expressible within them, effectively giving rise to different types of representation for different categories of grammatical information. So the issue of how many types of representation to advocate within grammar formulation cannot be reduced to the decision as to how many levels of representation are considered to be required. Decisions over how many levels of analysis to posit are driven by the objective of capturing generalisations in the most revealing manner. But the issue of how many types of representation to invoke is a claim as to the richness of the ontological base of natural language as a cognitive phenomenon, hence is intrinsically a foundational issue.

The issue of the relative independence of syntax and semantics has constituted the core of the representationalism debate as argued within linguistic theorising, if simply because of the live issue of whether generalisations about natural language interpretability require the articulation of representations of content at all. As we
shall see, as conventionally expressed, this turns out to be the wrong question to ask. But let us first examine the issue as it is usually formulated.\(^6\)

4 REPRESENTATIONS IN SEMANTICS?

The argumentation about distinguishing syntax and semantics that we considered in the previous section is usually taken to show that syntax requires a securely distinct type of representation, one that, following Chomsky [1965], acts as the core of the grammar. Its uniquely defining vocabulary, involving tree structures, category specifications, and whatever other features are taken to be needed to express appropriate syntactic generalisations, articulates a system of rules that is intended to license all and only the grammatical strings of some language independently of context. Such a stance is broadly in line with the formal language conception of syntax.

Leaving on one side more general problems associated with the concept of defining grammaticality in terms of the well-formedness of sentences [Pullum and Scholz, 2001], there are significant problems with this view of language when we turn to a fuller consideration of semantics. Right from the outset of the formal semantics programme, the need for some characterisation of context-dependence was recognised, but at least initially the problem was taken to be essentially peripheral to the core insight that truth-conditions can be presented as a basis for articulating natural-language semantics relative to orthodox assumptions of syntax [Lewis, 1972]. However, as research on the problems posed by context-dependence have deepened (from [Kamp, 1981] onwards), there has been increasing emphasis on the on-line dynamics of language processing and update; and such moves are almost diametrically opposed to the perspective in which syntax and semantics are defined in tandem, with reference only to sentences. Yet, as we saw informally in section one, there is evidence that in conversation interlocutors rely on some notion of representation that allows them to interpret an elliptical utterance, extend or complete an utterance, clarify some aspect of what has been said, and so on. Moreover this notion of representation must involve attribution of structure in order, for example, for reflexives to be used coherently in split utterances.

In coming to grips with the challenges posed by such context-dependence of content, linguists have begun to ask the further question: what is it in the nature of linguistic expressions that enables their apparent intrinsic specification

\(^6\)Morphology, and the status of morphological sub-word generalisations, is another case in point. We do not address this here, but note the debate within minimalism as between those who take the set of morphological phenomena to require wholly disjoint vocabulary specific to morphology, e.g. the positing of \textit{morphological templates}, and those who take the set of phenomena constitutive of morphology to be broadly syntactic, with explanations involving movement relative to appropriate feature-specified triggers. There is also the stance more pertinent to the issues raised in this chapter in which it is argued that the major morphological generalisations thought to provide evidence for such morphology-specific phenomena such as morphological templates can be expressed solely as a mapping from phonological sequences onto representations of content (see [Chatzikyriakidis and Kempson, to appear, 2011]).
of content to interact with aspects of context to determine their particular con- 
strual in a given linguistic situation (see in particular [Cooper, 2005; this volume; 
Larsson, 2008])? In other words, what sort of underspecified content can be at-
tributed to lexical items and what sorts of updates can be provided by linguistic 
and non-linguistic context to derive the meanings of words and expressions that 
interlocutors understand in particular conversational exchanges? This provides 
another source of tension with the formal-language specification of the relation 
between syntax and semantics, as the assumptions of that methodology are no 
more set up to express a concept of underspecified meaning or its update than 
they are to express dependence on a structured concept of context. Moreover, the 
static and context-independent conception of linguistic competence is destabilised 
by any attempt to accommodate underspecification and update, since these no-
tions require that the model of competence takes account of processes which apply 
when language is used in context.
The first moves in the direction of this form of analysis were taken by seman-
ticists modelling the nature of information update as the basis for language un-
derstanding. These moves have led to ever-increasing richness of structure in the 
formal specifications of both content and context, in particular in order to explain 
phenomena such as anaphora and ellipsis. As we shall see, it is this trend which 
enabled the question of relative dependence between syntax and semantics to be 
re-opened, with the potential for new answers.
As the initiator of this movement, Kamp [1981] set out two points of view of 
what meaning is, each of which has been assumed in different parts of the linguistic 
literature:

(25) (a) meaning is what determines truth conditions; the view taken by 
truth-theoretic and model-theoretic semantics
(b) meaning is what language users GRASP when they understand the 
words they hear. This representationalist view of meaning is in, 
principle, what the majority of psychologists, computer scientists, 
linguists, and others working in cognitive science, aim to 
characterise.

The first view of meaning is what Montague Grammar promulgated; and many 
philosophical accounts concerned with reference and truth can be taken as espous-
ing this view. It articulates a concept of a language as a set of interpreted symbols 
wholly independent of any agent’s knowledge of it. On the other hand, the rep-
resentationalist point of view which Kamp advocates within semantics involves 
defining formal constructs which are assumed to model the mental representations 
humans employ in response to linguistic processing. This is the view taken by var-ious cognitive science and linguistic approaches, with perhaps the most prevalent 
being the so-called computational theory of mind. On this view, the mind is a 
system for manipulating symbols according to syntactic rules which determine the 
recursive complexity of thoughts a so-called ‘language of thought’ [Fodor, 1975; 
1983]. On this view, human cognition operates systems of internal cognitive rep-
resentations (possibly more than one) enabling individuation of objects for the purposes of mind-internal reasoning [Fodor, 1983] and subsequently, [Hamm et al., 2006; Baggio et al., this volume].

Discourse Representation Theory (DRT) was the first theory which aimed to combine the two approaches to meaning listed in (25); DRT is motivated by the need to give an account of how discourse processing leads to the generation of a representation of the semantic content of a discourse.

4.1 Discourse Representation Theory

The immediate objective of Kamp [1981] was to articulate a response to the challenge of modelling anaphoric dependence in a way that enables its various uses to be integrated, contrary to the simple postulation of ambiguity to account for different modes of interpretation (as in the indexical, bound-variable, or E-type interpretations of pronouns exemplified section 2). Sentences of natural language were said to be interpreted by a construction algorithm for interpretation which takes the syntactic structure of a string as input and maps this onto a structured representation called a Discourse Representation Structure (DRS). This constitutes a partial model for the interpretation of the natural language string which contains named entities (discourse referents) introduced from natural language expressions, and predicates taking these as arguments (conditions on referents). The sentence relative to which such a partial model is defined is said to be true as long as there is at least one embedding of the model so constructed into an overall model. So, for example, for a simple sentence-sequence such as (26), involving an E-type use of a pronoun,

(26) John loves a woman. She is French.

the construction algorithm induces a DRS for the interpretation of the first sentence containing discourse referents corresponding to the name and the quantifying expression, together with a set of predicates corresponding to the verb and noun:

```
x, y
John = x
loves(x, y)
woman(y)
```
Such a DRS can then be extended by applying the construction algorithm to the second sentence, extending the initial DRS to an expanded DRS:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$x, y, z$</td>
<td></td>
</tr>
<tr>
<td><strong>John</strong></td>
<td>$=$ $x$</td>
</tr>
<tr>
<td><strong>loves</strong></td>
<td>$(x, y)$</td>
</tr>
<tr>
<td><strong>woman</strong></td>
<td>$(y)$</td>
</tr>
<tr>
<td><strong>$z = y$</strong></td>
<td></td>
</tr>
<tr>
<td><strong>French</strong></td>
<td>$(z)$</td>
</tr>
</tbody>
</table>

On this account, indefinite NPs are defined as introducing a new discourse referent into a DRS, while definite NPs and pronouns require that the referent entered into a DRS be identical to some discourse referent already introduced. Names, on the other hand, require a direct embedding into the model providing the interpretation. As noted above, once constructed, a DRS is evaluated by its embeddability into a larger model, being true if and only if there is at least one embedding of that partial model within the overall model. Notice that the basis for semantics is relatively conservative, in that the grounding model-theoretic assumptions are preserved unaltered.

Even without investigating further complexities that license the embeddability of one DRS within another (and the famous characterisation of *If a man owns a donkey, he beats it*), an immediate pay-off of this approach is apparent. E-type pronouns fall under exactly the same characterisation as more obvious cases of co-reference: all that is revised is the domain across which some associated quantifying expression can be seen to bind. It is notable in this account that there is no structural reflex of the syntactic properties of the individual quantifying determiner; indeed, this formalism was among the first to come to grips with the name-like properties of quantified formulae (see also [Fine, 1986]). It might seem that this approach obliterates the difference between names, quantifying expressions, and anaphoric expressions, since all lead to the construction of discourse referents in a DRS. But these expressions are distinguished by differences in the construction process. The burden of explanation for natural language expressions is thus split: some aspects of their content are characterised in essentially dynamic terms, in the mode of construction of a representation (the intervening DRS), while some aspects are characterised in more traditional semantic terms, through the embeddability conditions of that structure into the overall model.

The particular significance of DRT lies in the Janus-faced properties of the DRS’s thus defined. On the one hand, this intervening level is a partial model — or, more weakly, a set of constraints on a model — defined as true if and only if it is embeddable in the overall model, and hence essentially the same type of construct. On the other hand, it is argued [Kamp and Reyle, 1993] that the specific structural properties of the DRS are needed in defining the appropriate antecedent-pronoun relation, hence such a level constitutes an essential intermediary between
the natural language string and the denotations to be assigned to the expressions it contains. Nonetheless, this intermediate level has a fully defined semantics, via its embeddability into an overall model.

Despite the explicit model-theoretic characterisation provided by Kamp and Reyle, the DRT account of anaphoric resolution sparked an immediate response from proponents of the model-theoretic tradition. Groenendijk and Stokhof [1991] argued that the intervening construct of a DRS was not only unnecessary but illicit, in making compositionality of natural language expressions definable not directly over the natural language string but only via this intermediate structure (see also [Muskens, 1996]). Part of their riposte to Kamp involved positing a new Dynamic Predicate Logic (DPL) with two variables for each quantifier and a new attendant semantics: one of these variables is closed off in ways familiar from predicate-logic binding, but the second remains open, bindable by a quantifying mechanism introduced as part of the semantic combinatorics associated with some preceding string. This was argued to obtain cross-sentential anaphoric binding without any ancillary level of representation as invoked in DRT (see Asher, this volume). Such a view, despite its novel logic and attendant semantics, sustains a stringently model-theoretic view of context-dependent interpretation for natural language sentences commensurate with e.g. Stalnaker [1974; 1999]): the progressive accumulation of interpretation across sequences of sentences in a discourse is seen exclusively in terms of intersections of sets of possible worlds progressively established, or rather, to reflect the additional complexity of formulae containing unbound variables, intersection of sets of pairs of worlds and assignments of values to variables (see [Heim, 1982], where this type of program is set out in detail). However, the DPL account fails to provide an explanatory account of anaphora, in that it merely articulates a semantics for the outcome of the interpretation process, and it does not address Kamp’s objective of modelling what it is that pronouns contribute to interpretation that makes such diversity of resulting interpretations possible. It was this assigned objective which led Kamp to postulate a “DRS construction algorithm” with subsequent evaluation rules determining the embedding of the resulting structure within the overall model. It was only this way, he argued, that one could capture both the denotational diversity of anaphoric expressions in context and the intrinsic uniformity of such expressions within the language system.

Over and above the departure from orthodox semantic theories in shifting to representationalist assumptions, there is a further sense in which DRT constitutes a radical departure from previous theories. In defining a formal articulation of the incremental process of building an interpretation of discourse, relative to some previously established context, there is an implicit rejection of the severe methodology whereby no reflex of performance should be included in any specification of aspects of natural language competence. Indeed the construction algorithm for building DRS’s yields a formal reflection of the sentence-by-sentence accumulation

---

7The DPL account explicitly relies on some presumed independent indexing which the DPL characterisation is defined to reflect [Groenendijk and Stokhof, 1991].
of content in a discourse (hence the name *Discourse Representation Theory*). So DRT not only offers a representationalist account of natural language meaning, but one that reflects the incrementality of utterance processing, albeit one that takes clauses as the basic units of information and organisation in a text.

There remains a lurking problem in this move, however, of direct relevance to the representationalism issue. In assigning meaning to a sentence string, the DRS construction algorithm was defined to take as its input a syntactic structure for that string as articulated in some independently defined syntactic framework (for example some form of Phrase-Structure Grammar as in [Kamp and Reyle, 1993]), and progressively replace that structure with a DRS. Hence the explicit advocacy of more than one type of representation, semantic as well as syntactic. Yet, looked at as a basis for a uniform account of structure in language, DRT lacks only some characterisation of those phenomena which linguists take to be within the purview of syntax; and many of these involve just the kind of interaction with anaphora which one might expect should fall within the remit of DRT-based explanations, given the assigned aim to provide a unitary account of anaphora. There is, for example, no characterisation of long-distance dependency, expletives or quantifier scoping; and so, equally, there is no account of their systematic interaction with anaphora. If, however, such phenomena could be reconstructed in terms that build on the intermediate level of representation which DRT does indeed provide, the situation would be very different; for in that hypothetical scenario it would not be the level of content representation that would be redundant, but the level of structure as inhabited by the string. It is this scenario we shall shortly seek to actualise.

In the meantime, the debate between advocates of DPL and DRT is far from over; there has been continuing debate as to whether any intervening level of representation is justified over and above whatever syntactic levels are posited to explain structural properties of natural language expressions, preserving orthodox assumptions of syntax. Examples such as (27)-(28) have been central to the debate [Kamp, 1996; Dekker, 2000]:

\[
(27) \quad \text{?Nine of the ten marbles are in the bag. It is under the sofa.} \\
(28) \quad \text{One of the ten marbles isn’t in the bag. It is under the sofa.}
\]

According to the DRT account, the reason why the pronoun *it* cannot successfully be used to refer to the one marble not in the bag in (27) is because such an entity is only inferrable from information given by expressions in the previous sentence. No representation of any term denoting such an entity in (27) has been made available by the construction process projecting a discourse representation structure, on the basis of which the truth conditions of the previous sentence are compiled. So although in all models validating the truth of (27) there must be a marble not in the bag described, there cannot be a successful act of reference to such an individual, using the pronoun. By way of contrast, the anaphoric resolution is successful in (28), despite its being true in all the same models in which (27) is true, because
the term denoting the marble not in the bag is specifically introduced and hence represented in the DRS. Hence, it is argued, not only is the presence of an intermediate level of representation essential, but the concept of context from which expressions can pick up their interpretation must be representational in kind also.

4.2 Ellipsis

Even outside DRT debates, semanticists continue to argue over whether there is definitive justification of the representational nature of content and context (see [von Heusinger and Egli, 2000; Ginzburg and Cooper, 2004]). In particular, ellipsis provides a case of a particularly rich and structured concept of context-dependence, as we have already hinted in section 2; and there the dynamically evolving nature of both content and context is amply demonstrated. Indeed, ellipsis arguably provides a window on context, relying as it does on information that is in some sense manifestly available in the immediate context within which the fragment is produced/interpreted.

Consider again (8), repeated here:

(8) John was over-wrought about his results though Bill wasn’t.

As we saw, ellipsis is like anaphora in allowing multiple interpretations, determined by what information is accessed from context and how. This means that there is no reason to anticipate an algorithmic relation between an ellipsis site and assignable interpretation, nor between the fragment expression at that site and the assigned interpretation. So interpretation is not algorithmically determinable at either the level of the fragment itself or at the level of the propositional construal derived from it. Without some formal characterisation of how the context and content of sequences of utterances can evolve in tandem, one possible response is that the best that can be done is to analyze the relation between ellipsis site and the string from which its interpretation is built up, if there is one, as a grammar-internal phenomenon, leaving the nondeterminism of contextual factors aside. In principle, indeed, this has been the prevailing methodology. Until research into formal dialogue was developed (as, recently, by [Ginzburg, forthcoming; Ginzburg and Cooper, 2004; Purver, 2004; Cann et al., 2005]), the only forms of ellipsis that were addressed were indeed those where the ellipsis site can in some sense be analyzed sententially — either as the second conjunct of a compound conjunctive form or as an answer to a question. Relative to this restriction, the grammar-internal dispute has been whether a model-theoretic alternative is available for all such instances of ellipsis, with that semantic characterisation defined on the surface form of the string. Such an account would be non-representationalist, in the sense of invoking no additional semantics-internal concept of representation.

For some cases in particular, a model-theoretic account of this kind has seemed competitive. The debate is thus analogous to that between DRT and Dynamic Predicate Logic, but in this case played out over predicate meanings. For example, following Dalrymple et al. [1991], ellipsis construal has been defined as an
operation of (higher-order) abstraction on the model-theoretical content defined for an immediately antecedent conjunct. The resulting lambda-abstracted predicate can then be applied to the newly provided subject of the elliptical second conjunct.

By this means, as is minimally needed, a semantic explanation is available for cases such as (8) where, for a single antecedent form and assigned interpretation, ambiguity nonetheless arises. The process of abstraction over the content provided by the antecedent (‘John is overwrought about John’s results’) yields two types of abstract: abstraction over solely the subject of the antecedent, as in (29), or abstracting over the subject and all other references to the individual it denotes, as in (30):

\[(29) \quad \lambda x. x \text{ overwrought about Bill's results} \]
\[(30) \quad \lambda x. x \text{ overwrought about } x \text{'s results} \]

The two resulting predicates apply to the lambda term assigned to the elliptical fragment to yield the so-called strict/sloppy ambiguity. Hence, it is claimed, the semantic basis for explanation does not require any intermediate DRT or other representational construct, contrary to the conclusion informally evidenced in section 1.\(^8\) Indeed, on this view, there is no parallelism with anaphora: the process is simply an operation on the denotational content of one conjunct to provide a fully specified construal for the second, elliptical conjunct.

Yet there is competing evidence that sensitivity to structure is essential to the way in which elliptical fragments are reconstructed. In particular, the very restrictions taken to be diagnostic of syntactic phenomena constrain some cases of ellipsis. So-called antecedent-contained ellipsis displays sensitivity to the presence of a relative clause boundary, as expressed in the complex NP constraint of Ross [1965], for no such boundary can intervene between the relative pronoun of the containing structure and the ellipsis site [Fiengo and May, 1994; Merchant, forthcoming]:

\[(31) \quad \text{John interviewed every student who Bill already had.} \]
\[(32) \quad \text{*John interviewed every student who Bill ignored the teacher who already had.} \]

This is exactly as though the structure in question were present, not elided:

\[(33) \quad \text{John interviewed every student who Bill already had interviewed.} \]
\[(34) \quad \text{*John interviewed every student who Bill ignored the teacher who already had interviewed.} \]

\(^8\) The identification of the subject of predication is stipulated within the semantic vocabulary, and the restriction to only two such bases for abstraction is controversial.
On the syntactic account that such data are said to indicate, such structure is indeed present at some level of abstraction (only to be deleted through some low-level “PF Deletion” operation). Hence, so the argument goes, at least some types of ellipsis require syntactic explanation, involving full projection of clausal structure at the ellipsis site, with subsequent deletion of phonological material — hence an ineliminably representationalist account for at least some types of ellipsis. This brings its own problems, the first of which is that the VP apparently provided as the antecedent of the ellipsis site contains that site, threatening circularity. This immediately requires increased complexity in the syntactic analysis. A more general problem, however, is that once ellipsis is recast in structural terms, with full determinism of structure underpinning some required interpretation, the postulation of multiple ambiguities becomes inevitable. For example, the existence of sloppy and strict interpretations of ellipsis, as in (8), not only imposes distinct analyses of the ellipsis site, but it forces the use of different characterisations of coreference within the antecedent conjunct from which the construal is built up: one in terms of a binding mechanism, the other as free indexing, yielding the postulation of ambiguity in the source string even on a single interpretation. The matching of structure at ellipsis site and antecedent then forces the need to posit many different types of ellipsis [Merchant, forthcoming].

We thus appear to have arrived at the division of ellipsis into semantic types and an array of syntactic types, all supposedly displaying sentential structure/content even though in fact realised in a fragmentary form, with the parallelism of anaphora and ellipsis construal totally lost. Over and above this, there are yet further cases where, arguably, there is no determinate linguistic basis for assigning interpretation to the fragment [Stainton, 2006]:

(35) A: [coming out of lift] McWhirter’s?
   B: Second left.

Stainton argues that, contrary to both syntactic and semantic analyses, such cases do not allow any analysis as sentential reconstructions but have to be seen as a speech act that is performed without recourse to a sentential structure.9 These cases are redolent of the very broad use of fragments in dialogue (section 1), where a rich array of aspects of context may determine the way the fragment is understood:

(36) Father: We went to parents’ evening. The teacher
    Child: Mrs Smith?
    Father: No, your other teacher. Mr Jones. He said you were doing fine.

These fragmentary cases pose problems not merely for grammar-internal syntactic accounts of ellipsis, which effectively presuppose a single non-split utterance phenomenon, but also for the denotational account of ellipsis. On that analysis, the

---

9Nevertheless, Stainton’s argument also rests on the assumption that ellipsis mainly requires a sentential basis: his argument for the distinctiveness of what he calls pragmatic ellipsis is an argument to the effect that only these fragments have no such sententially based construal.
interpretation of the fragment would be expected to follow the general rule that its denotational content is provided by some lambda-abstract constructed from the content of the preceding utterance. This would require that the fragment *Mrs Smith?* in (36) must be interpreted as asking whether Mrs Smith went to parents’ evening; yet this is not even a possible interpretation. The clarification requested has to be about the identity of the teacher just introduced into the conversation. This illustrates the special — and, for conventional approaches to ellipsis, highly problematic — properties of cases where the fragment occurs very early in the interpretation process that is under way: there is an unexpected and unexplained asymmetry between early and late use of fragments relative to the context from which they build up their interpretations. Worse than this, whatever the fragment provides may have to eliminate what has just been processed in context, violating monotonicity, rather than constituting a re-usage or development of that context. This is so even in cases where the interpretation is accumulative, and not a correction:

(37) Every doctor — John, Bill AND Mary — signed the certificate.

Again, such cases are problematic for denotational accounts of ellipsis. Denotationally, the extension of *every doctor* has to be eliminated before any interpretation can proceed, since the denotation of that phrase cannot be taken, simpliciter, to be equivalent to the set made up of John, Bill and Mary. Consequently, this kind of approach cannot capture the sense in which the addition of *John, Bill and Mary* is an elaboration on, or extension of, the original, quantified expression.

At this point, it might seem that there is an inevitable drift towards recognizing that ellipsis simply is a complex set of phenomena. Indeed, ellipsis has been characterised as displaying “fractal heterogeneity” [Ginzburg and Cooper, 2004], calling on arbitrary types of information across the grammar-internal spectrum for reconstructing the requisite construal. The challenge of providing a uniform account of ellipsis understanding is thus signal not met: all that is achieved is an itemisation of disjunctive sets of ambiguities, some of them involving representations ineliminably, others said to be reconstructable from denotational contents directly. This amounts to a counsel of despair, admitting that no explanation can be found of what it is about elliptical fragments that enables them to yield such diverse interpretations, and to do so relative to context. Nor is any basis provided for why there is such widespread parallelism between anaphoric and elliptical forms of construal. If we are to find a solution to the puzzle of ellipsis in a way that reflects this parallelism, ellipsis has to be taken as illustrative of the much broader challenge of how to articulate the basis for context-dependence of natural language construal within formal linguistic theory.
5 A DYNAMIC SOLUTION: FROM REPRESENTATION TO CONSTRUCTION

To address this challenge, we take up the underlying methodology of DRT and pursue it further. To achieve a fully integrated account of anaphora and ellipsis resolution, we seek to incorporate the dynamics of how interpretation is built up relative to context, while articulating a concept of representation of content that is structurally sufficiently rich to allow the expression of structural constraints as required. We contend that the overarching problem lies in the assumption that representations are statically and independently defined. Another approach however is conceivable, one that is equally representationalist, but adds another dimension — that of update from one representation to another, hence a concept of construction of representations. The development of DRT, we suggest, pointed in the right direction with its modelling of anaphora resolution as reflected in the dynamics of how users build up interpretations in context; and to capture ellipsis, we generalise this to all structural dependencies, whether sentence-internal or supra-sentential.

The problem for DRT was the limit it imposed on the remit of its account. This reached a ceiling with forms of anaphora that are subject to syntactic explication. As these were, as sub-sentential structural dependencies, these were to be subject to grammar-internal characterisation, hence falling outside the DRT account. With ellipsis, the very same problem is replicated: the data are split into grammar-external and grammar-internal characterisations and then the latter into syntactic and semantic characterisations. In order to cross this hurdle, we need an account that is not restricted to the remit of the sentence-boundary, not restricted to arbitrary sub-types of structural dependency, and is essentially representational in its perspective on meaning specifications.

One way of extending the DRT form of analysis promises to remove these obstacles, thereby making possible an integrated account of context-dependent phenomena. This is to define a concept of time-linear incrementality within the grammar formalism itself, reflecting that of processing, so that both subsentential and supra-sentential dependencies are expressed as a process of update. The core syntactic notion becomes incremental growth of semantic representation and the concept of structure intrinsic to the string is replaced by the dynamics of building a progressively enriched semantic representation from a linear sequence of words, relative to a context that preserves a record of this growth process. With this shift, as we shall see, there is no intrinsic syntax-semantics divergence, no arbitrarily distinguished types of representation separating the two, and a structurally rich concept of context through which we can express an integrated characterisation of ellipsis. One framework, in particular, provides the requisite formal tools to effect this perspectival and conceptual shift: Dynamic Syntax [Cann et al., 2005; Kempson et al., 2001; Gargett et al., 2009; Kempson and Kiaer, 2010].
5.1 Dynamic Syntax

Dynamic Syntax (DS), like conventional syntactic frameworks, provides a theory of language form and its interpretation. Unlike in conventional frameworks, however, this theory of form emerges from a model of interpretation, and more particularly from the growth of interpretation. DS is avowedly representationalist, in the sense that it depends upon explicit representations of both semantic structure (prior to model-theoretic interpretation) and the ways in which this structure is progressively induced from the incremental processing of strings of words.

In this model, interpretations are represented as binary tree-structures of functor-argument form, and these are built up relative to context. Individual steps in this building process reflect the incrementality with which hearers (and speakers) progressively build up interpretations for strings, using information from context as it becomes available. It is like DRT in spirit, in that local predicate-argument structures are induced in a way that reflects the time line of processing, and such structured representations are taken to be essential to the expression of anaphora and ellipsis resolution. But it goes considerably further than DRT in a number of ways.

First, the mechanisms for building up such structures are presumed to apply in a strictly incremental way. Following the dynamics of on-line processing, representations of meaning are built up on a word by word basis — as opposed to the sentence by sentence approach of DRT. Second, this process of building up structure is taken to be what constitutes the syntactic component of the grammar: with the dynamics of structural growth built into the core grammar formalism, natural-language syntax is a set of principles for articulating the growth of structured semantic representations. Syntactic mechanisms are thus procedures that define how parts of interpretation-trees can be incrementally introduced and updated; they are therefore causally related to, but not constitutive of, the representations themselves. Third, reflecting the systemic context-dependence of natural language construal, all procedures for structural growth are defined relative to context; and context is defined to be just as structural and just as dynamic as the concept of content with which it is twinned. Context, by definition, constitutes a record not merely of the (partial) structures built up, with the typed formulae that decorate them, but also the procedures used in constructing them [Cann et al., 2007]. In short, the general methodology is a representationalist stance vis-à-vis natural-language construal [Fodor, 1983], with the further assumption that concepts of underspecification and update should be extended from semantics into syntax. But the bonus of such explicit adoption of representationalist assumptions with respect to content is the avoidance of any further levels or types of representation — a clear application of Occam’s razor.

The tree logic and tree-growth processes

The general process of parsing is taken to involve building as output a tree whose nodes reflect the content of some uttered formula — in the simple case of a sentence
uttered in isolation, this is a complete propositional formula. The input to this task, in such a simple case, is a tree that does nothing more than state at the root node the goal of the interpretation process to be achieved, namely, to establish some propositional formula.

For example, in the parse of the string *John upset Mary*, the output tree to the right of the $\mapsto$ in (38) constitutes some final end result: it is a tree in which the propositional formula itself annotates the top node, and its various subterms appear on the dominated nodes in that tree, rather like a proof tree in which all the nodes are labelled with a formula and a type. The input to that process is an initial one-node tree (as in the tree representation to the left of the $\mapsto$ in (38)) which simply states the goal as the requirement (shown by $?Ty(t)$) to establish some content, which is to be a formula of appropriate propositional type $t$ (there is also a pointer, $\diamond$, which indicates the part of the tree that is under development at any given time).\(^{10}\)

(38) John upset Mary.

\[
?t, \diamond \mapsto (\text{Upset}'(\text{Mary}'))(\text{John}') : t, \diamond
\]

\[
\begin{align*}
\text{John}' & : e \\
\text{Upset}'(\text{Mary}') & : e \rightarrow t \\
\text{Mary}' & : e \\
\text{Upset}' & : e \rightarrow (e \rightarrow t)
\end{align*}
\]

Parsing *John upset Mary*

The parsing task, using both lexical input and information from context, is to progressively enrich the input tree to yield an output of appropriate type using general tree-growth actions and the sequence of words of the string. In order to talk explicitly about how such structures grow, trees need to be defined as formal objects; and DS adopts a (modal) logic of finite trees (LOFT: [Blackburn and Meyer-Viol, 1994]).\(^{11}\) DS trees are binary with the argument always appearing by convention on the left and the functor on the right. These are defined over

\(^{10}\)This exegesis omits all indication of tense construal and quantification. In brief, the language of the formula representations is that of the epsilon calculus, with all quantified terms of type $\langle e \rangle$, matching the arbitrary names of predicate logic natural-deduction (of which the epsilon calculus is the formal study). The essential property of arbitrary names is that their structural representation is simple, but their semantics complex, by definition reflecting their containing environment. The advantage of this, in the linguistic application, is that this makes available a natural basis for name growth, following the general pattern of the DS framework, so that initial underspecification of a name under construction and its subsequent extension is naturally expressible. Tense construal then projects an epsilon event-term, with tense, adjuncts and aspect all taken to add to the restrictor specification of such terms, again a concept of term extension (see [Cann, forthcoming]).

\(^{11}\)There are two basic modalities, ways of describing node relations: $\langle \downarrow \rangle$ and $\langle \uparrow \rangle$. $\langle \downarrow \rangle \alpha$ holds at a node if $\alpha$ holds at its daughter, and the inverse, $\langle \uparrow \rangle \alpha$, holds at a node if $\alpha$ holds at its mother.
mother and daughter relations, indicating a possible sequence of mother relations, or a possible sequence of daughter relations. The LOFT language makes available a basis for structural underspecification, using Kleene star (*) operators, making concepts of dominate and be dominated by expressible for a tree relation even before the fixed number of such mother or daughter relations is fixed. For example, ⟨↑∗⟩Tn(a) is a decoration on a node indicating that somewhere dominating it is the node Tn(a). All that is determined is that the node in question must always be dominated by the Tn(a) in any future developments of the tree.

A corollary of structural underspecification, and another essential feature of the model, is the existence of requirements for update. This is central to the ability to reflect the time-linearity involved in building up trees in stages (i.e. through partial trees). For every node, in every tree, all aspects of underspecification are twinned with a concept of requirement, represented as ?X for any annotation X on a node. These are constraints on how the subsequent parsing steps must progress. Such requirements apply to all types of decoration, so that there may be type requirements, e.g. ?Ty(t) (or ?t), ?Ty(e) (or ?e), ?Ty(e → t) (or ?e → t); treenode requirements, e.g. ?∃xTn(x) (associated with underspecified tree-relations); and formula requirements, e.g. ?∃xFo(x)(associated with pronouns and other anaphoric expressions). These requirements drive the subsequent tree-construction process, because unless they are eventually satisfied the parse will be unsuccessful.

Such structural underspecification and update can then be used to define core syntactic notions in a way that follows insights from parsing, and the time-linear dimension of processing in real time. For example, they notably lend themselves to analysis of the long-distance dependency effects which, since the late 1960’s, have been taken by most to be diagnostic of a syntactic component independent of semantics. When first processing the word Mary in (39) below, it is construed as providing a term whose role isn’t yet identified. The parse is then taken to involve the application of a computational action which introduces a structural relation to the topnode (the initial root node decorated with ?t) which is underspecified at this juncture: it is identifiable solely as being dominated by the topnode, and requiring type ⟨e⟩, i.e. bearing the requirement ?e:

(39) Mary, John upset.

The expression Mary is thus taken to decorate an as-yet unfixed node: this is step (i) of (40). Accompanying the underspecified tree relation is a requirement for a fixed treenode position: ?∃x.Tn(x). All of this technical apparatus provides a formal reflection of the intuitive sense in which a string-initial expression of this kind finds its proper role in relation to form and meaning only after other aspects of the overall structure have been put in place.

The update to the relatively weak tree-relation in (40, (i)) becomes possible only after processing the subject and verb, which jointly yield the two-place predicate

12This is a standard tree-theoretic characterisation of dominate, used in LFG to express functional uncertainty; see [Kaplan and Maxwell, 1988]
structure as in step (ii) of (40). The simultaneous provision of a formula decoration for this node and update of the unfixed node is provided in the unification step indicated there, an action which satisfies the update requirements of both nodes to be unified, leaving no trace in the output of the word Mary having been string-initial.

(40) Parsing Mary, John upset:

\[
\begin{align*}
?t, Tn(0) \\
&\downarrow \\
&?x, Tn(x), \\
&\langle\uparrow \ast\rangle Tn(0), \\
&\triangleright step (i)
\end{align*}
\]

\[
\begin{align*}
&\downarrow \\
&?t, Tn(0) \\
&\downarrow \\
&Mary' : e, \\
&\langle\uparrow \ast\rangle Tn(0), \\
&\triangleright step (i)
\end{align*}
\]

\[
\begin{align*}
&\downarrow \\
&John' : e, \\
&\langle\uparrow \ast\rangle Tn(0), \\
&\triangleright step (ii)
\end{align*}
\]

\[
\begin{align*}
&\downarrow \\
&?e, \triangleright Upset' : e \rightarrow (e \rightarrow t)
\end{align*}
\]

This process feeds into the ongoing development in which, once all terminal nodes are decorated, bottom-up modalised application of labelled type deduction leads to the creation of the completed tree indicated in (38). Note that this is the same intuition that lies behind the syntactician’s notion of ‘displacement’, but captured without resort to abstract notions of movement or other syntax-specific mechanisms.

Such an account of structural underspecification and update is indeed not contentious as a parsing strategy; what is innovatory is its application within the grammar-mechanism to provide the central core of syntactic generalisations and the characterisation of wellformedness. Discontinuity effects can now be seen on a par with anaphora construal, the latter an underspecification of interpretation to be resolved as part of the interpretation process, the former an underspecification of structure, equally to be resolved as part of that process. In the case of discontinuity, this construction of partial representation is a generally available option for tree growth. In the case of anaphora, it is the lexical projection of a place-holding formula value along with a requirement for its update which induces an element of underspecification, with both types of underspecification requiring update within the construction process.

This account might seem in principle to be skewed by focussing on parsing, but this is only superficial. Production follows the very same processes, with but one further assumption: that at every step in production, there must be some richer tree, a so-called goal tree, which the tree under construction must subsume, in the sense of being able to be developed into that goal tree, according to the system defined. So parsers and producers alike use strategies for building up representations of content, either to establish an interpretation for a sequence of words, or to find words which match the content to be conveyed. Both of these activities are alike also in being context-dependent, so that structural and content choices may be determined on the basis of what has just been processed.
To achieve the basis for characterising the full array of compound structures displayed in natural language, DS defines in addition the license to build paired trees, so-called linked trees, linked together solely by the sharing of terms, such as may be established by encoded anaphoric devices like relative pronouns. Consider the structure derived by processing the string John, who smokes, left:

\[(41) \text{Result of parsing John who smokes left:}\]

\[
\begin{array}{c}
Tn(0), Leave'(John') \wedge Smoke'(John') : t \\
\downarrow \\
Tn(n), John' : e \\
\downarrow \\
\langle L^{-1}\rangle Tn(n), Smoke'(John') : t, ?(\downarrow) John'
\end{array}
\]

The arrow linking the two trees depicts the link relation. The tree whose node is pointed to by the arrow is the linked tree (read \(\langle L^{-1}\rangle\) as ‘linked to’). Such linked trees may be conceived of as subproofs over a term shared with the ‘host’ tree, whose content must be established in order to establish some property of that term, which is expressed by the overall proposition. In the above, non-restrictive case, the content of the linked tree merely adds the information that John smokes to the information that he left. But such structures may also provide restrictions on the shared term, as in restrictive relative clauses, or constrain the context within which some term is to be construed, as in the so-called Hanging Topic Left Dislocation structures [Cann et al., 2005]:

\[(42) \text{As for John, I dislike his style of painting.}\]

Within any one such linked tree, the full range of computational, lexical and pragmatic actions in principle remain available, depending solely on the type requirements relative to which the pairs of linked structures are developed.

With this flexibility to allow the incremental projection of arbitrarily rich compound structures, the result is a formal system combining lexical, structural and semantic specifications, all as constraints on the growth of trees. As argued in [Kempson et al., 2001; Cann et al., 2005; Kempson and Kiaer, 2010], this leads to the comprehensive DS claim that the syntax of natural languages does not involve a separate level of representation besides what is needed for semantics (Cann et al. [2005], and elsewhere), not because there is no level of semantic representation, but because there is no independent level of syntactic representation.\(^{13}\)

\(^{13}\)Analogous arguments apply to morphological structure [Chatzikyriakidis and Kempson, to appear, 2011] but we do not pursue these here.
So, despite the assumption that this progressive build-up of a semantic representation is a basis for doing syntax, syntax in this model is not taken to include a level of representation where there is structure defined over a string of words. DS trees are not inhabited by words and there is no notion of linear ordering expressed on the tree; the annotations on the tree are solely representations of conceptual content. Lexical specifications are defined in exactly the same terms of tree growth, as constraints on tree growth. Such tree growth actions can take place only if the condition triggering these actions matches the decorations on the node which the pointer has reached in the parse — this is a major determinant of word order effects.

A consequence of this methodology is the way concepts of structural underspecification and subsequent update replace the need to postulate multiple types of representation. Through the building and updating of unfixed nodes, a multi-level account of syntax is replaced with progressive growth of a single representational level; and this level turns out to be nothing more than the representation of content, as established from processing the linguistic string in context. The characterisation of lexical specifications in the same terms enables seamless integration of lexical and syntactic forms of generalisation, so that discrete vocabularies for lexical, syntactic or semantic generalisation are unnecessary (and, indeed, precluded).

Constraints taken to be specific to natural-language syntax and not reducible to semantic generalisations are analysed as constraints on the same growth process. For example, the complex NP constraint, which precludes the dependency of an expression outside a relative clause sequence with some site within that relative, is analysed in DS via the the licence to build linked-tree pairings. This imposes its own locality restrictions, in terms of limits on the direction of further tree growth. Any expression characterised as decorating an unfixed node, e.g. a relative pronoun, has to be resolved within the tree which that unfixed node construction step initiates. Hence it cannot be resolved in a tree that is merely linked to that tree. Thus, the island constraint is captured, not in terms of notions of subjacency (however realised) that are defined over putative hierarchical structures over strings of words, but in terms of the discreteness of subproofs within an overall proof of the content expressed by such a string.

5.2 Ellipsis and context

With this sketch of the DS framework, we can now return to ellipsis, and see how a multiplicity of unrelated ellipsis types can be avoided within a system if it articulates a dynamic account of content accumulation. Recall the central problem regarding ellipsis: model-theoretic accounts are too weak to handle syntactic constraints, while syntactic accounts, required to feed interpretation rather than interact with it, freely posit ambiguity. In DS, though, syntax is expressed as growth of representations of propositional content relative to context. Within

\footnote{A relative pronoun in English is lexically defined to induce a copy of its antecedent ‘head’ at an unfixed node.}
such a system, VP ellipsis construal and pronoun construal work in essentially the same way. Both project a place-holder for some value with its requisite type specification (provided by the lexical specification of the pronoun or auxiliary), for which the formula value is taken to be provided from the context [Purver et al., 2006].

Here, one’s notion of context is clearly crucial. In DS, context is an evolving record of representations of meaning plus the process of their building (strictly, a sequence of triples: a decorated structure, a sequence of words and the update actions used in establishing the given structure over that word sequence [Cann et al., 2007]). Given this notion of context, any aspect of it is expected to be re-usable as a basis for the construal of ellipsis, and this encompasses all that is required to account for the various kinds of ellipsis.

First there is the availability of meaning annotations from some context tree, re-using a formula just established by a simple substitution process. This direct re-use of a formula from context is illustrated by the strict readings of VP-ellipsis, where the meaning assigned to the ellipsis site matches that assigned to the antecedent predicate (see section 1). In the sloppy readings, where there is parallelism of mode of construal but not matching of resultant interpretation, it is the structure-building actions that are replicated and applied to the newly introduced subject. (43) provides such a case:

(43) A: Who hurt himself?
   B: John did.

Processing the question in (43) involves the construction of a two-place predicate, as indicated by the verb, plus the construction of an object argument; and then, because this object contains a reflexive pronoun, it is obligatorily identified with the argument provided as subject. Re-applying these very same actions in the new tree for B’s reply, gives rise to a re-binding of the object argument to john, which already decorates the subject node of the new tree, thanks to the elliptical fragment. The effect achieved is the same as the higher-order unification account but without invoking any mechanism beyond what has already been used for the processing of the previous linguistic input. All that has to be assumed is that the meta-variable contributed by the anaphoric did can be updated by some suitable selection of a sequence of actions taken from the context. This license to re-use actions stored in context is equally made use of in anaphora construal, giving rise to the so-called “lazy” use of pronouns (see section 1).

Finally — and now falling within just the same general mode of explanation — there are those cases from dialogue where what the context provides is structure, to which the words of the follow-on speaker provide an extension. Canonical cases of this are question-answer pairs, the answer providing the update to the very structure provided by the question.¹⁵

¹⁵This is on the assumption that wh expressions project a particular form of metavariable [Kempson et al., 2001].
A: Who did John upset?
B: Himself.

But this pattern is characteristic of dialogue: quite generally, as we saw in section 1, one speaker can provide words which induce some structured representation of meaning, often one that is in some sense incomplete, to which their interlocutor can provide an extension. Here the modelling of production in DS, using the very same procedures as in parsing, comes into its own. This predicts the naturalness of split utterance phenomena, of which question and answer are a subtype, since both the speaker and the hearer in an exchange are presumed to be building comparable structures, continuing to use the very same procedures whether in production or in parsing. So when an interlocutor switches roles, they continue with the very same structure which they have just been building in the previous role. The immediate consequence of this is that tight coordination between the parties is expected, as is the availability of a procedure of setting up the first part of some structural dependency within one speaker/hearer role which is subsequently resolved in the other role. Thus, ellipsis construal can take aspects from immediate context, whether these be representations of meaning, or the actions used to build such representations, or indeed the partial structures that formed that context. The breadth of effects achieved in ellipsis construal need not be stipulated; it is grounded in the richness of the dynamic, structural concept of context.

Perhaps the most significant part of this rich attribution of structure to context, in relation to issues of representationalism, concerns the interaction of structural constraints with the general process of building interpretations in an evolving, structured context. For example, the supposed island constraints displayed in antecedent-contained ellipsis are naturally captured without any syntactic level of representation. Recall that the complex NP constraint is said to underlie the contrast between (31) and (32), repeated here, in which the construal of the ellipsis site precludes interpretation across any additional relative clause boundary:

(31) John interviewed every student who Bill already had.

(32) *John interviewed every student who Bill ignored the teacher who already had.

The crucial point here is the nature of the relative pronoun, which initiates the construal of the expression containing the ellipsis site. From a processing point of view, an English relative pronoun intuitively does two things: it initiates some type of sub-structure and promises a term within that sub-structure that is co-referent with the term that the relative clause sequence is taken to modify. The implementation of this in DS is that the relativiser triggers both the construction of a linked tree and the presence of an initially unfixed node within that linked tree, to which it adds a copy of the term that the relative clause modifies (informally, the ‘head’ of the NP+relative sequence). It then follows from general properties of the structural dominance relations defined in DS that an unfixed node must ultimately be fixed locally within that structure; in particular, it cannot ‘look beyond’ any
additional link relation for a place to fix itself [Cann et al., 2005]. The effect that (in more conventional vocabulary) a relativiser cannot be coindexed with a trace across another relative clause boundary thus follows from the parsing mechanism. The fact that such constraints are apparently non-semantic therefore need not be taken as evidence for a distinct syntactic level of representation: in this new perspective they are locality constraints on tree growth. Accordingly, examples like (32) do not, on a dynamic view, preclude a unified account of ellipsis, even though on other accounts they are taken to be diagnostic of forms of ellipsis requiring syntactic analysis, while others require quite distinct semantic analysis.

Overall, then, the DS perspective indicates the potential to meet the threefold challenge posed by ellipsis: capturing the very diverse forms of interpretation, providing a unitary base from which this diversity can be obtained, and enabling the articulation of a concept of context that is rich enough to make an integrated account of ellipsis possible. We have developed the case of ellipsis here in preference to that of anaphora, not only because it brings out the issues of representationalism more strikingly, but also because of the extensive parallelisms between anaphora and ellipsis: both are taken to involve inputs of meaning which are lexically defined as being underspecified with respect to an output meaning, and rely on context to provide the necessary update. This is made possible precisely because the core concept of the natural-language grammar formalism is that of growth of representations of content; and all explanations are accordingly expressed in these terms. The diversity of structure-sensitive interpretation mechanisms which have on the conventional syntactic account of ellipsis to be expressed as independent structures, hence as stipulations of ambiguity of the strings themselves, can be seen as the application of processes to some underspecified input, with different procedures for growth from that input giving rise to the divergent construal. So the shift into the dynamic perspective is essential to the resulting integrated account.

While here we have pursued the significance of anaphora and ellipsis specifically, we should not lose sight of the endemic nature of context dependence in natural language, which extends far beyond these two particular phenomena. Indeed, context dependence is now recognised as permeating the whole of the lexicon (and not just some quasi-grammatical subset of lexical items such as pronouns and auxiliary verbs). In recent years, important work has begun on the general phenomenon of lexical meaning and the extreme flexibility of construal which content words in context seem to allow. In particular work being developed within Type Theory with Records [Cooper, 2005; this volume; Larsson, 2008], building on Martin-Löf’s type theory, provides a general framework for natural language interpretation which, like DS, takes a basic semantic vocabulary and gives it a proof-theoretic twist, so it is model-theoretically grounded but makes essential use of proof-theoretic dynamics (see [Cooper, this volume] for detailed exegesis). Notwithstanding the proof-theoretic underpinnings of the framework, one important application has

16We set aside here cataphoric effects, but these can be handled without stipulation in the framework, using appropriate delay mechanisms. Initial lexically determined specifications of type but not content allow for later specification of the content, as with expletive pronouns.
been to provide a formal articulation of what it is about a word that enables it both to be the bearer of an identifiable concept of meaning (in some sense) but nevertheless to model the full range of variation in truth conditional contents which the word makes expressible.

We now find ourselves in a situation that turns on its head the general view of linguistic representation from the later twentieth century. The position developed in the early nineteen-seventies onwards was one of indispensable representations of syntactic structure defined over strings of words, over which semantic interpretations can be directly stated. From this position, we have reached one in which syntactic representations are replaced by processes that articulate semantic representations — and these are, in their turn, conceived of as being the only necessary level of representation. So it is syntax, traditionally conceived, that has become superfluous, in the sense of not requiring any special vocabulary other than that of inducing the growth of representations of meaning — a view which is notably close to the ontology of categorial grammar [Morrill, 1994], but avoids the Montagovian limitations of the latter with respect to issues of context dependence in natural language. Thus, we are reaching the ability to express directly the folk intuition that in understanding utterances of language, it is representations of content that are being built up in context, and that language is a vehicle for recording or expressing our thoughts.

6 IMPLICATIONS FOR DYNAMIC PERSPECTIVES

6.1 Compositionality

With this move away from direct mappings of natural language strings onto denotational contents, in much the same spirit as DRT, it is important to address potential objections in the same style as those levelled at DRT by Groenendijk and Stokhof [1991]: that the account presented violates the compositionality of meaning for natural languages (see also [Dekker, 2000], and the response by Kamp [1996]).

Compositionality of content has indeed been very generally presumed to be sacrosanct as a working methodology for formal specifications of natural-language semantics. However, consider again the standard construal of compositionality: ‘The meaning of an expression is a function of its component parts and the way they are put together’. Far from being a tight restriction on natural languages, this general form of compositionality is (as has intermittently been noted) extremely

[17]It might be argued that this position can be attributed to the minimalist conception of transformational grammar where LF (Logical Form) is the only level of representation [Chomsky, 1995]. The difference between this approach and that of Dynamic Syntax is that minimalism retains its structuralist foundations by defining LF to be inhabited by words, categories and their hierarchical combinations, rather than by concepts and their proof theoretic ones. As we have argued, an inability to accommodate the context dependence of natural language phenomena ultimately follows from these foundations.
weak, requiring at least a specific definition of ‘be a function of’, since functions may delete or arbitrarily re-order symbols or other object language objects.

In essence, the principle of compositionality is an attempt to capture the intuitive idea that meanings are not computed arbitrarily on different occasions, but are constrained to the meanings of the basic expressions of a language (words or morphemes, in natural languages) and some known and determinate means of combining them to construct propositional meanings. A minimal assumption (often unexpressed, but see [Cann, 1993]) is that no mapping from syntax to semantics may delete already established information, so that any compositional function should be monotonic. Additionally, there is typically an assumption that each word contributes something to the meaning of the expression in which it appears. But this notion of ‘contribute’ is also in need some interpretation, given the existence of fully grammaticalised expressions like the complementizer that in English (which at most operates as an identity function over the proposition expressed by its associated clause) and pleonastic expressions like it in weather verb constructions or expletive uses (like that in (17), above). Arguably, these contribute nothing at all to the interpretation of a string; at most they can be claimed to contribute to some form of ‘constructional meaning’, and this is likely to be of a sort that does not fall under the purview of conventional denotational semantics. In general, therefore, the precise interpretation of compositionality depends on the theory in which it is taken to apply and to prior assumptions made about the nature of the relations between words, syntax and expressible meanings.

Furthermore, the familiar rule-by-rule concept of compositionality threatens to preclude any characterisation of the systemic potential of all natural language expressions for context-dependent construal. Yet this is arguably the core property of the expressivity of natural languages and as such should be diagnostic of successful characterisations of natural language content. It follows that the common understanding of compositionality has to be modified, if it is to be sustainable as part of a properly explanatory account of natural language interpretation.

In the face of this challenge, we suggest that discussions of compositionality of content for natural language strings have conflated two concepts. There is, on the one hand, the essential contribution to be made by each word to the business of interpretation. This, we have argued, should be conceptualised as a contribution to a process: the progressive construction of a structural representation from a sequence of words. On the other hand, there is the compositionality of the content of each structure that results from that process. In teasing these two notions apart, we have two relatively simple concepts of compositionality, one defined in terms of monotonic incrementality, but lacking any notion of content; the other defined in terms of compositionality of content, but making no reference to the contributions of individual words. The first of these is defined in [Kempson et

---

18Cf. Fodor’s [2001] position, in which the representation of meaning (the Language of Thought) is similarly claimed to be the locus of compositionality, rather than natural languages themselves displaying compositionality. Fodor remains inexplicit on the matter of just how words do make systematic contributions to representations of meaning.
Ronnie Cann, Ruth Kempson, and Daniel Wedgwood

al., 2001, chapters 2-3, 8] as monotonic growth over partial trees. The second requires a concept of semantic content defined for whichever logical system is used to represent the structure of the Language of Thought, taken in DS to be a typed version of the epsilon calculus.\textsuperscript{19}

In any case, it is important to bear in mind that the criticism of DRT with respect to compositionality primarily concentrated on the introduction of a level of representation that is intermediate between the structure inhabited by strings and the level of assignable denotational content, thus apparently flouting Occam’s Razor. On the DS perspective, this criticism is simply deflected, since DS posits but a single level of representation: that of meaning which is, in principle, amenable to model-theoretic interpretation.

What is abandoned in this move is the assumption of there being independent syntactic representations: these are replaced by systematic incremental processes which are defined for all wellformed strings of some language. The order of words as given must induce at least one monotonic process of tree growth to yield a complete, and compositionally defined, representation of content as output. Knowledge of language, then, resides in the systematic capacity to build up representations of thought from sequences of words, relative to the context within which such incremental processing takes place.

6.2 Knowing how and knowing that

The view of grammar that we have advocated in this article has implications for another fundamental assumption that underpins linguistic methodology. Conventional models of syntax and semantics assume that knowledge of language consists in ‘knowing that’ language is certain way, and not in any sense ‘knowing how’ to use language. A general distinction between knowing-that and knowing-how was introduced into the philosophy of mind by Ryle [1949]. Though Ryle himself advocated a certain kind of knowing-how view, the characterisation of knowledge in a knowing-that sense has overwhelmingly held sway in linguistics and in cognitive science more widely (it has also been dominant in philosophy, though see [Bengen and Moffett, 2007]\textsuperscript{20}).

\textsuperscript{19}See Meyer-Viol [1995] for discussion of formal properties of the epsilon calculus in relation to predicate logic.

\textsuperscript{20}From a philosophical perspective, Stanley and Williamson [2001] claim that the concept of knowledge-how is simply a species of knowledge-that — in both cases, a relation between an agent and a proposition. However, a significant part of the Stanley and Williamson case depends on parallel assumptions within linguistics of overwhelmingly sentence-based methodology, in which putatively nonsentential/non-truth-conditional phenomena are analysed wherever possible as sentential structures and truth-theoretic contents respectively. This leads to a near-circularity of reasoning: the conclusion that ‘knowing how’ phenomena are reducible to a sub-species of ‘knowing that’ is derived from syntactic and semantic methodologies which presuppose a strictly ‘knowing that’ approach, being themselves reflections of static, truth-based methodology. In addition, in order to explain their distinctiveness, they have to invoke sub-divisions of types of ‘knowing that’ based on pragmatic factors of very unclear provenance, for which no pragmatic theory as currently envisaged would provide substantiation.
In contrast to this, the DS conception of grammar, as a process of constructing representations of meaning, clearly constitutes a model of ‘knowing how’, in a certain sense. In some ways, this is a striking departure from long established assumptions in linguistic theory, and this difference is key to the potential of the framework to account both for the general properties of language as an inherently context-dependent system and for more specific phenomena such as ellipsis and anaphora. However, it is important to clarify the particular ways in which DS is a knowing-how approach, in order to pre-empt certain potential objections.

The view of grammar that we have advocated is a knowing-how view in the following sense: knowledge of language consists in having a set of mechanisms for language use, specifically a means of retrieving a meaning from a string of words (or their sounds) uttered in context (and, by the same token, a means of producing a string in context to convey some meaning), reflecting a ‘commonsense’ view of language (see also [Phillips, 1996; 2003]). Linguistic competence, then, is no more than the possession of mechanisms that make possible the doing of what one does in language performance.

This view ostensibly flies in the face of Chomsky’s articulation of the competence/performance distinction, which espouses a knowing-that concept of the capacity for language, and which is wedded to a purely static kind of representationalism, with no reflection of the dynamics of language processing. Yet the suggested shift does not amount to a collapse of a competence model into a performance model, with the loss of necessary scientific abstraction that this would imply. The strict concentration on knowledge of language that Chomsky advocated is not lost simply by reconceptualising the nature of this knowledge. The system as defined in DS articulates a set of constraints on tree growth, a set of principles proposed as underpinning processes of interpretation in real time. It does not bring the application of these principles into the domain of competence. There remains a sharp distinction between grammatical knowledge and independent, potentially intractable extra-grammatical factors, though the dynamic approach opens up new and explanatory possibilities for interaction between the two. In so doing, this approach makes available ways of articulating underspecification and update in natural language interpretation. Crucially, the DS formalism makes no attempt to model what determines the particular choices made by language users in resolving such indeterminacies on particular occasions; these matters do indeed belong to performance.

Therefore, the way in which DS constitutes a knowing-how approach does not entail the abandonment of a competence theory, in the sense of a theory that aims strictly at characterising knowledge of language. The notion of competence that we have criticised elsewhere in this article is an altogether more specific notion, encompassing a sentence-based methodology, a commitment to static representations of linguistic structure, and a self-imposed blindness to the pervasive context-dependence of natural languages. One of the lessons of re-assessing the abstract notion of representation in grammatical theory has been to show that a grammar, as a characterisation of knowledge of language, need not have these features.
Just as ‘knowing how’ in our sense must be distinguished from performance theory, so it must also be distanced from certain other possible associations, in particular any connection to behaviourism. Ryle’s original concept of ‘knowing how’ was grounded in terms of dispositions and this has led to the perception among some analysts that Ryle advocated some form of (quasi-)behaviourism. Whether or not this is a fair reading of Ryle (an issue that we leave to philosophy), it would be wrong to tar all kinds of knowing-how approach with the behaviourist brush. This is arguably what has happened in the development of modern linguistics, in which the adoption of a mentalist perspective on language is often equated, without argument, with a knowing-that conception of knowledge of language. For example, Chomsky [2000, 50–52] contrasts “the conception of language as a generative procedure that assigns structural descriptions to linguistic expressions”—an intrinsically knowing-that characterisation of linguistic knowledge — with the idea that knowledge of language can be reduced to an ‘ability’. No other possibility is entertained in this discussion, despite the fact that Chomsky actually notes that “knowing-how […] cannot be characterised in terms of abilities, dispositions, etc.” [2000, 50–52, italics in original]. Logically, this leaves open the possibility of a knowing-how model which characterises not dispositions, abilities and behaviours but the principles underlying them, which is suitably abstracted from the data and which in all other ways satisfies the need for a coherent and tractable account of linguistic knowledge. More generally, one may be committed to the study of language as an abstract system of knowledge but still question the best ways to conceive of and represent this knowledge. Formal modelling of the structural dynamics of language is as far from behaviourism as any more familiarly static representationalist model is.

Nevertheless, one thing that we have tried to convey is the fact that the choice between kinds of knowledge representation is a significant one. Different conceptions of linguistic knowledge are not mere ‘notational variants’ of one another, nor is the choice between them merely a matter of perspective. To the contrary, different approaches have very different empirical coverage, display different degrees of overall theoretical economy and contrast in how they relate to the reality of language as a system that is interpreted relative to context. As we have argued, it is not only possible to define a coherent representationalist system that gives the dynamics of structure a central role; it is also highly desirable from both empirical and conceptual points of view. Therefore, while we would distance ourselves from some aspects of Ryle’s characterisation of knowing how — and certainly from some associations it has gained over the years — we maintain that a knowing-how model, at least in the limited sense that we require, is still very much available for linguistic theory to pursue. There remain two different types of perspective for natural language analysis. One broad type is a characterisation of language as a declarative multi-level body of knowledge for which numbers of stipulations and imposed ambiguity postulations are required even to express the data. The other perspective is one in which a natural language is characterised as a set of mechanisms for proposition construction. Recalcitrant problems of the former
perspective promise to dissolve away in the latter.

We close with the observation that our arguments, for all their questioning of conventional methodologies, have led us to a position that is highly intuitive and has a long and respectable (even in some ways Fregean) heritage: that thoughts have denotational content and constitute the sole focus of semantic enquiry. Languages are used to express our thoughts, and these have the content they do through a semantics of resulting structures defined to yield denotational contents grounded in the primitive concepts of individual and truth. What is new is the explanation of how language performs this function. Language is neither identical to thought nor ‘mapped onto’ thought via a series of representations of uncertain status. It is instead intrinsically dynamic, a vehicle for the construction of objects over which inference is definable. As a mechanism, it is built to interact with the environment in which it is employed, but is not defined by it. A representationalist commitment, then, is ineliminable. Yet it is minimal, in invoking only the forms themselves and representations over which inferences are derivable from their use. In sum, language is a vehicle for constructing thoughts from the building blocks which our words provide. But, at one and the same time, given its reflection of real-time language use, it is a vehicle that allows interaction with others in the construction of such thoughts, hence a vehicle for the interactive and coordinated construction process that constitutes communication.

BIBLIOGRAPHY


Note that this does not in itself entail a particular concept of individual. Hence this might take the form of the individual constant beloved of Russell or some concept of type \( \langle e \rangle \) terms that are constructed for the purpose of inference-drawing to general conclusions, with arbitrary witnesses as their denotation (in the manner of [Sommers, 1982; Fine, 1986], and others).


Representationalism and Linguistic Knowledge


THE PHILOSOPHY OF PHONOLOGY

Philip Carr

INTRODUCTION: THE NATURE OF PHONOLOGICAL KNOWLEDGE

I will assume, in what follows, that phonology as a discipline has, as its object of inquiry, a form of knowledge. On that assumption, the central questions that arise in the philosophy of phonology are: what is its nature? Is it entirely mind-internal? Is it intersubjective in nature? Is it, somehow, both of those? How do we come to possess such knowledge? If phonological knowledge is mind-internal, is there innate phonological knowledge, present in the mind at birth? What form would such innate knowledge take? Can such putatively innate phonological knowledge be coherently subsumed under a Chomskyan version of innate linguistic knowledge? If there is innate phonological knowledge, how much of phonological knowledge is acquired, rather than innate? Why? Is phonological knowledge distinct in kind from syntactic, morphological and semantic knowledge? What assumptions might lead us to think that it cannot be distinct? Or, if it is distinct, how and why does it differ from other kinds of linguistic knowledge? These issues are complex; they are connected to issues in (among other things) child development, neurophysiology, and issues in phonological theorising concerning (1) the phonetics/phonology distinction, (2) the interpretation of the notion ‘the linguistic sign’, and the notion ‘groundedness’, (3) the acquisition of phonological knowledge and the factors involved in it, including social interaction, (4) the status of social conventions, unconscious knowledge and implicit learning, (5) competence, performance and usage-based approaches to phonological knowledge, and (6) internalist and externalist conceptions of phonological knowledge. While I will endeavour to deal with each of these in turn, we will see that they are inter-connected in complex ways; discussion of any of these topics requires discussion of the others.

1 THE PHONETICS/PHONOLOGY DISTINCTION

There is no consensus in the phonological literature as to whether it is possible to adopt a clear distinction (or indeed, any distinction) between phonetics and phonology, and among those who do adopt such a distinction, there is no clear shared sense in the phonological community of what the relationship between the two might be (see the contributions to [Burton-Roberts et al., 2000] for extended discussion of the issues from a variety of viewpoints). The relationship between
phonetic phenomena and phonological objects has been described in a variety of ways: instantiation, realisation, interpretation, implementation, representation, encoding, exponence, expression, transduction and even transmogrification. I now discuss some of these, and relate them to the question of the status of phonological objects (if such exist).

**Realisation, instantiation and the type/token distinction**

Talk of instantiation is, arguably, type/token talk: tokens are said to instantiate their types. A token of a type is an *instance* of that type. To begin with a simple example, assuming that there is such a thing as the type ‘sheep’, a token of that type is an instance of ‘sheep’; more simply, it is a sheep. But the type/token distinction is interpretable in more than one way, and its application to human language is far from simple. I begin by suggesting that it is type/token thinking (originating in the work of the philosopher C.S. Peirce; see [Peirce, 1933]) which underlies most linguists’ appeal to the ‘emic’/‘etic’ distinction, though not, perhaps, to the original conception of that distinction proposed by Pike [1954], which is behavioural in nature. When applied to the field of phonology, the emic/etic distinction results in (among other things) the phonemic/phonetic distinction, often indicated by means of forward slash brackets for phonemes (e.g. English /i:/) and square brackets for phonetic segments (e.g. English [i:]). Phonemes are often taken to be types, and specific speech sounds uttered on specific occasions to be their tokens. Thus Trask [1996: 364]: ‘type: a single abstract linguistic object, such as a phoneme.’ (I will return later to the senses in which phonological objects may be said to be ‘abstract’). See too Trask [1996: 355]: ‘token: a single pronunciation of a linguistic form by a particular individual on a particular occasion’. But there are conceptual problems with this way of defining the phonemic/phonetic distinction, and those problems arise for postulated phonological representations regardless of whether they take the form of phonemes or not. The square phonetic brackets are, in fact, used in two different ways in the phonological literature. They are used to transcribe specific speech sounds uttered on a particular occasion. This is the use made of square brackets in much corpus-based work, such as work in the PAC project (for an overview, see [Carr et al., 2004]). In this project, recordings are made of sets of informants who speak a range of varieties of contemporary English; the utterances are then phonetically transcribed and analysed. Tokens are spatio-temporally unique: a particular utterance of an [i:] by a particular speaker on a specific occasion is thus interpretable as a token of the type ‘i:’. We can therefore plausibly refer to ‘i:’ as a *speech sound type*, and the [i:] uttered on that occasion as a token of that type.

I have used the notation ‘i:’ here to denote the speech sound *type* in question, and [i:] (a long, high, unrounded vowel) to denote that specific *token*, uttered on that occasion. In the literature on phonetics and phonology, square brackets are used to denote both the speech sound type *and* its tokens: much of the ‘data’ that one encounters in phonological research contains transcriptions in square brackets.
that do not represent speech sounds in specific utterances; rather, they represent types of speech sounds. This is the second kind of use of square brackets in the literature. It is often not at all clear whether the term ‘phone’ is used in that literature to refer to speech sound types, or to specific utterances of such types. If it is reasonable (I claim that it is) to apply the type/token distinction to speech sound types and their tokens, where does this leave the phoneme? One might suggest that, on type/token assumptions, phonemes (if they exist: I claim that they exist as perceptual categories) may be conceived of as types which have speech sound types as their sub-types.

Bromberger & Halle [2000] argue that phonological theory has, among its objects of inquiry, tokens (such as tokens of words). But they adopt an instrumentalist interpretation of the notion ‘type’: ‘talk of types is just a façon de parler’ ([Bromberger & Halle, 2000: 35]; see [Carr, 1990] for discussion of instrumentalism in linguistics). Types, they suggest, simply do not exist (but they leave open the possibility that they might exist). However, it is arguable that one cannot have tokens without types, any more than one can have children without parents, or parents without children; type and token are defined in relation to each other (see [Burton-Roberts et al., 2000]). And indeed Bromberger & Halle, despite their apparent rejection of the notion ‘type’, end up arguing that certain tokens can be classified as ‘being of the same type’ or ‘being of different types’ [Bromberger & Halle, 2000: 32]. They insist that such classifications do not presuppose the existence of types, but it is difficult to take the notions ‘being of the same type’ and ‘being of different types’ as not involving appeal to types. Bromberger & Halle object to the notion ‘type’ because they take types, if they exist, to exist outside of space and time, an idea they understandably find ontologically questionable. But types need not necessarily be interpreted in this Platonic manner: we can interpret types as mental categories, and argue, as many have, that categorisation is at the core of human perception. If types are mental categories, then we can argue plausibly that human speech sound types are perceptual categories, in terms of which we perceive the speech signal. The question then arises whether phonemes, if they exist, are perceptual categories. There is certainly empirical evidence that subphonemic differences in specific languages begin to be ignored by infants during the first year of life: this is the ‘learning by forgetting’ phenomenon, identified by Jacques Mehler [1974] and colleagues. I now briefly exemplify this phenomenon.

It is known that human infants can discriminate aspirated stops (such as the aspirated stop found at the beginning of pit in most varieties of English) and unaspirated voiceless stops (such as the stop in the English word spy), probably from birth. So too can chinchillas [Kuhl & Miller, 1982]. But infants exposed to English, where the difference between aspirated and unaspirated voiceless stops is not phonemic, begin to fail to discriminate between these two categories during the second half of the first year of life. Infants exposed to languages such as Hindi and Thai, where aspiration is phonemic, continue to discriminate between the two sorts of stop during the second half of the first year of life. This strongly suggests that, even before (monolingual) infants utter their first attempts at words, they
are ‘tuning out’ from sub-phonemic properties of the speech signal: phonemic categories, construed as perceptual categories, are already being established at this early stage of development.

The issues surrounding the emic/etic distinction are not made any easier to resolve by the controversy over whether phonemes exist, and if so, what their ontological status might be. A purely instrumentalist interpretation of the notion ‘phoneme’ is possible: the term has been used instrumentally as a mere ‘façon de parler’, and this is one of the senses in which phonemes have been said to be ‘abstract’. Among the purportedly realist interpretations of the concept ‘phoneme’, Hyman [1975] offered three: (a) the phonetic reality interpretation (associated with the work of Daniel Jones), (b) the phonological reality interpretation and (c) the psychological reality interpretation. The Jones [1950] definition of the phoneme is arguably problematical: the phoneme defined as ‘a family of sounds’ is a set-based definition: phonemes here are taken to be sets of phonetically similar speech sound types. But sets, even phonetic sets, cannot be heard. We can coherently speak of phonetic sets, but that is not to say that the set itself has phonetic properties, either articulatory or acoustic. Sets, even phonetic sets, cannot be articulated or heard: such is the ontological nature of sets (the set of dogs, for instance, does not bark, since it is a set, but its members bark, and of course the members of a phonetic set can be heard). The question is what the cognitive status might be of the sets in question. The question also arises what Hyman’s ‘phonological’ reality interpretation might be. Hyman claims that the ‘purely phonological reality’ interpretation of the term ‘phoneme’ is characteristic of the Prague School. He claims that a phoneme, on this ‘phonological reality’ interpretation ‘is not a sound, or even a group of sounds, but rather an abstraction, a theoretical construct on the phonological level’ [Hyman, 1975: 67]. This appears to constitute a non-realist interpretation of the ‘phonological’ view of the phoneme: if phonemes are ‘abstractions’, nothing more than theoretical constructs, interpreted instrumentally, then they are not to be understood in realist terms. The expression ‘phonological reality’ here is thus, paradoxically, not to be interpreted in realist terms.

Regarding the putative psychological reality of the phoneme (Hyman’s third interpretation): this is often associated with the work of Edward Sapir [1933] and that of Chomsky & Halle [1968]. But there is more than one way of interpreting what ‘psychologically real’ might mean. In Carr [2000] I distinguish strong realism from weak realism: strong realism amounts to claiming that constructs, such as ‘the phoneme’, correspond directly to phonological representations stored in the mind/brain, and constructs such as ‘phonological derivation’ (appealed to by [Bromberger & Halle, 2000]) correspond to on-line processes occurring during acts of speaking. ‘Weak realism’ is the claim that our theoretical constructs are some kind of ‘indirect’ characterisation of mental states, with no commitment to what happens during on-line processing. In Carr [2000], I claimed that both strong and weak versions of psychological realism could be found in the work of Noam Chomsky. A strong version of psychological realism would amount to the claim that we possess phoneme-like mental representations: representations of phone-
mic categories, which figure in the ‘decoding’ of the speech signal. Under weak realism, the phonologist’s postulated phonemic entities are taken to be a way of characterising the native speaker’s mentally real lexicon, without claiming that it contains phoneme-like representations. Under this weak version of realism (which comes close to instrumentalism), we claim that there is a real mental lexicon, and that talk of phonemes is talk of the similarities and differences between gestalt-like stored acoustic images.

However we interpret ‘phoneme’, it is often claimed that phonemes are ‘realised’ phonetically. Exact characterisations of what ‘realisation’ might mean are difficult to find. The notion seems to bring with it the idea that postulated phonemes are somehow ‘less real’ or ‘more abstract’ than their ‘realisations’, that phonemes are ‘made real’ via the process of realisations. But if one is to adopt some version of ontological realism with respect to phonemes, then it makes little sense to speak of real entities being made ‘real’, or made ‘more real’. It may be the case that talk of ‘realisation’ is equivalent to talk of ‘instantiation’, in which case, all of the conceptual problems inherent in the ‘instantiation’ account of phonological entities re-appear on the ‘realisational’ account. However, it may be the case that some appeals to ‘realisation’ are appeals to the idea that speech sounds are externalizations of something which is mind-internal (I will return to this idea below).

Burton-Roberts ([in preparation]; henceforth BR) makes an interesting attempt to unpack some of the intended meanings of ‘realisation’ as used in generative linguistics. An important point in BR is the distinction between what he calls the generic conception of language and the realist (naturalistic) conception. Under the generic conception, the term ‘language’ is a cover term for all particular languages: the study of language is thus the study of human languages. This is perhaps the lay, common sense, understanding of what language is, and is probably also the interpretation of ‘language’ adopted by many researchers in linguistics. It is quite distinct, BR insists, from the realist/naturalistic conception, associated with the work of Chomsky, according to which ‘language’ is a real object in the natural world, an innate endowment which is quite distinct from any particular socio-politically determined language. Chomsky uses the term ‘E-language’ to refer to the latter, and denies that E-language constitutes the object of linguistic inquiry; BR agrees with Chomsky on this score, but suggests that Chomsky’s thinking is inconsistent, in that aspects of the generic conception creep into Chomsky’s naturalism. I lack the space to pursue BR’s critique of Chomsky here, but I will return to the generic vs. naturalistic distinction in what follows.

To return to ‘realisation’: BR identifies two interpretations of the term. The first is an indexical (semiotic) interpretation. The second is a type/token interpretation. BR adopts the generative view that mentally constituted grammars generate linguistic objects, whereby ‘generate’ simply means ‘defines’. As is often (but inconsistently) done in the generative literature, he distinguishes sharply between these objects and the speech sounds uttered by speakers. He argues that, notwithstanding the resistance in generative linguistics to a semiotic conception of
language, if ‘realisation’ is implicitly conceived of as a semiotic relation, then the semiotic relation that most accurately reflects what generativists seem to mean by ‘realisation’ is indexical in the sense of Peirce [1933]. Indexical signs are natural in the sense that they are causally related to what they are signs of. Examples are smoke as a sign of fire, or human footprints in snow as a sign of the presence of a human being. Since indexical signs are natural, they are unlike symbolic signs (such as traffic lights), which are based on conventions. They are also unlike iconic signs (such as Magritte’s painting of a pipe), which are based on perceived resemblance. One of BR’s points is that the use of the term ‘realisation’ in generative linguistics often makes implicit or explicit appeal to the idea that speech sounds are causally determined by the workings of a mentally constituted grammar, since the objects generated by the grammar are said to have phonological properties which, for most practitioners of generative phonology, are taken to have intrinsic phonetic content. Thus, assuming a computational theory of mind [Fodor, 1998], the speech sounds emitted during an utterance are viewed as the end result of a linguistic computation. Bromberger & Halle [1996: 446] are explicit about this when they suggest that ‘the motions of speech organs are shaped by grammar’. BR argues that this conception of the ‘realisation’ relation is at odds with the widely held view that sound-meaning relations are symbolic (conventional, more specifically arbitrary, i.e. non-natural), rather than indexical. He seeks to sustain the claim that speech sound tokens function as signs, but not indexical signs of a linguistic computation.

In discussing the interpretation of the ‘realisation’ relation as some version of the type/token relation, BR distinguishes between what he calls the ‘classical’ conception of type/token and his own ontologically ‘sorted’ conception of type/token. On the ‘classical’ Peircean conception, tokens are spatiotemporally unique mind-external physical phenomena, as noted above, and types are not spatio-temporal. If the objects generated by a mentally constituted grammar are implicitly or explicitly conceived of as types, and if they are taken to be realised in speech, then utterances are taken to be tokens of those types. BR argues that this classical interpretation of the type/token relation is inapplicable to linguistic objects, conceived of as the objects generated by a mentally constituted grammar as conceived of by him. The argument runs as follows. On the classical interpretation, BR argues, types are types of mind-external physical phenomena (note that BR is assuming here a mind-internal vs mind-external distinction: see the final section for an alternative to this distinction). BR refers to these kinds of physical realities as ‘E-physical’, meaning external to the mind/brain, whether conceived of as entirely physical or not. But the linguistic objects generated by the grammar are not, for him (and he argues, not for Chomsky), types of E-physical phenomena. The relation between linguistic objects and spatiotemporally unique speech events cannot therefore be the classical type/token relation.

BR therefore proposes a novel conception of the type/token relationship: the ontologically ‘sorted’ conception. Syntactic structure, he suggests, is not the sort of structure that can be heard or seen: it is abstract in the sense that it is generated
by a wholly mind-internal grammar. By ‘wholly mind-internal’, BR intends a rad-
ically mind-internal state, distinct from what may be called weakly mind-internal
states and processes such as perceptual processes and internalised acoustic and
visual images. The radically mind-internal state in question is not internalized
since what is internalised was initially mind-external. It is arguably this version
of internalism that Chomsky [2000] has in mind when he speaks of the ‘austere’
inner component of mentally constituted grammars, as distinct from weakly mind-
internal performance systems. Given the ontological status of linguistic objects as
postulated by BR, and given a conception of those objects as types, it is im-
possible that they may have tokens of an E-physical sort. ‘Sorted’ type/token
distinguishes between types whose tokens are spatio-temporally E-physical (such
as tokens of sheep or tables) and types whose tokens are radically mind-internal,
such as linguistic types. This leaves BR with a question to answer: what, then,
is the relation between E-physical tokens, such as speech sounds, and linguistic
objects? His response is that the sequences of speech sounds we produce are pro-
duced as E-physical representations (not tokens) of the objects generated by the
grammar. This sense of ‘representation’ differs from Chomsky’s: it is ‘represen-
tation’ used in the everyday two-place predicate sense. BR is at pains to point
out that this relation of representation is quite distinct from the ‘realisation’ rela-
tion: an E-physical representation of an object is not an instance of that object:
linguistic objects, for BR, have no phonetic or phonological properties. Rather,
phonology is for the physical (phonetic) representation of phonology-free linguistic
objects.

BR’s stance presupposes at least the following things: (a) the validity of the
computational theory of mind, (b) a mind-internal vs mind-external distinction,
(c) a distinction between that which is radically mind-internal and that which
is weakly mind-internal, (d) the idea that linguistic objects are generated by a
grammar, (e) a distinction between that which is E-physical and that which is
mental. All of these are open to debate: the mind may not be computational (see
below on ‘distributed language’ for a claim that it is not); that which is mental
may be conceived of as intersubjective (‘distributed’, in Cowley’s [2007b] sense:
again, see the final section for discussion); there may be no radically internal
states; linguistic objects may not be generated by a grammar (assuming that such
objects exist), and (e) might be taken by some to constitute a version of Cartesian
dualism, which has been regarded by most as unsustainable (but see [Sampson,
2005] for a defence of such a dualism).

**Representation**

One occasionally encounters the claim that phonetic segments ‘represent’ phonemes,
without explicitly defining the notion ‘represent’. An example of this comes
from Wells’ [1982] classic three-volume work on a wide variety of accents of En-
lish: ‘It was the great achievement of phoneticians and linguists in the first
sixty years of the twentieth century to develop the concepts of phoneme and
allophone...sounds constituting a set of this kind may indeed differ physically (‘phonetically’, ‘allophonically’) from one another, because they are different allophones of the phonemes in question; but they are the same linguistically (‘functionally’, ‘phonemically’, ‘phonologically’), because they represent the same phoneme’. [Wells, 1982: 41]

In the same volume, Wells [1982: 73] adopts the view that phonemes are phonetically realised. It is not clear, however, whether it is conceptually coherent to claim that phonetic segment types represent phonemes, while at the same time claiming that they are realisations of phonemes. If ‘represent’ is used in one of its the everyday senses, as a two-place predicate (as in ‘That flag represents the Welsh nation’), then it would appear incoherent to say that phonetic segment types both represent phonemes and realise them: a token of the type ‘The Welsh flag’, if it represents the Welsh nation (leaving aside the ontological status of nations), is surely not a realisation of the Welsh nation. It is also worth noting that Wells adopts the same set-based interpretation of the phoneme as was adopted by Jones. This means that he is committed to arguing that individual members of a set of sounds represent that set. It is not clear that this argument is coherent, but in the absence of any clear definition of what ‘represent’ is intended to mean here, it is difficult to assess Wells’ claims.

**Transduction**

A transducer converts a specific kind of input into a distinct, but related, kind of input. An example is the mouthpiece on a telephone: it converts acoustic input into electrical output. That output is transmitted and is then converted back into an acoustic signal. I will take it that transduction, conversion and transmogrification are the same thing: the conversion of a given kind of input into a different kind of output. For transduction to take place, the two kinds of signal have to be of the same physical type: one cannot enter water into the mouthpiece of a telephone and have it converted into an electrical signal (one might be able to transducer the sound of the poured water into an acoustic signal, but not the water itself). Let us distinguish between what is technologically impossible for transducers and what is impossible in principle. The earliest telephone mouthpieces were technologically incapable of converting visual input into a transmissible electrical signal. Today’s digital mobile phones can now do this. But no telephone will ever be capable of converting, say, the meaning of the word *however* into a visual image, or visual images stored in the mind. And no transducer will ever be capable of converting non-E-physical objects (in BR’s sense) such as concepts into acoustic or visual signals. When I say ‘non-E-physical’ here, I mean ‘incapable of being perceived by any of the five senses’: concepts (which I simply assume to exist; I will not attempt to engage with the complex literature on how they might be defined: see [Fodor, 1998] for discussion) are not perceptible. Such is their ontological status. Whether phonological objects may be said to be capable of undergoing transduction into an acoustic signal will depend on one’s conception of the ontological status of phono-
logical objects. Note that transduction is non-arbitrary and thus bi-unique (e.g. acoustic > electric > acoustic).

An appeal to transduction in phonology is offered by Hale & Reiss ([2000]; henceforth H&R), who adopt the naturalist/realist conception of language. They take language, understood in this naturalistic sense, to include phonology. They insist that mind-internal phonological representations are ‘substance-free’, by which they mean not grounded in phonetics. Adopting a computational theory of mind, they argue that phonological objects are substance-free symbols manipulated during mental computations in a manner no different from syntactic symbols. This is at the root of some of their objections to Optimality Theory (OT), a widely-adopted constraint-based version of generative phonology in which rule-based derivations are replaced by constraint evaluations. H&R object that many of the phonological constraints postulated in OT are substance-based, and thus fall outside of what they take to be a strictly linguistic conception of phonology, in which ‘strictly linguistic’ is understood in the naturalistic interpretation of the terms ‘language’ and ‘linguistic’. One of the merits of H&R’s stance is that they are explicit about their conception of human language, unlike many of the practitioners of OT: very little is said in the OT literature about the underlying assumptions regarding human language. There appear to be two general types of interpretation of the nature of constraints in the OT literature: one takes constraints to be grounded in phonetics, while the other appears to be agnostic on this matter. H&R, having adopted a naturalistic/realist conception of language, have understood that phonetically-grounded phonological knowledge has no place in an innately-endowed language module which is distinct from our articulatory and perceptual capacities. Phonology must, therefore, be substance-free. In order to give an account of how those putatively substance-free phonological objects relate to speech events, they postulate that a process of transduction converts them into articulatory, and thus acoustic, events. But this is impossible: the postulated objects, being substance-free mental representations, are of a completely distinct ontological category from articulatory and acoustic events. Acoustic events can be perceived by one of the five senses; phonetic-free mental representations are unperceivable. Such mental representations can no more be converted into phonetic events than can concepts. The rôle of transduction in Hale & Reiss’s conception of phonological knowledge is, I suggest, deeply problematic: transduction is non-arbitrary, but Hale & Reiss insist that the relation between phonological objects and phonetic events is arbitrary.

Transmogrification

Bromberger & Halle ([2000]; henceforth B&H) conceive of phonological objects as mental representations, which are said by them to have intrinsic phonetic content. By this is meant, not that one can literally hear phonological representations (one could not, in principle, since mental representations, by their very nature, cannot be perceived by any of the five senses). What they mean, I take it, is that
these are representations of features of acoustic input: these are representations internalised from the mind-external environment. They then allow that these representations can be transmogrified (converted) into articulatory movements which create acoustic signals; those signals can then be perceived by our fellow speakers and mapped onto their mind-internal phonological representations, which are assumed to be the same as those of their fellow native speakers (this is an assumption which has been challenged in various quarters). This conversion and transmission model appeals to the twin ideas of encoding and decoding: phonological objects are said to be encoded during the conversion process and then decoded by the listener. B&H interpret phonetically-grounded phonological representations as both internalised from acoustic input, but also as constituting articulatory intentions of some sort. This conception of the relation between phonological objects and the acoustic events encountered in human speech is perhaps sustainable, but, when coupled with a rule-based, derivational conception of phonological knowledge, it is harder to sustain, since B&H have to postulate that a set of articulatory intentions is transformed into a distinct set of articulatory intentions by a phonological rule, and then transformed again by the application of the next phonological rule, and so on until the derivation is complete. As a picture of how speech planning works, this seems implausible. It also runs into the problems encountered in the history of transformational generative grammar with respect to the derivational theory of complexity: since B&H’s model is a model of on-line production of speech events, the model predicts that, the longer the phonological derivation for a given utterance, the longer it should take to produce. Equally, the longer the phonological derivation, the longer a given utterance should take to be ‘decoded’ by the hearer. It seems unlikely that any psycholinguistic evidence will be forthcoming to corroborate these claims. B&H’s view of phonological representations as articulatory intentions might fare better if coupled with a model of phonology that is non-derivational.

The status of the notions ‘phoneme’ and ‘phonetic segment’

The notion ‘phoneme’, often said to date back to the work of the 19th century Polish linguist Jan Baudoin de Courtenay, has been under attack for decades, at least since the time of the debates between Jones and Firth in the mid-twentieth century. One objection is that linguists who adopt (some version of) the notion ‘phoneme’ are the victims of ‘alphabetism’ [Silverman, 2006: 203]: because most linguists are literate in an alphabetic writing system, they are said to have the intuition that the stream of speech is made up of sequences of phonetic segments, which may be regarded as the realisations of underlying sequences of phonemes. This, Silverman argues, is an illusion: the stream of speech cannot be segmented into phones. Silverman, like Port [2010a], is arguing against the reality, not just of phonemes, but of phonetic segments. He argues that ‘the consonant and vowel sequences that we think we observe are simply artefactual, and it is the transitions between them that are most relevant, since these are the most informationally
rich and often the most auditorily prominent components of the speech signal’ [Silverman, 2006: 216]. But talk of transitions between consonants and vowels presupposes that consonants and vowels exist. The same worry arises with respect to Port’s [2010] rejection of the idea that consonants and vowels exist, since he subsequently goes on to show a spectrogram of the syllables di and du, in which he speaks of ‘the burst into the vowel’ in the utterance of the word *dew*.

Port (p.c.) suggests that he often uses terms such as ‘consonant’ and ‘vowel’ for the sake of making sure that his readers know what he means, but without commitment to the idea that consonants and vowels are psychologically real entities. This is the instrumentalist *façon de parler* defence often offered by linguists who deny that certain constructs in linguistics actually correspond to anything real: talk of consonants and vowels is to be interpreted instrumentally. But there is something worrying about this defence. Port and Silverman both take themselves to be doing science. But, in the history of science, when a consensus emerged that phlogiston didn’t exist, physicists simply abandoned the term (even as a *façon de parler*): there was simply no more use of the term in scientific discourse. If we find it impossible to talk of phonetics and phonology without talking of consonants and vowels, isn’t that perhaps because there are consonants and vowels, if only in our perceptual systems, independently of knowledge of writing systems, constituting part of ‘the furniture of the world’ (to use Bromberger and Halle’s expression)? Where is the alternative metalanguage in the work of Silverman and Port, free of talk of consonants and vowels? It appears not to exist. Note too that one of the first things one encounters in Silverman’s fascinating book (the basic tenets of which deserve to be taken seriously) is the chart of symbols which constitutes the IPA: what place does such a segment-and-phoneme-based chart have in a book which denies the existence of segments? It is rather like writing a book on the present-day elements (depicted in the periodic table) beginning with a depiction of the four elements earth, air, water and fire, only to go on using those terms while arguing that talk of those elements is a mere *façon de parler*. One might argue that the analogies with phlogiston and ‘the four elements’ is less persuasive than other instrumentally-interpreted scientific terms, such as, say, ‘force’ and ‘atom’. Let us take talk of ‘forces’: one might argue that, if the equation $F = ma$ (force equals mass times acceleration) is true of the world, then talk of forces is mere shorthand (‘*façon de parler*’) for talk of accelerating masses, and indeed such an instrumentalist interpretation of ‘force’ was adopted in the twentieth century [Mach, 1893/1966]. But physicists continued to talk of forces, and this fact is good grounds for assuming that forces are part of the furniture of the world: who now would deny that electromagnetic and gravitational forces are real? If talk of gravitational force is instrumentalist talk, why do planes stay in the sky, rather than exiting the atmosphere? The concept ‘atom’ was also attributed a purely instrumental interpretation by some scientists, but physicists continued to talk of atoms, for the good reason that atoms almost certainly exist, as do sub-atomic particles. I suggest that phonetics/phonology specialists will go on speaking of consonants and vowels because they have no alternative, and that they have no
alternative because consonants and vowels are part of the furniture of the world.

Now consider Port’s [2010a] objections to phones and phonemes (I will return, in the final section, to more general issues addressed by Port). Port argues that a language is a set of conventions shared by members of a community. This is not a new claim (see Itkonen 1978 for a fully elaborated conception of linguistic conventions), but what is new in Port’s work is the combination of this view with a consideration of the richly represented nature of human memory traces, as we will see. The social conventions in question, Port argues, ‘are represented independently in the memory of each community member’ [Port, 2010a: 45]. The memories stored by speakers are said to be rich, in the sense that they contain fine-grained phonetic detail, as well as, among other things, details about the voices of specific speakers. But those memories do not, according to Port, contain discrete segment-like objects. Nonetheless, Port allows for the existence of speech sound types: ‘Apparently, human speakers have the ability to hear a novel linguistic stimulus and to find the appropriate linguistic categories (e.g. word and phrase identities, the identities of various speech sound types, etc) by searching a large, personal utterance memory for all the many kinds of closest matches (see [Johnson, 2006])’ [Port, 2010a: 48]. But if speech sound types exist, so must speech sound tokens, and it is not clear what a speech sound token might be if it is not a speech sound, i.e. a segment. The question here is whether Port can sustain a definition of ‘speech sound token’ that is not segment-like.

Fowler [2010] counters Port’s position with several arguments. Firstly, speakers make spontaneous speech errors involving what look a lot like segments, as in the case of Spoonerisms. Note too that these abound in child speech, and are very plausibly interpreted as errors in speech planning. Secondly, she points out that it is common to find, in naturally occurring systems, particulate organisation, and this appears to be a defining feature of human languages: the three segments [a], [t] and [k] can be combined to form act, tack or cat. Thirdly, as I argue below, alphabetic writing systems could not have been elaborated unless the people who elaborated them perceived segments in the stream of speech: if, as Port and Silverman argue, it is solely knowledge of alphabetic writing that induces the purported illusion of segments, then the originators of alphabetic writing systems could not have suffered that illusion. Fourthly, how could we interpret the various tokens of a given word (say, cat) as tokens of that word unless there were a type of which those were tokens? And how could we interpret the various tokens of a given speech sound type unless there were a speech sound type of which those were its tokens? Finally, Fowler points out that, even given the rich memory indicated by Port, memories of words will be stored in vectors of features. There is no reason not to assume that, in those vectors, phonetic features will be represented.

In his rejoinder to Fowler’s argument about the invention of alphabetic writing, Port argues that ‘I certainly would not say that our impression that speech has segment-sized units is due solely to alphabet training’ ([Port, 2010b: 61]; emphasis in the original). And yet in the article Fowler is discussing, Port says ‘the seeming directness of segmental speech perception comes about only when we learn to
read, not when we learn to speak’ ([Port 2010a: 48]; emphasis mine). A further puzzling aspect of Port’s position is that he assumes that, in mental storage, ‘There is apparently no way to separate linguistic from nonlinguistic information’ [Port, to appear]. If this is so, how can I identify that a given speaker is male, even if I don’t understand a word of the language he is speaking? In such cases, I can access very little (or no) linguistic information, but I can tell that a man, not a woman, is speaking.

The claim that phonetic segments ‘do not exist’ appears to be based solely on articulatory and acoustic facts: if you can’t see them in a spectrogram, they do not constitute part of the furniture of the world. There is, I suggest, an unsustainable externalist physicalism underlying this view, incompatible, I suggest, with a cognitive interpretation of phonology. The main perceptual point made by those who accuse phonologists of ‘alphabetism’ (let us call them ‘segmental eliminativists’) is that phonologists are suffering a perceptual illusion that they can hear consonants and vowels, induced entirely by the fact of being literate in an alphabetic writing system. I suggest that alphabetic writing came about for the same reasons that word-based and syllable-based writing systems came about: because human beings perceive words, syllables and segments in the stream of speech, even in societies with no writing. The fact that there are no clear divisions between segments in a spectrogram does not show that these are not units of human perception.

Supporters of the ‘alphabetism’ charge are also at a loss to explain the striking success of alphabetic writing. Port [to appear] argues that ‘learning to read one’s language via an arbitrary set of graphic shapes is intrinsically difficult. It takes systematic training for several years.’ This is undeniably true. For Port, this is because reading and writing in an alphabetic system does not correspond to the phonetic reality of human speech. But the difficulty may arise because the child’s phonological representations are not normally accessible to conscious awareness: becoming literate in an alphabetical writing system is difficult because it forces the child to become consciously aware of unconscious representations. This, I suggest, is also why linguistics in general, and phonetics/phonology in particular, is a difficult area of study for most human beings: it is humanly abnormal to seek to consciously contemplate that which is largely non-conscious. Given that learning to read and write in an alphabetic system is so difficult, why have alphabets been so successful? Perhaps because, despite the inherent cognitive difficulty in learning such systems, they map onto (some of) our perceptual categories. There is clear evidence that humans who know how to read and write in an alphabetic system have two different mechanisms for reading. One is a holistic visual word recognition system of the sort that Port (quite rightly) postulates: it is simply impossible to read properly without engaging a ‘snapshot’ written word recognition mechanism. The other is the set of grapheme-to-phoneme correspondences which we have been trained in at school. Evidence that both of these mechanisms are available to alphabetically literate speakers is clear: under conditions of acquired dyslexias, the two doubly dissociate. In subjects with acquired ‘phonological’ dyslexia, the grapheme-to-phoneme correspondences are either lost or inaccessible,
but the holistic visual word memories are spared. Such subjects cannot properly pronounce words they have never seen before: given a made-up English word such as *blug*, they either cannot pronounce it at all, or they utter some similar word, such as *bug* or *blue*. For sufferers of acquired ‘surface’ dyslexia, the reverse is the case: they are unable to recognise written words as holistic units. They can pronounce words with regular spelling, because they still have access to the grapheme-to-phoneme correspondences, but with words with irregular spellings, such as *yacht*, they can fall back only on the grapheme-to-phoneme correspondences, so that pronunciations such as [jatSt] are produced [Ellis & Young, 1996]. The main point here is that it is impossible for humans to internalise grapheme-to-phoneme correspondences unless phoneme-sized, or at least segmental speech sound sized representations, are already stored in the mind/brain: graphemic units are put into correspondence with pre-existing segment-sized units. The child cannot learn the spelling of *cat* without pronouncing [k] – [a] – [t]. Nor can a French child learn consciously that there are two syllables in *bateau* without pronouncing the two syllables [ba]-[to]. The reality of such syllables is abundantly evident in French speech phenomena such as liaison-plus-hesitation, as in Les..euh.. [z]espagnols, where the liaison consonant appears after the hesitation vowel. Such phenomena are inexplicable if consonants, vowels and syllables are not phonetic realities.

To support the claim that units such as segments are real units of speech perception, even if they cannot be found in spectrograms, consider the case of stress-timing in (most varieties of) English: it has been shown that the isochronicity between stressed syllables in English which has been postulated by phonologists does not show up in physical measurements. That, I suggest, does not demonstrate that the perception of isochronicity is not psychologically real for speakers of English. Furthermore, the tendency towards the production of isochronous metrical feet in English is supported by production phenomena such as the reduction and elision of vowels and consonants in unstressed syllables. One might ask what underlies the tendency in humans to perceive sequences of stressed and unstressed vowels, or the production of sequences of consonants and vowels. I suggest that there is a perceptual preference (which I informally call ‘The Yin-Yang Preference’) for sequences of alternating opposites in the acoustic input: sequences such as consonant-vowel-consonant-vowel aid in our perception of speech (and are rooted in infant babbling), just as sequences of stressed-unstressed-stressed syllables do. Widely attested phenomena such as Word Stress Clash Avoidance and phenomena such as Iambic Reversal in English are, I suggest, rooted in this preference: Iambic Reversal, exhibited in prominence reversal in pairs such as *Piccadilly vs Piccadilly Circus* exhibit a similar preference for sequences of strong-weak-strong metrical feet, as opposed to weak-strong-strong sequences. Of course, many other factors intervene to create less eurythmic sequences, so that human languages can indeed exhibit word stress clashes, and sequences other than consonant-vowel-consonant-vowel.
2 PHONOLOGY, GROUNDEDNESS AND THE INTERPRETATION OF ‘THE LINGUISTIC SIGN’

If there is such a thing as phonological knowledge, is it distinct in kind from syntactic knowledge? Are there parallelisms between phonological and syntactic knowledge? One response to the first question, proposed by Bromberger & Halle [1989; 2000] is that phonology is different from syntax (see below on the work of Joan Bybee for an opposing claim). Adopting the kind of derivational, rule-based conception of phonological computations associated with the early days of generative phonology [Chomsky & Halle, 1968], they argue that, while syntax is no longer to be conceived of as involving the application of a series of rules in a derivation, phonology is thus to be conceived. They also adopt the view that phonological representations have intrinsic phonetic content, a view that goes back to the beginnings of generative phonology. That view is not universally upheld, however: there are phonologists who argue that, in order to sustain a coherent mentalistic conception of phonological knowledge, it is essential to conceive of phonological representations as devoid of phonetic content. This view can, arguably, be traced back to the work of Louis Hjelmslev (but see Anderson to appear for a suggestion that Hjelmslev was inconsistent on this count). Present-day proponents of the ‘phonetics-free phonology’ view are certain practitioners of Government Phonology and related frameworks [Kaye et al., 1985]. Hale & Reiss [2000] also argue against what they call ‘substance abuse’ in phonology, as we have seen.

Most phonologists, however, still support the view that phonological objects are grounded in phonetics. However, if one adopts that view, and if one also adopts the Chomskyan view that linguistic knowledge is knowledge without grounds [Chomsky, 2000], then it would appear to follow that phonological knowledge is not linguistic knowledge, precisely because it is conceived of as grounded in phonetics (see [Carr, 2000] for discussion of this point). This is the view of Burton-Roberts [2000], who, as we have seen, argues that linguistic objects possess no phonological properties. BR accepts the view that phonology is grounded in phonetics, but denies that the phonological is linguistic. In fact, he denies that phonological objects exist: for BR, a phonological system is a system of conventional rules to be followed in the E-physical representation of linguistic objects, conceived of as purely syntactic-semantic in nature.

John M. Anderson ([2006; to appear], and elsewhere), like BR, argues that phonology is grounded in phonetics, and syntax grounded in semantics. Anderson also adopts a version of the structural analogy principle: the view, again associated with the work of Hjelmslev, that there are, subject to certain limitations, structural analogies between syntactic structure and phonological structure. The reason that Anderson expects to find such analogies is that he takes linguistic knowledge, pace BR and Chomsky, to be grounded in facts about domain-general human cognitive capacities. An example of a syntactic/phonological analogy postulated by Anderson is as follows. The head-dependent relation is said by Anderson to be contracted by both syntactic and phonological objects. For instance, transitive
verbs are said to contract head/dependent relations with their arguments, as are the constituents of syllables. Furthermore, the complement/adjunct distinct is said by Anderson to be applicable in both syntax and phonology: in a word such as pact, the vowel is said to be the head of the syllable (it is the head in the phonetically-grounded sense that it is the perceptually most salient constituent in the syllable). That head, Anderson argues, takes the /k/ as its complement. Similarly, the transitive verb kicked in The man kicked the dog takes the dog as its complement. Here, the verb is said to be the head of the verb phrase: it is the semantically most salient constituent in the verb phrase. In English verb phrases, adjuncts (modifiers) normally have to be stacked outside of complements, as in The man kicked the dog on Saturday, as opposed to *The man kicked on Saturday the dog. Anderson analyses the /t/ in pact as an adjunct, which may not occur between the nucleus of the syllable and its complement: */patk/.

It is widely assumed that human languages contain sound/meaning pairings, often referred to as ‘linguistic signs’. My aims here will be to clarify exactly what this might mean, to show that there is more than one way of interpreting the notion ‘linguistic sign’, to elaborate the conceptual problems with those interpretations, and to consider the relationship between phonology and syntax (since this is a topic which arises from a consideration of linguistic signs). The notion ‘linguistic sign’ is often associated with the work of Saussure (1916), but Saussure’s notion ‘linguistic sign’ is often interpreted in more than one way, and these different interpretations are relevant for the status of phonological knowledge. One interpretation (perhaps the one intended by Saussure) is that a linguistic sign consists of some kind of coupling, or association, between an acoustic image, stored in the mind, and a concept (I will not dwell here on the complex issue of what exactly concepts might be: see Fodor 1998 for extensive discussion). On this definition of ‘linguistic sign’, signs are entirely mind-internal: both the acoustic image and the concept are located in the minds of the speakers of a specific language. It is this interpretation of ‘sign’ that John M. Anderson (e.g. [Anderson, 2005; to appear]): adopts. It is this central claim that leads him to reject notions such as empty categories, since such postulated entities are not, he argues, linguistic signs: they lack a phonology. Anderson, as we have seen, argues that human language is rooted in general cognitive capacities; he thus rejects the Chomskyan claim that human beings are born with an innate, specifically linguistic, module of mind, distinct from general cognitive capacities. He adopts a version of the structural analogy principle, the view that there are significant analogies between phonological structure and syntactic structure. Among the analogies postulated by Anderson are the putative presence of headhood, dependency, agreement and spreading of features in both syntax and phonology. It is crucial for Anderson that the notion ‘linguistic sign’, as interpreted by him and many others, is sustainable.

However, a sustained attack on the Saussurean notion of sign can be found in work by Noel Burton-Roberts (henceforth BR). In Burton-Roberts [to appear], BR argues that, for Saussure, the acoustic image (the mentally represented phonological representation) and the concept are two ontologically distinct objects which
are nonetheless said to combine to make a third object, the linguistic sign. It is this claim that BR objects to. He points out that the relation between an acoustic image and a concept is, for Saussure, a part-part relationship: the concept and the acoustic image are the two parts of the sign. BR uses the term ‘mereology’ for the study of part-whole and part-part relationships: the Saussurean conception of the relationship between the concept and the acoustic image is mereological. However, for Saussure, it is also semiotic: the relation between the acoustic image and the concept is a semiotic one (the acoustic image signifies the concept). BR notes that these two features of the Saussurean sign are often taken to be mutually compatible, but he contests this. A semiotic conception of the relation between phonological representations and concepts does not require (and, he argues, is inconsistent with) the postulating of an entity which consists of both a phonology and a concept. BR denies that there are linguistic objects which consist of a phonology and a concept: the respective contents of phonological representations and concepts are sortally distinct and thus cannot combine to form some third kind of entity. It is this sortal distinctness that underlies and explains Saussurean arbitrariness. With his notion of physical representation BR can give an account of what the relation between phonology and concepts is while denying that there are linguistic signs which combine the two. For BR, it is the phonological representation (or rather speech events implementing phonologically encoded instructions) which constitute signs, in the mind-external Peircean sense. BR argues that one of the consequences of Saussure’s view that concepts do not pre-exist signifiers is an extreme version of the Sapir-Whorf hypothesis: the claim that all thinking is carried out in some particular language. This version of Sapir-Whorf faces difficulties. Consider work by Brulard et al. [in preparation] on two siblings exposed to two languages (French and English) from birth. In the diary data and digital audio recordings collected by Carr & Brulard, these siblings, when asked a question in French, may reply in English or French (or in a code-switched utterance with both languages), and when asked a question in English, may reply in either French or English (or, again, in a code-switched utterance containing both French and English). A simple example is this: Parent: ‘T’as vu le chien?’ (‘Have you seen the dog?’) Child: ‘What dog?’ Since the siblings respond appropriately to the questions, they have clearly understood those questions. But the level at which they have understood those questions simply must be a conceptual level, which must be something other than English or French: the grasping of the meanings of the questions must be carried out in a purely conceptual vocabulary, something other than the vocabulary of English or the vocabulary of French, each with their own phonological representations for the concept ‘dog’. This is not to deny that there can be such a thing as ‘Thinking in French’ and ‘Thinking in English’, or Slobin’s [1996] ‘thinking for speaking’: it is simply to deny that all thinking is conducted in some particular language. Whether the purely conceptual vocabulary appealed to here is supplied by a Fodorian Language of Thought is a question I will leave open here. The main point is that, on BR’s conception of the linguistic sign, phonological objects are not linguistic objects (where ‘linguistic’ is to be understood in what
BR calls the ‘naturalist/realist’, rather than the ‘generic conception of language). Note, however, that BR’s approach depends on a clear distinction between that which is mind-internal and that which is mind-external. I will turn later to work which questions this distinction.

3 THE ACQUISITION OF PHONOLOGICAL KNOWLEDGE

I will consider here the status of phonological development with respect to the Chomskyan Rationalist (often referred to as nativist) conception of the child’s linguistic development, and alternative conceptions adopted by Gopnik [2001], Karmiloff-Smith ([1998] and elsewhere), Sampson [2005], Tomasello ([2001] and elsewhere) and Vihman ([1996] and elsewhere). I firstly address some terminological matters which are simultaneously conceptual matters. Under the Chomskyan conception, the child is said to be born with a module of mind that is dedicated to language and is species-specific. This module, as we have seen, is often referred to by Chomsky as ‘the language faculty’, where ‘faculty’ is synonymous with ‘module’ (thus Fodor’s [1983] talk of ‘faculty psychology’). The term ‘Universal Grammar’ (UG) was used for many years by Chomsky to refer to this putative module; while Chomsky now uses the term ‘UG’ to refer only to the study of the putative language module (rather than the module itself), many of his followers continue to refer to the module as ‘UG’. The term ‘the language organ’ has also been used to refer to this postulated module; I will use this term and the term ‘the language module’. While Chomsky allows, as he must, that the child requires input from speakers of the ambient language(s), that input (which he often refers to as ‘primary linguistic data’) is said by Chomsky to have a merely ‘triggering’ rôle (rather than a determining rôle) in the child’s development: the input is said to trigger a biologically pre-determined path of development. The child’s development is conceived of as biological growth, parallel to the growth of other organs [Chomsky, 1976: 76] Thus the idea of a ‘language organ’. As Chomsky puts it, the child’s linguistic development, because it is conceived of as a kind of biological growth, is not best characterised as something the child does: it is something that happens to the child [Chomsky, 2000: 7], rather in the way that puberty is something that happens to humans. It is worth noting that this is a purely passive conception of linguistic development, incompatible with the idea that the child is actively formulating and testing hypotheses about the ambient language (see discussion below of the work of Vihman, and of Tomasello, on this idea). The term ‘language acquisition’ is not, I suggest, a terribly accurate expression to use in characterising most aspects of Chomsky’s conception of child linguistic development: that which is acquired cannot be innate, and that which is innate is, by definition, not acquired. This terminological point relates to the distinction, discussed above, between the naturalistic and the generic conceptions of ‘language’: what is acquired, according to Chomsky, is a specific language, as appealed to on the generic conception of ‘language’, whereas what is not acquired is ‘language’ in the naturalistic sense. I will use the conceptually more neutral term ‘linguistic development’ in discussing
Chomsky’s position. Note that, on a naturalistic conception of ‘language’, the term ‘correct’ has no place in a Chomskyan conception of linguistic development: correctness relates to norms (conventions), but conventions have no rôle to play in the purely biological Chomskyan view of linguistic knowledge: conventions are social, intersubjective, in nature, and Chomsky denies that language is a social, intersubjective reality (this conception of language being characterized by Chomsky as ‘E-language’). For the sake of discussion, I will use the more neutral term ‘well-formed’, though I personally believe that Esa Itkonen’s [1978] term ‘correct’ is the most appropriate term, since I agree with Tomasello that the child is, among other things, actively engaged in acquiring linguistic conventions (on the generic interpretation of ‘language’ and ‘linguistic’) which are social in nature; I will elaborate on this below. One final terminological point: the term ‘mastery’ is inappropriate for a Chomskyan conception of the child’s linguistic development, since ‘mastery’ relates to skills, capacities to do certain things, and Chomsky insists that linguistic knowledge is not a matter of knowing how to do anything.

Both Chomsky and Fodor allow for an innate ‘Language of Thought’ (henceforth LOT), and claim that this contains a set of semantic primitives which form the basis for all conceivable lexical meanings. If this assumption is made, then Chomsky has to allow that, even though this set of semantic primitives is not acquired, the child nonetheless has to acquire the phonological labels for concepts, since those labels are language-specific and arbitrary, varying from one language to another. At least some of what we might call phonological knowledge is thus, on Chomskyan assumptions, acquired, internalised from the ‘primary linguistic data’. The question then arises: are there any aspects of phonological knowledge that are innate, and thus not acquired, and if so, what are they, and are they a part of the specifically linguistic, species-specific language module?

Chomsky is surely right to argue that it is unquestionable that human beings are born with innate cognitive capacities: as Chomsky points out, the question is what they might be. Even Empiricist philosophers such as Locke [1689] did not deny that there are innate cognitive capacities: Locke’s objection was to innate conceptual content, such as Descartes’ [1642] ‘innate ideas’. Recall that Chomskyan innate linguistic knowledge is distinct in kind from general cognitive capacities, such as the capacities to (a) categorise, (b) form inductive generalisations, (c) form analogical generalisations, and (d) perceptually distinguish figure from ground. Chomsky does not deny that capacities such as these exist, but he downplays their rôle in the child’s linguistic development, and he must exclude all of these capacities from the ‘language organ’. It might also be argued that he must exclude all capacities from the language organ, since linguistic knowledge is said by Chomsky not to be a matter of knowing how to do something, or even knowing that something is the case: for Chomsky, linguistic knowledge is a cognitive state, distinct in kind from cognitive capacities to perform various cognitive tasks such as passing the Sally-Anne test, or passing conservation tests: pathologically normal infants initially fail these tests, but later come to pass them. The child’s developing capability of passing such tests does not, for Chomsky (as opposed to Piaget),
form part of the child’s specifically linguistic development. However, since there is so much talk of cognitive capacities in the literature, I will (reluctantly) allow that Chomsky and his followers may attribute certain cognitive capacities to the language module and I will insist that, for any capacity or form of knowledge to be located in the Chomskyan language module, it must be both specific to our species and specific to language (see [Carr, 2000] for discussion). Bearing these general points in mind, let us consider some of the capacities the child brings to bear on its phonetic and phonological development (I will return below to a possible distinction between phonetic development and phonological development).

Firstly, it is widely agreed that the child is born with certain innate perceptual capacities. For instance, as we have seen, human infants appear to be capable of discriminating aspirated and unaspirated voiceless stops, and voiced stops as in the syllables [pa], [pʰa] and [ba]. It is also known that chinchillas have this capacity. It cannot, therefore, be the case that this perceptual capacity forms a part of the language organ: it is a capacity that is neither specific to language nor specific to our species. It has not yet been established whether or not human infants are capable, at birth, of discriminating between all of the phonetic distinctions which can form the basis of phonological contrasts in human languages: it is entirely possible that infants initially fail to discriminate certain segmental contrasts (such as, say, [n] and [l]), and they may also fail, initially, to discriminate certain segmental sequences, such as, say, English [tɹiːz] and [tʃiːz] as in trees vs. cheese. If the child slowly comes to discriminate such differences, that aspect of the child’s development cannot be said to be driven by a Chomskyan language module.

Secondly, it is entirely plausible to suggest that children avail themselves of the capacity to form analogical generalisations: a ‘U-shaped’ developmental curve has often been attested in children’s development, in which the child begins by accurately uttering irregular forms such as went, and then overgeneralises to create forms such as goed, and then subsequently returns to the well-formed irregular forms. The stage at which the child overgeneralises is the stage at which the child is able to pass the wug test [Berko, 1958], in which the child is shown a picture of an unknown object, is given a made-up name for that object (such as wug) and the regular plural form (wugs, with a phonetically accurate allomorph) is elicited. These aspects of the child’s development cannot be driven by a Chomskyan language module, since the capacity to form analogical generalisations is not specific to language.

Thirdly, the child is capable of inductive generalisations, which are also plausibly said to underlie the child’s overgeneralisations. For instance, French children, after repeated exposure to forms such as the infinitive sortir (‘to go out’) and its participial form sorti, will often produce utterances such as J’ai couri très vite (‘I ran very quickly’), rather than the well-formed participle couru. It is very plausible to suggest that it is the repeated exposure to high-frequency forms such as sortir and sorti which underlies the child’s production, via inductive generalisation, of ill-formed participles such as couri. The capacity to form inductive generalisations is not restricted to our species, and inductive generalisations are not restricted to
language. Any aspect of the child’s linguistic development which relies on inductive generalisation cannot be driven by the putative language module. Note the rôle of token frequency in such phenomena: it is plausible to suggest that the infinitive and participial forms of verbs such as *sortir* are uttered frequently in the child’s environment, and this will help establish the inductive generalisation which gives rise to forms such as *couru*. Frequency effects such as this have no place in a Chomskyan conception of linguistic development, since such effects indicate that (aspects of) the input play a determining (rather than a merely triggering) rôle in the child’s development.

Fourthly, while mimicry cannot plausibly be the sole basis for the child’s linguistic development, there can be no doubt that it plays a rôle in the child’s capacity to utter adult-like words and phrases: a child who utters, say, [g\k] on exposure to adult utterances of [d\k] (*duck*) is plausibly said to be attempting to mimic the adult utterance. The capacity for mimicry must also be excluded from the Chomskyan language module, for several reasons. Firstly, it is the capacity to do something, so knowing how to imitate an adult does not constitute linguistic knowledge, for Chomsky. Secondly, mimicry is not limited to speech sounds: children will imitate the sounds of coffee grinders, ducks and dogs. Thirdly, mimicry is not specific to our species: several species of bird engage in mimicry (starlings and lyre birds, for instance).

It is often claimed (e.g. by Smith [2010]) that ‘knowledge of’ phonetic features, such as voicing or place of articulation, is innate. If, by this, Smith and others mean that the child can discriminate voiced and voiceless sounds from birth, that capacity, as noted above, cannot be a part of the Chomskyan language module: it is simply a part of our general perceptual capacities, and, as noted above, it is not restricted to our species. It is possible to distinguish between phonetic development and phonological development, but any attempt at such a distinction will run into the issue of whether phonology can be distinguished from phonetics, how the distinction can be drawn, and what the relation between the two might be. As we have seen, this is an area fraught with difficulties: there is simply no consensus in the literature on these matters. One might argue that the child’s developing articulatory and auditory perceptual capacities are a matter of purely phonetic capacities: if the child cannot initially produce, say, dental fricatives, that is a matter of an under-developed capacity to engage in the fine-grained motor control required to produce such speech sounds. It is possible, on Chomskyan assumptions, to argue that this purely phonetic development falls outside of the domain of strictly linguistic development. One could argue that this is distinct in kind from properly phonological development, such as the acquisition of the phonological contrasts and the phonological generalisations of the ambient language. If phonological development proper is taken to fall within linguistic development, then phonetic, but not phonological, development is non-linguistic. I conclude here that much, perhaps all, of the child’s phonetic/phonological development cannot be said to be driven by the putative ‘language organ’: much of this development involves capacities which are not specific to language; it also constitutes the mas-
tery of certain skills, both in production and in perception. Let us turn, therefore, to alternative accounts of the child’s development.

Sampson’s [1997; 2005] version of Empiricism is based on the work of Karl Popper ([1963] and elsewhere), notably the hypothetico-deductive method, which Popper claims is central to scientific reasoning. Stated briefly, the hypothetico-deductive method is the method whereby the scientist formulates falsifiable hypotheses, and then engages in deduction in order to see what is predicted by the hypothesis. The predictions are then tested in order to establish whether the hypothesis is falsified; if so, it is modified or abandoned. Sampson rejects Chomsky’s Rationalism in its entirety, and has painstakingly spelled out a critique of linguistic rationalism/nativism. For Sampson, there is no innate language module; the child learns the ambient language, including its phonology, and learns it the way children learn anything: using domain-general learning capacities, such as the capacities to form inductive and analogical generalisation, and the capacity to master certain tasks via repetition. Gopnik’s [2001] ‘theory theory’ of linguistic development is agnostic with respect to the Rationalist vs Empiricist divide: her aim is to establish a plausible version of the Sapir-Whorf Hypothesis with respect to the child’s conceptual and linguistic development. However, both Gopnik and Sampson claim that the child is formulating and testing (not necessarily consciously) hypotheses about the structure of the ambient language, including its phonological system. This is the view of ‘the child as a little scientist’; it is important for Sampson, since the capacities to form hypotheses and to engage in deduction (and thus testing) are not specific to language.

There are, it seems to me, inconsistencies here in both the Chomskyan Rationalist tradition and in Sampson’s Empiricism. If linguistic development is, as Chomsky claims, something that happens to the child, not something that the child does, then hypothesis formation can play no rôle in the child’s linguistic development, since forming and testing hypotheses amounts to actively doing something. As Smith [2004], who fully supports Chomsky’s Rationalism, puts it, ‘Learning of the sort that takes place in school or in the psychologist’s laboratory typically involves association, induction, conditioning, hypothesis formation and testing, generalisation and so on. None of these seems to be centrally involved in (first) language acquisition, which takes place before the child is capable of exploiting these procedures in other domains’ [Smith, 2004: 120-121]. Smith is hedging his bets here by using the term ‘centrally’, since he allows that there are areas of linguistic development where general learning capacities do play a restricted rôle, namely the production of the kinds of over-generalised forms cited above. But, as a Chomskyan, he is surely right to play down the rôle of general learning capacities: if these can be shown to play a central rôle, then Chomsky Rationalism is greatly weakened. The inconsistency I see in the Chomskyan tradition is this: that the putative innate language module (at the time referred to as UG) has been said at times by Chomskysians to constrain the range of linguistic hypotheses that the child entertains: Chomskynes have, in the past, conceived of the child as a little scientist: innate constraints were said to constrain the range of hypotheses that
the child would entertain during the course of linguistic development. Similarly, what Chomskyans call ‘the logical problem of language acquisition’ centres on the rôle of deduction: the child is said to deduce which value in a set of putatively innate, specifically linguistic, parameters is the right value for the ambient language. This sounds like hypothesis formation and deduction: it appears that the child is being said to hypothesise which is the right setting for each parameter.

The problem I see for Sampson is this. Being an Empiricist, he is bound to reject the idea of an innate LOT: for an Empiricist, the child comes to grasp concepts in interaction with the environment, via a process of general learning. The question then arises: if the child is formulating hypotheses, what are the hypotheses formulated in (and indeed, what are they hypotheses about)? There would have to be a conceptual vocabulary in place prior to the formulating of the hypotheses, and that vocabulary would have to be sufficiently rich to allow the formulating of all possible hypotheses. Such a conceptual vocabulary looks suspiciously like an LOT, as Fodor claims, but Sampson rejects the claim that the child is born with an LOT. Fodor’s claim that the child is born with access to the entire range of possible concepts is surely wildly implausible. But something akin to the LOT is remarkably difficult to get rid of, at least if we adopt the assumption that the child is formulating hypotheses.

Perhaps a half-way house can be established between Chomsky’s Rationalism and Sampson’s Empiricism: perhaps we can allow for innate capacities, some of which may be domain-specific (where one of the domains may relate to some aspect of language) and others domain-general, while allowing that the input plays a determining rôle in shaping the child’s linguistic development. This, I believe is a fair (partial) characterisation of the work of Annette Karmiloff-Smith ([1992] and elsewhere), whose ideas can be said to support a view of linguistic modularity as developmentally emergent linguistic modularity, as opposed to innate linguistic modularity. Karmiloff-Smith makes a subtle distinction between domain-relevant mechanisms and domain-specific mechanisms, allowing that the former may develop into the latter: ‘Unlike the domain-general theorist, this position does not argue for domain-general mechanisms simply applied across all domains. Rather, it suggests that biological constraints on the developing brain might have produced a number of mechanisms that do not start out as strictly domain-specific, that is, dedicated to the exclusive processing of one and only one kind of input. Instead, a mechanism starts out as somewhat more relevant to one kind of input over others, but it is usable — albeit in a less efficient way — for other types of processing too. This allows for compensatory processing and makes development channelled but far less predetermined than the nativist view. Once a domain-relevant mechanism is repeatedly used to process a certain type of input, it becomes domain-specific as a result of its developmental history’ [Karmiloff-Smith, 1998, reprinted 2001: 332-333]. If there were a mechanism that is best suited to processing sequences in the input, as distinct from a mechanism that is best suited to processing holistic input, in which the input is not broken down into its component parts, then the former would be well suited to the processing of sequences of speech sounds,
whereas the latter would be well suited to, for instance, the recognition of familiar faces (which relies on both an innate capacity to recognize face-like visual input, but which also requires training on specific faces).

Tomasello’s approach to the child’s linguistic development partially resembles Sampson’s empiricist approach, in that environmental input is not said to play a merely triggering rôle in development. But Tomasello stresses the child’s social-pragmatic interaction more than Sampson’s child-as-a-little scientist. For Tomasello, the child is acquiring linguistic symbols, and this acquisition is seen as ‘a kind of by-product of social interaction with adults, in much the same way that children acquire many other cultural conventions’ [Tomasello, 2001: 135]. Tomasello stresses the social world into which the child enters, a world full of structured social interactions. Central to Tomasello’s view is the rôle of intentions: ‘the child must always do some social cognitive work to determine the adult’s referential intentions’ [2001: 136]. Tomasello claims, entirely plausibly, that the child is not born knowing that other people have intentional relations towards the world: the child must come to appreciate that other people have intentions, and must develop the capacity to have intentions towards those intentions (which leaves open the question of whether the child must be said to be born with a capacity to enable this development). Tomasello differs from both Gopnik and Sampson in rejecting the picture of the child-as-little-scientist: rejecting Markman’s [1989] claim that there are innate constraints on what hypotheses the child-as-little-scientist will entertain, he claims that ‘word learning is not a hypothesis-testing procedure needing to be constrained at all, but rather it is a process of skill learning that builds upon a deep and pervasive understanding of other persons and their intentional actions (i.e. social cognition in general)’ [Tomasello, 1992: 148-149]. For Tomasello, when the child, during the second year of life, imitates adult behaviour, the child is becoming aware of the intentions behind the behaviour: the child is a mentalist, not a behaviourist. After the first year of life, the uttering of speech sounds by adults is taken by the child to be related to intentions, such as the intention to draw the child’s attention to something, and thus engage the child in acts of joint attention. Central to Tomasello’s view of the child’s linguistic development is the idea that the child is acquiring knowledge of linguistic symbols which are constituted as social conventions: I now turn to this idea that linguistic knowledge is knowledge of socially constituted norms/conventions.

Vihman’s ([1996; 2009] and elsewhere; see also [Vihman et al., 2009]) approach to the child’s linguistic development (more on which below) is radically anti-Chomskyan: she rejects the claim that humans are born with an innate language module. For Vihman, the child begins with no phonological system, but develops its own transitional production system during the course of development (a claim opposed by Smith [2010], who denies that the child has a system of its own). Smith [2010] regards child pronunciation errors as performance errors; for him, the child’s mentally stored phonological representations during the one-word stage are adult-like. Those representations are, for Smith, part of the child’s competence (in the Chomskyan sense): the child’s system is located entirely within
competence. For Smith, there are no production representations, distinct in form from the postulated adult-like representations. For Smith, the child’s phonological representations are accessible to conscious awareness [Smith, pc]. Vihman argues for a close-knit interconnection between perception and production in child phonology which, she argues, may have its neurological underpinnings in mirror neurons [Rizzolati & Arbib, 1998], neurons which fire when one engages in different kinds of articulatory acts, but which also fire when the child sees and hears others engaging in those same kinds of act. In studying what she takes to be the production systems of individual children during their transition to an adult-like phonology, Vihman stresses the rôle of individual words and the rôle played by the child’s attentional resources. The child will, Vihman suggests, focus its limited attentional resources on some selected aspects of the input: the child cannot focus its attention on all aspects of the input at the same time. On this view, the input, far from being impoverished, is so rich that the child is obliged, at any given stage in development, to pay attention to selected aspects of that input. One of the factors that will determine how this selection in perception works will, Vihman suggests, be the range of articulatory patterns that the child has mastered: the child is more likely to pay attention to, and attempt to articulate, patterns which form part of the child’s production repertoire at any given stage in development. The child will develop vocal motor schemes which it has mastered in production (say, consonant + a syllables) and will select from the input words which roughly correspond to such schemas (e.g. car, banana). The child will also modify its production of adult targets to make them fit with such schemas. For instance, a child which has mastered a CV[l]V sequences will adapt its productions of adult targets to force them into that mould; an example, drawn from Carr & Brulard’s [2003] work on their bilingual child Tom, is the production of the word cardy as [ka/li], with a CV[l]V template, found elsewhere in his repertoire, and a stress template with the stress on the final syllable, applicable to all of his French and English words during much of the one-word stage of production. These kinds of pattern can be remarkably systematic in child speech, which is what leads Vihman to postulate a production system. Note that, when Tom started to ‘fixed up’ his mis-stressed English word patterns, he did not engage in hyper-correction: forms such as [bi/to] (beetle) were changed to [bito], with the correct stress pattern, whereas forms such as the imperative [k∧m] (come in!), which we took to count as holistic one-word expressions during Tom’s one-word period, were not hyper-corrected to [k∧mm]. This suggests that, as far as Tom’s English word stress patterns were concerned, he had, at some level, adult-like representations, along the lines suggested by Smith. Smith’s argument against separate representations for articulatory productions is an argument from Occam’s Razor, but it seems to me that the sheer systematicity of much of the child’s transitory productions points towards a transitory production system. If we are to argue for an underlying set of representations that are shared with the community, and production systems which are specific to individual children in the early stages of development, then this could be taken to suggest an intersubjective (public) status for the underlying
system, but an individual status for the developing production systems, which is reminiscent of Saussure’s view that *langue* is ‘a social fact’ (depending on what ‘social’ means) whereas *parole* is individual. However, the production systems in a given community converge in the course of development, and thus become shared, intersubjective systems.

4 NORMATIVITY, UNCONSCIOUS KNOWLEDGE AND IMPLICIT LEARNING

Perhaps one of the most clearly worked-out conceptions of linguistic knowledge (understood in the generic sense of ‘linguistic’ and thus taken to subsume phonological knowledge) as knowledge of socially-constituted norms is that of philosopher and historian of linguistics Esa Itkonen [1978]. Itkonen distinguishes between spatiotemporal events such as a thunderstorm, which have no intentional basis, and actions, which are intentional, and carried out by agents. In addition to these, Itkonen postulates socially constituted norms. The distinction between events and actions leads Itkonen to distinguish between observable regularities such as the movement of waves in the sea and the spatiotemporal manifestation of social norms. Itkonen is an anti-reductionist: while he allows that actions have a spatiotemporal aspect, he denies that human actions can be reduced to spatiotemporal events. Itkonen argues that ‘It is possible to abstract from every action the intentional element which, properly speaking, constitutes and action *qua* action.’ And ‘intentions, which are necessary constituents of actions, must be at least potentially conscious: to do something, one must be able to know, at least under some description, what one is doing. Thus knowledge is, in principle, inseparable from action... knowledge is necessarily social’ [Itkonen, 1978: 122-123].

Itkonen’s conception of linguistic knowledge as knowledge of norms cannot be interpreted as a form of unconscious knowledge: the native speaker/hearer is said to know linguistic norms consciously. If Itkonen’s linguistic conventions (in the *generic* sense of ‘linguistic’) include phonological conventions, and if phonological knowledge is knowledge of socially-established conventions, then the question arises why speakers tend not to have conscious knowledge of those conventions. Let us consider some phenomena which might be describable in terms of phonological conventions. It is the norm for speakers of most varieties of English to utter dental fricatives (as in *thin*), but not front rounded vowels. Conversely, for speakers of most varieties of French, it is the norm to utter front rounded vowels (as in *lune*, *peu*, *soeur*) but not dental fricatives. A consequence of this is that most English speakers are not in the *habit* of uttering front rounded vowels when speaking their native language, while most French speakers are *not* in the habit of uttering dental fricatives. These conventions (assuming that they *are* conventions) concern the segmental inventories of particular languages. Other language-specific phonological phenomena are suprasegmental, such as Iambic Reversal in English, whereby, as we have seen, prominence levels are reversed in sequences of metrical feet. Consider the prominence levels in an expression such as ‘I’m going to
Piccadilly’, uttered with default intonation, in which ‘Piccadilly’ contains two metrical feet, the first less prominent than the second: ‘Picca’dilly. In the expression ‘I’m going to Piccadilly Circus’, uttered with default intonation, the stressed syllable in ‘Circus’ is more prominent than either of the stressed syllables of ‘Piccadilly’, but the prominence levels of the two metrical feet in ‘Piccadilly’ are reversed. If this kind of phenomenon, often analysed by phonologists, is describable as a social convention, found in English but not in other languages, why can speakers spend their lives adhering to the convention without having any conscious awareness of doing so?

As we have seen, Burton-Roberts [2000], unlike Itkonen, allows that there is an innate Chomskyan language module, but claims that it excludes phonological knowledge, which he conceives of as knowledge of phonological conventions, which he takes to be conventions of physical representation of the linguistic objects generated by the language module. Despite the differences between Itkonen and Burton-Roberts with respect to linguistic rationalism, both allow for socially-established conventions as playing a rôle in acts of uttering, even if, for Itkonen, the relevant conventions are linguistic, whereas those conventions are, for BR, not strictly linguistic in nature (since, on BR’s naturalistic interpretation of ‘linguistic’, they are said to be distinct in kind from the objects generated by the putative language module). For both Itkonen and BR, the child can be said to be internalising phonological conventions. A question which arises here is this: is it definitional of conventions that they are known consciously? Consciously-known social conventions certainly exist: I know that the convention for PhD orals in France is that the members of the jury stand up when the president of the jury formally delivers the candidate’s grade. That is because I have either noticed that this is how the grade is delivered, or have had it explained to me: in either case, knowledge of the convention involves conscious learning. Abiding by this kind of convention means adapting one’s behavior to local norms. A question for phonology is this: does adapting one’s behavior to local speech patterns constitute abiding by conventions which, unlike the French PhD jury case, are not consciously known? Do observed regularities in speech behaviour necessarily indicate the following of conventions?

Consider the remarkably structured phonological regularities often attested in the speech of individual infants, such as consonant harmony (CH). While some children exhibit little or no CH, those who do produce utterances with CH will exhibit different patterns: some children will target words of the syllabic shapes CVC (e.g. [g∧k] for duck) and CVCVC (e.g. [b∧kk] for bucket), but fail to target any words of the shape CVCV (such as Peggy). Other children will target words of all three syllable shapes. Some children will systematically replace coronal consonants (such as [t], [d], [s] and [z]) with labial consonants (such as [p] and [b], [f] and [v]); others will systematically do the reverse. What is striking is how systematically structured these patterns are for many children.

Such regularities, given that they frequently differ from one individual child to another, are statable in just the same way as the generalisations that are said to form part of the adult grammar. And yet, because they are unique to individ-
ual children, they cannot be conceived of as evidence for the internalisation of conventions, precisely because they are not intersubjective in nature. It is true that different individual children in (broadly) the same environment will eventually converge on pretty much the same grammar, but why should it be that the developmentally intermediate patterns should so closely resemble adult-like generalisations (whether conceived of in terms of rules, constraints, or both)? Why don’t we find mostly haphazard intermediate pronunciations, without any particular structure, unanalysable in terms of generalisations statable in very precise terms? Could it be that the phonological generalisations we formulate for adult speech are neither conventions of the sort envisaged by Burton-Roberts, Itkonen and Tomasello, nor internally-represented rules/constraints of the sort envisaged in generative phonology?

To explore these issues, let us return to the simple case of the wug test in English. If adult speakers of English are asked how many plural markers there are in English, they are likely to reply that there is only one (I ignore here irregular plurals such as oxen and children). Linguists have pointed out that there are three allomorphs of the English plural morpheme, and that their occurrence is entirely predictable, with the relevant generalisation statable in purely phonetic terms, as follows:

the \([iz]\) allomorph will occur if the stem-final consonant is one of the following: \([s]\) \([z]\), \([ʃ]\), \([ʒ]\), \([tʃ]\), \([dʒ]\], as in *horses, mazes, ashes, mirages, witches*, and *badges*. The triggering consonants form a natural class: it is the class of coronal fricatives and affricates.

Otherwise, if the final sound is voiceless, the voiceless allomorph will occur, as in *cats*, and if the final sound is voiced, the voiced allomorph will occur, as in *dogs* and *bees*.

The rules for the \([s]\) and \([z]\) allomorphs are phonetically natural, involving assimilation for voicing state. The retention of the historical vowel in the \([iz]\) cases is explicable in terms of the perceptual difficulty with potential forms such as *horses* as *hɔːsəs*. Adult speakers of English, unless they are linguists, are unaware that there are three allomorphs, and unaware that their occurrence is entirely predictable for reasons that are historically explicable in terms of perception and production. We need not be concerned here with the different analyses proposed by linguists to account for the phenomena: what matters, for our purposes, is that the phenomena exhibit structured phonetic regularity. The traditional generative approach to such regularities is to say that the speaker internalises the regularity stated above as a generalisation which forms part of the phonological component of a mind-internal grammar, whose contents are largely represented below the level of consciousness. That grammar will contain all the other phonological generalisations internalised

---

1I concede that consonant harmony for major place of articulation (Labial, Coronal, Dorsal) is unattested in adult phonologies, which exhibit only consonant harmony for minor place of articulation (bilabial, labio-dental, etc). My point is that systematic regularity involving syllabic shape of words and consonant harmony is attested in both child and adult phonologies.
by speakers of the language, from phonotactic constraints, through foot structure
generalisations, word stress assignment generalisations, rhythmic generalisations
such as Iambic Reversal, through to generalisations concerning the structure of
intonational phrases. For supporters of knowledge of generalisations as knowledge
of conventions, the speaker is also said to have internalised the generalisations, but
the internalisation in question is a matter of internalising something which is in-
herently intersubjective: a community-based norm/convention. This convention-
based approach is at odds with the Chomskyan notion ‘I-language’, according to
which linguistic knowledge (understood in the naturalistic sense) is individual and
internal, not intersubjective and thus external.

To further pursue the question of whether there can be unconscious knowledge
of social conventions, consider the literature on memory and implicit knowledge.
A distinction can be drawn between declarative and procedural memory. Neuro-
scientist Steven Rose argues that ‘remembering how and knowing that seem to
be different types of process, and patients suffering from amnesia are generally
well able to learn new skills-how knowledge-from riding a bike to doing a jigsaw,
although they have difficulty in remembering the actual experience of learning’
([Rose, 1992: 119-120]; italics in original). Note that Chomsky insists that lin-
guistic knowledge is neither procedural nor declarative: it is not knowing how nor
knowing that. Note too that procedural memory is also known as habit or skill
memory, and that such memory is very likely to be implicated in the acquisition of
articulatory skills in speech, resulting in the establishing of speech habits (whose
status was arguably hugely over-stated by the mid-twentieth century behaviourists,
but whose existence is surely beyond doubt). Rose claims that ‘declarative mem-
ory can be further subdivided into episodic and semantic memory’ [Rose, 1992:
120]. Episodic memory is memory of specific events, whereas semantic memory
is independent of those specific events: ‘My knowing that there was a European
war in 1939-1945 is in this sense semantic memory; remembering my experience
of wartime bombing is episodic’ [Rose, 1992: 120]. Work on amnesia seems to
show that the hippocampus is involved in the establishing of episodic memory.
Antonio Damasio, a neurophysiologist who works on consciousness and emotions,
reports on one of his cases: that of David, who, at the age of forty-six, was struck
by encephalitis, caused by the herpes simplex virus attacking the brain. This
caus ed damage in the left and right temporal lobes, including the hippocampus,
with the result that David was subsequently unable to learn any new facts, unable
to learn any new words, unable to recognise new faces, and also unable to recall
most old facts and familiar faces. Although David can access general notions, he
cannot access specific notions or recall individual facts, events or people. He has
a short-term memory span of around 45 seconds, but his long-term memory is
almost entirely inaccessible (or non-existent). Damasio notes that David ‘observes
a good number of social conventions, as shown in the polite manner with which he
greets others, takes turns in conversation, or walks about in a street or hallway’
[Damasio, 2000: 121]. David can also play draughts/checkers and win. If asked
what the name of the game is, he cannot recall it, and he is unable to state a
single rule of the game of draughts [Damasio, 2000: 117]. And yet, he can play the game with a pathologically normal person, and win. What are we to make of this phenomenon? The rules of draughts are very plausibly viewed as conventions, and pathologically normal humans can be said to consciously learn those conventions, which are intersubjectively established: there can be no such thing as my own unique, private rule for playing draughts, known only to me. Cases such as David’s suggest that implicit learning can take place without the conscious internalisation of the rules/conventions of the game. It therefore appears that, while knowledge of rules-as-conventions is often conscious, unconscious knowledge of rules-as-conventions is possible. Perhaps, in David’s case, the conventions in question were initially learned consciously and then subsequently stored unconsciously, thus allowing him to access that unconscious knowledge of conventions despite having no conscious access to the conventions. But, as Rose points out, amnesiacs can be trained to acquire new skills, with no conscious memory of the training sessions.

Vihman and Gathercole (ms; henceforth V&G) provide an overview of experimental findings with respect to implicit learning, which show that both children and adults automatically tally distributional regularities in the environment; this happens automatically, and reveals probabilistic, rather than categorical, learning (they conceive of the latter as symbol manipulation). V&G point out that this sensitivity to statistical regularities in the environment is not specific to human speech: it is a general capacity to automatically learn any regularly recurring sequences in the environment. Strikingly, this kind of implicit learning occurs without any intent to learn, and without any attention being paid to the patterns in question. V&G note that this capacity is operational prior to birth, in the last trimester of pregnancy, whereas declarative learning of words, involving reference, attention and intention, does not begin until the first half of the second year of life. Our capacity for registering statistical regularities in the environment goes some way to resolving the bootstrapping problem in child development. The bootstrapping problem can be formulated thus: if spoken word recognition involves a process of mapping words in the acoustic input to stored representations of words in the mental lexicon, how can the child engage in such mapping in the absence of a mental lexicon? How does the child get started? How can the child extract utterances of words from the stream of speech prior to having a mental lexicon? It appears that there are different statistical probabilities regarding sequences of segments within words, as opposed to sequences across word boundaries: given that the child is innately set up to tune in to such probabilities, the child is born with a capacity that will help with the segmentation of the continuous speech stream into word-like units.

Given that V&G also allow for conscious learning (word learning, for instance), they propose three stages of learning in child linguistic development: (a) the implicit tallying of regularities mentioned above, which relies on procedural memory, followed by (b) the conscious learning of lexical items, which relies on declarative, rather than procedural, memory and then (c) a ‘secondary’ process of procedural
induction over the regularities manifested in the lexical items which have been acquired. Regarding type (b) learning, they point out that ‘the registering and recall of arbitrary form-meaning pairs depends on processing in both frontal lobes (known to be involved in the selection of percepts for focussed attention) and the hippocampus, which alone is capable, in adults, of rapidly learning conjunctions of associated elements of experience’ (V&G ms). The registration of regularities, in stage (a) and stage (c) learning, they point out, can occur even when the hippocampus is damaged: there are two distinct learning/memory systems in the human brain. If this dual learning claim is correct, then it is possible to claim that the very earliest accommodation to the ambient language(s) involves procedural learning based on regularities in the acoustic input, and that at least some, perhaps all, of the rules postulated by phonologists to account for adult grammars reflect, not knowledge of rules-as-conventions, but implicit knowledge based on stage (c), in which the child has already begun to establish a mental lexicon, based on devoting attentional resources to the input, and procedural learning again (automatically) sets in, but operates, not over the ambient acoustic input, but over the representations in the mental lexicon, extracting regularities in that lexicon. This is what Karmiloff-Smith refers to as ‘re-representation’. If the entirety of adult phonological knowledge can be acquired using these two learning mechanisms, then no appeal to a Chomskyan innate language module is necessary, at least as far as phonological knowledge is concerned. The main point here is that type (c) knowledge can be conceived of as procedural knowledge of regularities across the mental lexicon, rather than knowledge of rules-as-conventions. Thus, the ability to pass, for example, the wug test reflects the stage at which the child has enough of a mental lexicon to extract, procedurally, and indeed inductively, the relevant regularities. The question then arises how those regularities come to be there, and why they are more-or-less uniform across a speech community. Port [2010: 45] suggests that ‘a community is a complex system that can create a practical phonology in historical time. Meanwhile the individual speakers come and go, learning whatever methods allow them to interpret others’ speech and to produce their own’. Note that Vihman’s two types of learning is not to be equated with the ‘dual mechanism hypothesis’ proposed by Pinker and others (e.g. [Pinker, 1991]): that hypothesis says that, regular and irregular forms, such as *come/came* vs *walk/walked* are dealt with differently in the mind: irregulars are simply stored in memory, whereas regular forms are created by a rule located in a separate component from the lexicon. I return to this below in my discussion of the work of Joan Bybee.

A point worth stressing here is that, if Vihman and her colleagues are right, the acquisition of phonology involves learning (albeit of different sorts), particularly inductive learning, driven by brain mechanisms that are nor specific to language: this view of phonological development could not be further removed from the Chomskyan view, assuming that Chomsky and his followers wish to include phonological knowledge within specifically linguistic knowledge, which appears to be the case, although Chomsky takes phonological knowledge to constitute an in-
put/output performance system within the putative language module. Note that Burton-Roberts, in some respects more Chomskyan than Chomsky himself, has no difficulty in accommodating such a picture of phonological acquisition to his claim that there is an innate language module since, for him, phonology is excluded from that innate module on the grounds of its being for the representation of, in phonetic substance, of the objects generated by that module (this is his ‘Representational Hypothesis’).

Further issues arise here. Let’s assume that Vihman’s postulated spiral-shaped developmental path is an accurate picture of the child’s phonological development. If so, then the same neural mechanism is responsible for both her stage (a) and her stage (c). Nonetheless, the epistemological status of the two stages may differ: the implicit knowledge yielded at stage (a) may be said not to be mentally represented, whereas the (equally implicit?) phonological knowledge yielded (equally automatically) at stage (c) may be said to be mentally represented: one and the same neural mechanism may be said to give rise to two qualitatively distinct epistemic outcomes. Related to this is the question of whether the explicit/implicit distinction may be said to map onto the conscious/unconscious distinction: are we to argue that the mental representations yielded by a procedural neural mechanism at stage (c) are available to consciousness? Perhaps not. What are we to make of the fact that speakers typically have no conscious awareness of sub-phonemic variation in their native language? Are we to say that they have conscious awareness of phonemic, but not sub-phonemic, representations? Or that sub-phonemic phenomena are not mentally represented? This brings us back to issues relating to the extent to which phonological representations in the mind are phonetically rich (see above on Port’s work), and also brings us back to the question of whether frequency in the input plays a central rôle in the establishing of phonological knowledge. I therefore end this chapter with discussion of the kind of usage-based approach to phonological knowledge advocated by Joan Bybee.

5 COMPETENCE/PERFORMANCE, USAGE-BASED PHONOLOGY AND FREQUENCY EFFECTS

Phonologist Joan Bybee has been arguing, for some decades now, against some of the central claims of generative phonology with respect to the nature of phonological knowledge. Bybee [2001] is a synthesis of her main claims, but see too Bybee [2007]. Bybee does not entirely reject the competence/performance distinction, or a mentalistic conception of phonological knowledge, but she objects to the exclusion of language use from such a conception. Language use, she argues, shapes phonological knowledge. Her approach is non-Chomskyan in the sense that she, unlike Chomsky, does not adopt a non-behavioural conception of linguistic knowledge (under which she subsumes phonological knowledge). But this does not necessarily mean that she is a behaviourist, in the sense used to describe the approach of Skinner to language. Unlike Chomsky, she stresses the role of the input, allowing that it plays a determinining, and not a merely triggering, rôle
in shaping linguistic knowledge. Like Port [2010], she adopts an exemplar-based model of phonological memory, and because of this, she (like Port) argues that stored phonological representations are not phoneme-like, stripped of all phonetic detail. Rather, they are rich in phonetic detail. She is not, however, a segmental eliminativist (‘all languages utilise consonants and vowels’: [Bybee, 2001: 191]), and she does not abandon the phonemic insight that we perceive phonetically distinct sound types as ‘instances of the same thing’. Given her usage-based approach to phonology, she takes phonological knowledge to be procedural, rather than declarative in nature. Since she stresses the role of the input, she argues that frequency effects play a major role in the shaping of phonological representations. Let us consider some of these, and their implications for the way we conceive of phonological knowledge.

Presupposing some version of the type/token distinction, Bybee distinguishes between token frequency and type frequency. Token frequency, as it relates to words, is ‘the frequency of occurrence of a unit, usually a word, in running text—how often a particular word comes up. Thus broke (the past tense of break) occurs 66 times per million words in Francis and Kucera [1982], while the past tense verb damaged occurs 5 time in the same corpus. In other words, the token frequency of broke is much higher than that of damaged’ [Bybee, 2001: 10]. Type frequency, on the other hand, concerns the frequency with which a specific kind of pattern occurs in the lexicon. The irregular past tense form broke has a much lower type frequency than regular past tense forms such as damaged. The pattern found in broke occurs only in a few items in the lexicon, such as spoke, rode and wrote. Bybee argues that historical phonetic change often progresses more quickly in items with high token frequency. For instance, elision of a syllable in sequences of schwa-plus-/r/ or /l/ is more common in words with high token frequency, such as every, camera, memory and family than it is in words with lower token frequency, such as mammary, artillery and homily. Her point is that ‘sound changes are the result of phonetic processes that apply in real time as words are used’ [Bybee, 2001: 11]. Bybee argues that mentally stored representations ‘change gradually, with each token of use having a potential effect on representation’: the adopting of an exemplar-based view of phonological representations is at the core of her conception of the nature of phonological knowledge. Another kind of frequency effect proposed by Bybee concerns resistance to historical analogical change: irregular forms which are high in token frequency will tend to resist analogical leveling. This kind of effect is offered as an explanation for why certain irregular forms have survived so long. Irregular past tense forms such as wept and crept have survived for centuries because they are used so often. Irregular forms which are low in token frequency will tend to be regularised or lost during the historical evolution of the language.

Type frequency, Bybee argues, has an effect on productivity. The high type frequency of the regular past tense suffix in English guarantees productivity: newly-coined verbs will be uttered with regular past tense suffixes. Thus, the relatively new denominal transitive phrasal verb to mike up (to set someone up with a portable microphone) will have, as its past tense form, miked up, and not mik
up, by analogy with irregulars like bite/bit. An important point made by Bybee is that her conception of the mental lexicon is less static than the conception found in most work in generative phonology. Her conception of the mental lexicon is a dynamic one, compatible with connectionist modeling, in which mentally stored words form ‘neighbourhoods’, based on different kinds of similarity, and contract neural net-type associations of different strength levels. Activation of a given mentally-stored word triggers activation of semantically or phonologically similar words. Generalisations are taken to be emergent: speakers can extract them from activation links in the mental lexicon. Take word stress patterns in English: a traditional generative approach to English word stress patterns would be to strip off any predictable stresses from mentally-stored underlying representations and to express those generalisations as phonological rules which are assigned to a phonological component, distinct from the lexicon. Bybee argues that each word is stored with its word stress pattern, and that the relevant word stress generalisations emerge from relations between those stored word representations. Bybee rejects Pinker’s dual mechanism hypothesis: for her, ‘regulars and irregular are handled by the same storage and processing mechanisms’. Her proposal is that ‘what determines whether or not a morphologically complex form is stored in memory is its frequency of use, not its classification as regular or irregular’ [Bybee, 2001: 110]. According to Bybee, ‘the high-frequency forms have storage in memory and low-frequency forms do not’. But it is difficult to see why low-frequency forms should not be stored in memory at all: surely even infrequent forms can be somehow stored in memory, given the vastness of the storage space.

Bybee abandons any strict separation of the lexicon and the syntax: frequently recurring sequences of words are, she claims, stored as constructions in the mental lexicon. These subsume more than the kinds of sequence traditionally referred to as idiom chunks in the generative literature. Idiom chunks such as to keep tabs on have often been taken to be lexicalised by generative syntacticians: they have been said to be stored as wholes in the mental lexicon. Bybee allows for many other sequences to be stored this way, and this has consequences for her analysis of certain phonological phenomena. Take obligatory liaison in French, as in sequences such as C’est un chien (‘It’s a dog’), where a ‘t’ is pronounced at the end of c’est (whereas, in an expression such as C’est chaud (‘It’s hot’), no ‘t’ is pronounced). A simple account of liaison would be to say that the liaison consonant is pronounced if the following word begins with a vowel (put another way: if the following word has an empty onset in the initial syllable). But things are not so simple: there are many cases where the following word begins with a vowel, but no liaison takes place. In the sentence Mes amis arrivent (‘My friends are coming’), the liaison consonant ‘z’ in the plural marker is pronounced between mes and amis, but not between amis and arrivent. Various attempts have been made to explain this in terms of syntactic structure, the suggestion being that there is somehow a ‘closer syntactic link’ between determiners and their following nouns than between subjects and their following verb phrases. None of these attempts have been particularly convincing. Bybee points out that frequency of
occurrence is the key to understanding the phenomenon: the sequence est un is high in spoken French: ‘In 47% of the uses of est, it occurs in the construction est un ‘is a’ + NOUN. That means the transitional probability of un after est is quite high (.47). In this sequence, liaison occurs 98.7% of the time, much more frequently than with any other uses of est, which strongly suggests a construction in which est t un is a constituent that precedes a noun’ [Bybee, 2001: 186].

The suggestion that sequences such as est t un are stored as units in the mental lexicon constitutes a radical departure from the way syntax and the lexicon have been modeled in much of the generative literature. Traditionally, items such as est, un and chien are seen as being stored in the lexicon, while sequences of these are constructed by the syntax. Additionally, most syntacticians would analyse sequences such as C’est un chien as containing un chien as a constituent, but not est un. Bybee claims that both sequences are constituents, and thus allows for overlapping syntactic constituency, which is a departure from the ‘no crossing branches’ tradition in syntax and phonology. Not only are the lexicon and syntax not strictly separated in Bybee’s approach: they are said to be subject to the same organisational principles, and those principles can have no place in a Chomskyan conception of universal linguistic principles.

While Bybee is a mentalist of some sort, she at times sounds as though she is adopting a version of behaviourism in which phonology is behaviour: ‘If we conceptualise phonology as part of the procedure for producing and understanding language, the phonological properties of language should result from the fact that it is a highly practised behaviour associated with the vocal tract of human beings’ [Bybee, 2001: 14]. A key notion here is the idea that what we produce when we utter is language, a view which conflates language and speech (see below for discussion of this topic). Another, closely related, key notion is the claim that language is constituted as behaviour, which appears not to be compatible with the mentalism inherent in Bybee’s talk of the mental lexicon. In a section entitled ‘Language as Part of Human Behaviour’, Bybee perhaps goes some way to resolving this tension by speaking of language in terms of both cognitive and social behaviour: if Bybee’s conception of the mental lexicon is based on both of these kinds of behaviour, then she cannot be accused of behaviourism in the sense applied to the work of Skinner, since she does not restrict the term ‘behaviour’ to that which is strictly observable (overt behaviour).

Bybee’s conception of linguistic universals differs from Chomsky’s: linguistic universals, for her, are driven by mechanisms of use, such as the tendency towards the reduction of articularatory gestures, the tendency to lenite and then elide word-final and syllable-final consonants, the formation of groups of similar units in the mental lexicon and ‘the formation of constructions out of frequently used structures’ [Bybee, 2001: 190]. These mechanisms, which may conflict with each other, cannot be conceived as forming part of a Chomskyan innate language module, since they relate to neuromotor control, and perceptual capacities which are almost certainly not specific to language, such as the capacity to categorise, and to make inferences. The view of language that emerges here is one in which lan-
guage is taken to be a self-organising system, where the structure of the system is a by-product of the cognitive mechanisms postulated by Bybee.

6 PHONOLOGY, INTERNALISM AND EXTERNALISM

We have considered more than one version of linguistic internalism, ranging from Burton-Roberts’ very radical internalism, which takes phonology to be behavioural, and thus non-linguistic in nature (where ‘linguistic’ means ‘pertaining to a wholly mind-internal, innate, entirely biological module of mind’), through Chomskyan internalism, to positions such as that of Vihman, in which a mind-internal lexicon is internalised from the environment, but no innate language module is postulated. Common to all of these positions is the idea that individual speakers possess mental representations of phonological forms; this is true even of phonologists, such as Hale & Reiss, who argue that phonology is substance-free. These various views also share some version of realism: phonological representations are to be interpreted realistically, rather than instrumentally, and the reality in question is mind-internal, not intersubjective. Some variety of internalist realism is assumed by most phonologists, but we have seen that there is an alternative view, which shares some features of Tomasello’s conception of linguistic symbols as social conventions.

Port’s conception of language as a social institution, which deserves serious consideration, is such a version of externalism, since it is embedded in the context of ‘distributed cognition’. Distributed cognition is a notion stemming from the work of philosopher Andy Clark [Clark, 1997; Clark & Chalmers, 1998], who argues that some kinds of knowledge can be conceived of as being distributed over a community of individuals, rather than being represented in individual brains. This is a major departure from internalism. The notion of distributed cognition has been applied to language by members of the Distributed Language Group (see [Cowley, 2007] for an overview). The ‘distributed’ view of language abandons the comparison of human cognition with the processing of symbols by computers, that is, it abandons that particular version of the computational theory of mind. Instead, human cognition is said to be based on on-line intelligence, ‘generated by complex causal interactions in an extended brain-body-environment system; these processes are grounded in biology; they can be modeled by dynamical systems theory’ [Cowley, 2007: footnote 1]. As Port [to appear] puts it, ‘the term ‘distributed’ emphasises that linguistic structures are not located in each individual, but are distributed across a population’. Port claims that ‘language cannot be separated from the rest of human behaviour without severe distortion’, that ‘language is just one aspect of the intense interpersonal co-ordination exhibited by humans.’ Port even goes as far as to claim that ‘no fundamental distinction can be made between linguistic conventions and other cultural conventions such as culture-specific gestures and facial expressions’. Even if we accept a central role for linguistic conventions in our conception of human language, and also accept that there are culture-specific gestures which may be regarded as conventional in nature, it is not clear that linguistic
expressions based on linguistic conventions are strictly parallel to culture-specific gestures. Gestures have no syntax, whereas the structures governed by syntactic conventions do, by definition. Take Smith’s [2004] example ‘I speak fluently English’ which is said to be ungrammatical, where ‘ungrammatical’, interpreted according to a conception of linguistic knowledge as knowledge of social norms, means ‘incorrect’, in the sense used by Itkonen [1978]. Compare this with the French sentence ‘Je parle couramment anglais’, which is well-formed. Adopting a view of language in which syntactic, morphological and phonological conventions are central, we can argue that it is a matter of conventionality that the English sequence is ill-formed, whereas the French sentence is well-formed. The conventions in question are to do with sequencing of discrete elements, but there is no sequencing of discrete elements in a gesture such as a shrug of the shoulders: even if gestures are conventional, the two kinds of convention are surely distinct.

According to Port, ‘For a newborn, language is clearly just part of the environment—something it hears and may have a special interest in. The child must learn to use the language, but does not need to represent it explicitly’. Here, Port departs radically from Bybee, who insists that the child does explicitly represent language structures in the mental lexicon. Note too that Port is declining to adopt a distinction between language and speech: speech is language, for Port. This is arguably unsustainable; let us consider why. The language vs speech distinction has had a rather tortured history (see [Burton-Roberts & Carr, 1999] for extended discussion), in which the two are often said to be distinct, but are nonetheless conflated. Some of the arguments for distinguishing the two are as follows. Firstly, deaf/mute users of sign language have language, but not speech, so a distinction between language and speech appears to be indicated. This argument can be countered by arguing that such users have, not speech, but manual signing, and that signing can be taken to constitute language for such users. A more telling argument for taking language to be distinct from speech is this: the events produced during acts of speaking are acoustic events. Such events have only acoustic properties: they do not have visual properties, tactile properties, or conceptual properties, for instance. It is a category mistake to attribute to acoustic events any properties other than acoustic properties. It is, arguably, equally an error to attribute to acoustic events syntactic properties such as, say, the relationship between a head and a dependent: acoustic events do not have heads or dependents. Additionally, if the stream of speech is said to contain words, which it would have to if speech is indeed language, then we are forced to claim that a given word, say cat, has acoustic properties such as loudness. But it makes no sense to ask ‘How loud is the word cat?’ We could ask ‘How loud was an utterance of the word cat on a particular occasion?’ But here we are speaking of utterances of words, a notion that needs unpacking. If the word cat is taken to constitute a Saussurean linguistic sign, then it constitutes an association of a mind-internal acoustic image and a mind-internal concept: neither of these can be heard, since neither constitutes an acoustic event. However we conceive of the relation between the postulated acoustic image and specific utterances of the word, those utterances do
not constitute the word itself. We are obliged to acknowledge that words are not speech events. If words are central to language, and words are not acoustic events, then language is not speech. If externalism involves claiming that language is constituted as speech, then that variety of externalism seems unsustainable. Perhaps the most plausible version of externalism is the view that languages are (or perhaps, contain) sets of social conventions which are, by definition, intersubjectively constituted. But even if one accepts this view, it is hard to get away from some version of (weak) internalism, since knowledge of social conventions is plausibly argued to be internalised knowledge.

If language and speech ought not to be conflated (are to be distinguished), then we are required to supply some conception of what the relation between the two might be. If that relation is one of realisation, and if phonology is included in language, that would seem to indicate that speech science is a distinct discipline from phonology, a view which many would oppose, particularly members of the Laboratory Phonology community (see [Pierrehumbert et al., 2000] for an overview). If the relation between language and speech is not one of realization, what might it be? One suggestion, made, as we have seen, by Burton-Roberts [2000; to appear] is that speech stands in a relation of representation to an innate syntactic-semantic system: it is the latter which BR calls language. This suggestion has the merit of supplying an explicit and coherent account of what the relation between language and speech is. However, it requires us to accept a very radical version of internalism, more radical than (though arguably more consistent than) that of Chomsky.

7 CONCLUDING REMARKS

One often encounters working phonologists who believe that issues in the conceptual foundations of phonology are of no concern to their work as phonological theorists, that such issues are a matter of ‘mere philosophy’, an optional extra which can be ignored while one gets on with the business of ‘doing phonological analysis’. This is misguided. The issues in the foundations of phonology are not purely conceptual: they are intimately intertwined with research into human neurophysiology (particularly the neural systems subtending different kinds of memory), human cognitive neuropsychology, psycholinguistics, child language acquisition and the analysis of adult phonologies. One cannot divorce the business of phonological analysis from the kinds of issues discussed above: as we have seen, the kinds of foundational assumption one makes will often determine the way one does phonology. The way that Morris Halle and other generative phonologists have done phonological theorising and analysis is rooted in their conception of the nature of linguistic knowledge. The way Joan Bybee does phonological analysis is different in crucial respects from mainstream generative phonology, precisely because her underlying conception of the nature of phonological knowledge differs from that of practitioners of generative phonology. Those underlying conceptions lead her to a different kind of model of the relationship between phonology, mor-
phology, the lexicon and syntax. If Silverman’s and Port’s foundational claims are valid, then we will have to engage in phonological analysis in a different manner. The way that Marilyn Vihman studies child phonology is distinct from the way that Neil Smith studies child phonology, precisely because Vihman and Smith differ in some of their basic assumptions about the nature of phonological knowledge. Of course, Vihman and Smith often gather data in similar ways, but the differences between their approaches are a matter of what they take the child to be doing, how they take the child to be developing. Phonologists have to be clear as to what they take the nature of phonological knowledge to be. In the discussion above, I hope to have offered a relatively useful overview of the issues, and the complex inter-relations between those issues. I hope to have given a sense of what has been emerging in recent years with respect to these issues: among other things, important conceptual developments in the work of Burton-Roberts, Hale & Reiss, Port, and Silverman; a new emphasis on the role of alphabetic literacy with respect to phonetic and phonological knowledge; more emphasis on the role of the input in the emergence of phonological phenomena; an increased interest in research on the neurology of memory, and its implications for how we conceive of phonological representations as psychological realities; the adopting of research results on different kinds of memory in relation to theories of child acquisition of phonology; and a renewed interest in social cognition and dynamic systems with respect to the child’s linguistic development.

ACKNOWLEDGEMENTS

I thank the following for useful discussion of some of the issues addressed here: John M. Anderson, Noel Burton-Roberts, Nigel Love, Bob Port, Neil Smith, and Marilyn Vihman. I am particularly grateful to Noel Burton-Roberts for extensive comments on the first draft.

BIBLIOGRAPHY


The Philosophy of Phonology


COMPUTATIONAL LEARNING THEORY AND LANGUAGE ACQUISITION

Alexander Clark and Shalom Lappin

1 INTRODUCTION

Computational learning theory explores the limits of learnability. Studying language acquisition from this perspective involves identifying classes of languages that are learnable from the available data, within the limits of time and computational resources available to the learner. Different models of learning can yield radically different learnability results, where these depend on the assumptions of the model about the nature of the learning process, and the data, time, and resources that learners have access to. To the extent that such assumptions accurately reflect human language learning, a model that invokes them can offer important insights into the formal properties of natural languages, and the way in which their representations might be efficiently acquired.

In this chapter we consider several computational learning models that have been applied to the language learning task. Some of these have yielded results that suggest that the class of natural languages cannot be efficiently learned from the primary linguistic data (PLD) available to children, through domain general methods of induction. Several linguists have used these results to motivate the claim that language acquisition requires a strong set of language specific learning biases, encoded in a biologically evolved language faculty that specifies the set of possible languages through a Universal Grammar.\(^1\)

In fact, when the assumptions underlying these models are carefully examined, we find that they involve highly implausible claims about the nature of human language learning, and the representation of the class of natural languages. Replacing these models with ones that correspond to a more realistic view of the human learning process greatly enhances the prospect for efficient language learning with domain general induction procedures, informed by comparatively weak language specific biases. Specifically, various procedures based on the ideas of distributional learning show that significant classes of languages can be learned.

\(^1\)For a discussion of the relevance of current work in computational learning theory to grammar induction, see [Clark and Lappin, 2010a]. For a detailed discussion of the connection between computational learning theory and linguistic nativism, see [Clark and Lappin, 2010b].
2 LINGUISTIC NATIVISM AND FORMAL MODELS OF LEARNING

The view that a set of strong language specific learning biases is a necessary condition for language acquisition can be described as linguistic nativism. This view has been endorsed by, *inter alia*, [Chomsky, 1965; Chomsky, 1975; Chomsky, 1981; Chomsky, 1995; Chomsky, 2000; Chomsky, 2005], [Crain and Pietroski, 2002], [Fodor and Crowther, 2002], [Niyogi and Berwick, 1996], [Nowak *et al.*, 2001], [Pinker, 1984], [Pinker and Jackendoff, 2005], and [Yang, 2002]. It has been dominant in linguistics and cognitive psychology for the past fifty years. One of the central motivations for this view is the claim that if children were equipped only with domain general learning procedures of the sort that they employ to achieve many kinds of non-linguistic knowledge, they would not be able to acquire the complex grammars that represent the linguistic competence of native speakers. The argument takes domain general inductive learning of grammar to be ruled out by limitations on the primary linguistic data (PLD) to which children are exposed, and restrictions on the resources of time and computation available to them. This view is commonly known as the *argument from the poverty of the stimulus* (APS).

There are several different versions of the APS, each of which focuses on a distinct aspect of the way in which the PLD underdetermines the linguistic knowledge that a mature native speaker of a language acquires. In this chapter we are concerned with the APS as a problem in formal learning theory, and we adopt the computational formulation of this argument given in [Clark and Lappin, 2010b].

\[(1)\]

a. Children acquire knowledge of natural language either through domain general learning algorithms or through procedures with strong language specific learning biases that encode the form of a possible grammar.

b. There are no domain general algorithms that could learn natural languages from the primary linguistic data.

c. Children do learn natural languages from primary linguistic data.

d. Therefore children use learning algorithms with strong language specific learning biases that encode the form of a possible grammar.

Some linguists and psychologists have invoked learning theoretic considerations to motivate this version of the APS. So [Wexler, 1999], apparently referring to some of [Gold, 1967]’s results, states that

The strongest most central arguments for innateness thus continue to be the arguments from APS and learnability theory. ... The basic results of the field include the demonstration that without serious constraints on the nature of human grammar, no possible learning mechanism can in fact learn the class of human grammars.

---

2See, for example, [Laurence and Margolis, 2001], [Pullum and Scholz, 2002], and [Crain and Pietroski, 2002] for alternative statements of the APS.
As we will see in Section 3, Gold’s results do not entail linguistic nativism. Moreover, his model is highly problematic if taken as a theory of human language learning.

At the other extreme, several linguists have insisted that learning theory has little, if anything of substance to contribute to our understanding of language acquisition. On their approach, we must rely entirely on the empirical insights of psychological and linguistic research in attempting to explain this process. So [Yang, 2008] maintains that

In any case, the fundamental problem in language acquisition remains empirical and linguistic, and I don’t see any obvious reason to believe that the solution lies in the learning model, be it probabilistic or otherwise.

We suggest that computational learning theory does not motivate strong linguistic nativism, nor is it irrelevant to the task of understanding language acquisition. It will not provide an explanation of this phenomenon. As Yang observes, it is not a substitute for a good psycholinguistic account of the facts. However, it can clarify the class of natural language representations that are efficiently learnable from the PLD. There are a number of important points to keep in mind when considering learning theory as a possible source of insight into language acquisition.

First, as we have already mentioned, a formal learning model is only as good as its basic assumptions concerning the nature of learning, the computational resources with which learners are endowed, and the data available to them. To the extent that these assumptions accurately reflect the situation of human language learners, the models are informative as mathematical and computational idealizations that indicate the limits of learning in that situation. If they significantly distort important aspects of the human learning context, then the results that they yield will be correspondingly unenlightening in what they tell us about the formal properties of acquisition.

Second, at least some advocates of the APS as an argument for linguistic nativism conflate learnability of the class of natural languages with learnability of a particular grammar formalism. While a formalism may indeed be unlearnable, given reasonable conditions on data, domain general induction procedures, and computational resources, this does not, in itself, show us anything about the learnability of the class of natural languages. In order to motivate an interesting unlearnability claim of the latter sort, it is necessary to show that the formalism in question (or a theory of grammar formulated in this formalism) is the best available representation of the class of natural languages. Establishing such a claim is exceedingly difficult, given that we have yet to achieve even a descriptively adequate grammar for a single language. In its absence, attempting to support the

---

3[Berwick and Chomsky, 2009] identify language acquisition with achieving knowledge of a transformational grammar of a particular kind. See [Clark and Lappin, 2010b], Chapter 2 for a critical discussion of this and other theory-internal instances of the APS.
APS on the grounds that a particular grammar formalism is unlearnable from the PLD is vacuous.

Third, it is has often been assumed that the class of natural languages must be identified either with one of the classes in the Chomsky hierarchy of formal languages, or with a class easily definable in terms of this hierarchy.\(^4\) In fact, there is no reason to accept this assumption. As we will see in subsequent sections, there are efficiently learnable classes of languages that run orthogonal to the elements of the Chomsky hierarchy (or are proper subsets of them), and which may be candidates for supersets of the class of natural languages.

Fourth, it is necessary to impose reasonable upper and lower bounds on the degree of difficulty that a learning model imposes on the language learning task. At the lower bound, we want to exclude learning models that trivialize the learning task by neglecting important limitations on the learning process. As we shall see, it is easy to construct models in which almost any class of languages is learnable. Such models are both inaccurate and unhelpful, because they do not constrain or guide our research in any way. At the upper bound we want to avoid theories on which learning is impossibly difficult. Given that humans do achieve the task we seek to model formally, our learning theory must allow for acquisition. If our model does not permit learning, then it is clearly false.

Finally, it is important to distinguish the hypothesis space from which a learning algorithm can select candidate representations of a language, from the class of languages that it can learn. The learning model imposes constraints that (partially) specify the latter class, but these do not prevent the algorithm from generating hypotheses that fall outside that class. Indeed in some cases it is impossible for the algorithm to restrict its hypotheses so that they lie inside the learnable class. It is also possible for such an algorithm to learn particular languages that are not elements of its learnable class, with particular data sets. Therefore, the class of learnable languages is generally a proper subset of the hypothesis space (hence of the set of representable languages) for a learning algorithm.

It follows that it is not necessary to incorporate a characterization of the learnable class into a language learner as a condition for its learning a specified class of languages. The design of the learner will limit it to the acquisition of a certain class, given data sets of a particular type. However, the design need not specify the learnable class, but only a hypothesis class that might be very much larger than this class.

Moreover, as the set of learnable languages for an algorithm may vary with its input data, this set corresponds to a relational property, rather than to a data invariant feature of the algorithm. In particular, in some models, as the amount of data increases, the class of languages that an algorithm can learn from that quantity of data will also expand. Therefore, only a range of learnable classes of languages, rather than a particular learnable class, can be regarded as intrinsic to the design of a learner.\(^5\)

---

\(^4\)See [Wintner, 2010] for a discussion of the Chomsky hierarchy within formal language theory.

\(^5\)See [Clark and Lappin, 2010b] Chapter 4, Section 7 for a detailed discussion of the relation
The tendency to reduce the hypothesis space of a learner to its learnable class runs through the history of the APS, as does the belief that human learners are innately restricted to a narrow class of learnable languages, independently of the PLD to which they are exposed. Neither claim is tenable from a learning theoretic perspective. To the extent that these claims lack independent motivation, they offer no basis for linguistic nativism.

We now turn to a discussion of classical models of learning theory and a critical examination of their defining assumptions. We start with Gold’s Identification in the Limit paradigm.

3 GOLD’S IDENTIFICATION IN THE LIMIT FRAMEWORK

We will take a language to be a set of strings, a subset of the set of all possible strings of finite length whose symbols are drawn from a finite alphabet Σ. We denote the set of all possible strings by Σ∗, and use L to refer to the subset. In keeping with standard practice, we think of the alphabet Σ as the set of words of a language, and the language as the set of all syntactically well-formed (grammatical) sentences. However, the formal results we discuss here apply even under different modeling assumptions. So, for example, we might consider Σ to be the set of phonemes of a natural language, and the language to be the set of strings that satisfy the phonotactic constraints of that language.

[Gold, 1967]’s identification in the limit (IIL) paradigm provides the first application of computational learning theory to the language learning task. In this paradigm a language L consists of a set of strings, and an infinite sequence of these strings is a presentation of L. The sequence can be written s1, s2, ..., and every string of a language must appear at least once in the presentation. The learner observes the strings of a presentation one at a time, and on the basis of this evidence, he/she must, at each step, propose a hypothesis for the identity of the language. Given the first string s1, the learner produces a hypothesis G1, in response to s2. He/she will, on the basis of s1 and s2, generate G2, and so on.

For a language L and a presentation of that language s1, s2, ..., the learner identifies in the limit the language L, iff there is some N such that for all n > N, Gn = GN, and GN is a correct representation of L. IIL requires that a learner converge on the correct representation GL of a language L in a finite but unbounded period of time, on the basis of an unbounded sequence of data samples, and, after constructing GL, he/she does not depart from it in response to subsequent data. A learner identifies in the limit the class of languages L iff the learner can identify in the limit every L ∈ L, for every presentation of strings in the alphabet Σ of L. Questions of learnability concern classes of languages, rather than individual elements of a class.

The strings in a presentation can be selected in any order, so the presentation between the hypothesis space and the learnable class of an algorithm, and for arguments showing why even the specification of the algorithm’s learnable class cannot be treated as part of its design.
can be arranged in a way that subverts learning. For example, the first string can recur an unbounded number of times before it is followed by other strings in the language. In order for a class to be learnable in the IIL, it must be possible to learn all of its elements on any presentation of their strings, including those that have been structured in an adversarial manner designed to frustrate learning.

Gold specifies several alternative models within the IIL framework. We will limit our discussion to two of these: the case where the learner receives positive evidence only, and the one where he/she receives both positive and negative evidence.

3.1 The Positive Evidence Only Model

In the positive evidence only variant of IIL presentations consist only of the strings in a language. Gold proves two positive learnability results for this model. Let a finite language be one which contains a finite number of strings. This class is clearly infinite, as there are an infinite number of finite subsets of the set of all strings. Gold shows that

(2) Gold Result 1:

The class of finite languages is identifiable in the limit on the basis of positive evidence only.

The proof of (2) is straightforward. Gold assumes a rote learning algorithm for this class of languages. When the learner sees a string in a presentation, he/she adds it to the set which specifies the representation of the language iff it has not appeared previously. At point $p_i$ in the presentation, the learner returns as his/her hypothesis $G_i$ = the set of all strings presented up to $p_i$. If $L$ has $k$ elements, then for any presentation of $L$, there is a finite point $p_N$ at which every element of $L$ has appeared at least once. At this point $G_N$ will be correct, and it will not change, as no new strings will occur in the presentation.

We can prove a second positive result in this model for any finite class of languages. In contrast to the class of finite languages, these classes have a finite number of languages, but may contain infinite languages. We will restrict ourselves throughout this chapter to recursive languages which are defined by the minimal condition that an effective decision procedure exists for deciding membership in the language for any string.

(3) Gold Result 2:

A finite class of recursive languages is identifiable in the limit on the basis of positive evidence only.

To prove (3) we invoke a less trivial algorithm than the rote learning procedure used to demonstrate (2). Assume that $L$ is a finite class of languages, and its elements are ordered by size, so that that if $L_i \subset L_j$, then $L_i$ occurs before $L_j$. Initially the learning algorithm $A$ has a list of all possible languages in $L$, and it returns the first element in that list compatible with the presentation. As $A$
observes each string $s_i$ in the presentation, it removes from the list all of the languages that do not contain $s_i$. Eventually it will remove all languages except the correct one $L$, and the languages that are supersets of $L$. Given the ordering of the list, $A$ returns $L$, the smallest member of the list that is compatible with the presentation, which is the correct hypothesis.

The best known and most influential Gold theorem for the positive evidence only model is a negative result for supra-finite classes of languages. Such a class contains all finite languages and at least one infinite language. Gold proves that

(4) **Gold Result 3:**

A supra-finite class of languages is not identifiable in the limit on the basis of positive evidence only.

The proof of (4) consists in generating a contradiction from the assumptions that (i) a class is supra-finite, and (ii) it can be learned in the limit. Take $\mathcal{L}$ to be a supra-finite class of languages, and let $L_{inf} \in \mathcal{L}$ be an infinite language. Suppose that there is an algorithm $A$ that can identify $L_{inf}$ in the limit. We construct a presentation on which $A$ fails to converge, which entails that there can be no such $A$.

Start with the string $s_1$, where $L_1 = \{s_1\}$ is one of the languages in $\mathcal{L}$. Repeat $s_1$ until $A$ starts to produce a representation for $L_1$ (the presentation will start $s_1, s_1, \ldots$). If $A$ never predicts $L_1$, then it will not identify $L_1$ in the limit, contrary to our assumption. If it does predict $L_1$, then start generating $s_2$ until it predicts the finite language $L_2 = \{s_1, s_2\}$. This procedure continues indefinitely, with the presentation $s_1, \ldots, s_2, \ldots, s_3, \ldots$. The number of repetitions of each $s_i$ is sufficiently large to insure that $A$ generates, at some point, the corresponding language $L_i = \{s_1, \ldots, s_i\}$. This presentation is of the language $L_{inf}$, which is infinite. But the algorithm will continue predicting ever larger finite subsets of $L_{inf}$ of the form $L_i$. Therefore, $A$ will never produce a representation for the infinite language $L_{inf}$.

Notice that we cannot use the algorithm $A$ that Gold employs to prove (3) in order to establish that a class of supra-finite languages is identifiable in the limit. This is because a supra-finite class contains the infinite set of all finite languages as a proper subset. If these are ordered in a list by size, and the infinite languages in the class are then ordered as successively larger supersets of the finite elements of this infinite class, then, for any given infinite language $L_{inf}$, $A$ will never finish identifying its infinite set of finite language subsets in the list to arrive at $L_{inf}$.

### 3.2 The Negative Evidence Model

In Gold’s negative evidence (informant) model, a presentation of a language $L$ contains the full set of strings $\Sigma^*$ generated by the alphabet $\Sigma$ of $L$, and each string is labeled for membership either in $L$, or in its complement $L'$. Therefore, the learner has access to negative evidence for all non-strings of $L$ in $\Sigma^*$. Gold proves that
(5) **Gold Result 4:**

The class of recursive languages is identifiable in the limit in the model in which the learner has access to both positive and negative evidence for each string in a presentation.

Gold proves (5) by specifying an algorithm that identifies in the limit the elements of this class. He takes the enumeration of the class to be an infinite list in which the representations of the language class are ordered without respect to size or computational power. At each point \( p_i \) in a presentation the algorithm returns the first representation of a language in the list that is compatible with the data observed up to \( p_i \). This data includes labels for all strings in the sequence \( p_1 \ldots p_i \). A representation \( G_i \) of a language is compatible with this sequence iff it labels its strings correctly.

The algorithm returns the first \( G_i \) in the list that is compatible with the data in the presentation. Because the presentation contains both the strings of the target language \( L \) and the non-strings generated by its alphabet, at some point \( p_j \) one of the data samples will rule out all representations in the list that precede \( G_L \), and all samples that follow \( p_j \) will be compatible with \( G_L \). Therefore, this algorithm will make only a finite number of errors. The upper bound on the errors that it can make for a presentation corresponds to the integer marking the position of the target representation in the ordered list.

Assume, for example, that \( L_{fs} \) is a finite state language which includes the strings of the context-free language \( L_{cf} \) as a proper subset. This is the case if \( L_{fs} = \{a^n b^n | n, m > 0 \} \) and \( L_{cf} = \{a^n b^n | n > 0 \} \). Let \( G_{fs} \) precede \( G_{cf} \) in the list of representations for the class. At some point in a presentation for \( L_{cf} \) a string labeled as not in the language will appear that is accepted by \( G_{fs} \). As a result, the algorithm will discard \( G_{fs} \), and, by the same process, all other elements of the list, until it arrives at \( G_{cf} \). After this point all data samples will be labeled in accordance with \( G_{cf} \), and so the algorithm will return it. If only positive evidence were contained in the presentation of \( L_{cf} \), all of the data samples would be compatible with \( G_{fs} \), and the algorithm would not be able to identify \( G_{cf} \) in the limit.

The class of recursive languages includes the class of context-sensitive languages as a proper subset. To date no natural language has been discovered whose formal syntactic properties exhibit more than context-sensitive resources, and so it seems reasonable to conjecture that natural languages constitute a proper subset of this latter class. Therefore, (5) implies that, with negative evidence for all strings in a language, any natural language can be identified in the limit by the simple learning algorithm that Gold describes.

The negative evidence variant of IIL is an instance in which learning is trivialized by an excessively powerful assumption concerning the sort of evidence that is available to the learner. It is clear that the PLD to which children are exposed does not consist of sentence-label pairs in which every string constructed from the alphabet of the language is identified as grammatical or as ill formed. Whether or not negative evidence of any kind plays a significant role in language acquisition
remains a highly controversial issue in psycholinguistics.\textsuperscript{6} Even if we assume that certain types of negative evidence are available, it is clear that Gold’s full informant model of IIL does not offer a plausible view of the PLD that provides the basis for human language acquisition.

3.3 The Positive Evidence Only Model and Learning Biases

Some linguists have used Gold’s proof that a supra-finite class of languages is not identifiable in the limit as grounds for positing a rich set of prior constraints on the human language learning mechanism. So, for example, [Matthews, 1989] states

[pp 59-60] The significance of Gold’s result becomes apparent if one considers that (i) empiricists assume that there are no constraints on the class of possible natural languages (…), and (ii) Gold’s result assumes that the learner employs a maximally powerful learning strategy (…). These two facts … effectively dispose of the empiricist claim that there exists a “discovery procedure” capable of discovering a grammar for any natural language solely by analyzing a text of that language. This claim can be salvaged but only at the price of abandoning the empiricist program, since one must abandon the assumption that the class of possible languages is relatively unconstrained.

Advocates of linguistic nativism go on to insist that these learning biases must specify the hypothesis space of possible natural languages, and determine a task particular algorithm for selecting elements from this space for given PLD, as necessary conditions for language acquisition. [Nowak \textit{et al.}, 2001] claim the following.

Universal grammar consists of (i) a mechanism to generate a search space for all candidate mental grammars and (ii) a learning procedure that specifies how to evaluate the sample sentences. Universal grammar is not learned but is required for language learning. It is innate.

In fact, these conclusions are not well motivated. They depend upon assumptions that are open to serious challenge. First, Gold’s negative result concerning supra-finite languages is significant for language acquisition only if one assumes that the class of natural languages is supra-finite, as are the language classes of the Chomsky hierarchy. This need not be the case. A set of languages can be a proper subset of one these classes such that it is a finite class containing infinite languages. In this case, it is not supra-finite, but it is identifiable in the limit. Moreover, it may contain representations that converge on the grammars of natural language.

So, for example, [Clark and Eyraud, 2007] define the class of substitutable languages, which is a proper subset of the class of context free languages. The

\textsuperscript{6}See [Clark and Lappin, 2010b], Chapter 3, Section 3.2 for detailed discussion of this issue, as well as Chapter 6 for a proposed stochastic model of indirect negative evidence.
grammars of these languages can generate and recognize complex syntactic structures, like relative clauses and polar interrogative questions. [Clark and Eyraud, 2007] specify a simple algorithm for learning substitutable languages from well formed strings (positive data only). They show that the algorithm identifies in the limit the class of substitutable languages in time polynomial to the required data samples, from a number of samples polynomially bounded by the size of the grammar.

Second, Gold's positive evidence only version of IIL is not a plausible framework for modeling human language acquisition. It is both too demanding of the learner, and too permissive of the resources that it allows him/her. Its excessive rigor consists in the condition that for a class to be identifiable in the limit, all of its elements must be learnable under every presentation. Therefore, learning is required even when a data presentation is designed in an adversarial mode to sabotage learning. As Gold notes, if we discard this condition and restrict the set of possible presentations to those that promote learning, then we can significantly expand the class of learnable languages, even in the positive evidence only model. Children are not generally subjected to adversarial data conditions, and if they are, learning can be seriously impaired. Therefore, there is no reason to demand learning under every presentation.

Conversely, IIL allows learners unbounded amounts of computational complexity in time and data samples. Identification need only be achieved in the limit, at some bounded point in a presentation. This feature of Gold's framework is unrealistic, given that humans learn under serious restrictions in time, data, and computational power. In order to approximate the human learning process, we need to require that learning be efficient.

Third, as we noted in Section 2, the hypothesis space for a learning algorithm cannot be reduced to the class of representations that it can learn. A grammar induction procedure can generate hypotheses that represent languages outside of its learnable class. It may even learn such languages on particular presentations, but not on all of them.

Finally, the positive evidence only IIL paradigm is too restrictive in requiring exact identification of the target language. Convergence on a particular adult grammar is rarely, if ever, complete. A more realistic approach characterizes learning as a process of probabilistic inference in which the learner attempts to maximize the likelihood of a hypothesis, given the data that it is intended to cover, while seeking to minimize its error rate for this data. We will consider probabilistic learning theories in the next two sections.

\footnote{Impairment of learning due to an absence of data is particularly clear in the case of feral children, who are deprived of normal linguistic interaction. Perhaps the best known case of such a child is Genie, discussed in [Curtiss, 1977].}
4 PROBABILISTIC MODELS AND REALISTIC ASSUMPTIONS ABOUT HUMAN LEARNING

One of the limitations of the Gold model is that the learner must identify the target under every possible presentation. Therefore, he/she is required to succeed even when the sequence of examples is selected in order to make the learning task as difficult as possible, i.e. even when the teacher is an adversary who is trying to make the learner fail. This is a completely unrealistic view of learning. In the human acquisition process adults generate sentences in the child’s environment with an interest, in most cases, in facilitating child learning.

A consequence of the IIL is that it is difficult for the learner to tell when a string is not in the language. Absence of evidence in this model is not evidence of absence from the language. The fact that the learner has not seen a particular string does not permit him/her to conclude that that string is ill formed. No matter how short a string is, nor how long the learner waits for it, its non-occurrence could be due to the teacher delaying its appearance, rather than ungrammaticality. It is for this reason that, as we have seen, the presence or absence of negative data has such a significant effect on the classes of languages that can be learned within the IIL framework (see [Clark and Lappin, 2010b] Chapters 3 and 6 for extensive discussion of these issues).

Linguists have been mesmerized by this property of IIL, and they have frequently taken the absence of large amounts of direct negative evidence to be the central fact about language acquisition that motivates the APS ([Hornstein and Lightfoot, 1981] characterize this issue as the “logical problem of language acquisition”). It is worth noting that it is only in linguistics that the putative absence of negative evidence is considered to be a problem. In other areas of learning it has long been recognised that this is not a particular difficulty. The importance that many linguists assign to negative evidence (more specifically its absence) arises largely because of an unrealistic assumption of the IIL paradigm [Johnson, 2004]. From very early on, learning theorists realised that in a more plausible model a learner could infer, from the absence of a particular set of examples, that a grammar should not include some sentences.


A not unreasonable acquisition system can be devised with the operative principle that if certain structures or rules fail to be exemplified in relatively simple expressions, where they would expect to be found, then a (possibly marked) option is selected excluding them in the grammar, so that a kind of “negative evidence” can be available even without corrections, adverse reactions etc.

This sort of data has traditionally been called “Indirect Negative Evidence”. The most natural way to formalise the concept of indirect negative evidence is with probability theory. Under reasonable assumptions, which we discuss below, we can infer from the non-occurrence of a particular sentence in the data that the
probability of its being grammatical is very low. It may be that the reason that we have not seen a given example is that we have just been unlucky. The string could actually have quite high probability, but by chance we have not seen it. In fact, it is easy to prove that the likelihood of this situation decreases very rapidly to insignificance. But much more needs to be said. Clearly there are technical problems involved in specifying the relationship between probability of occurrence and grammaticality. First, there are an indefinite number of ungrammatical strings and it is not clear how the learner could keep track of all of these, given his/her limited computational resources.

Second, there are ungrammatical strings that do occur in the PLD. Suppose we have an ungrammatical string with a non-zero probability, say $\epsilon$. Since there are, in most cases, an infinite number of strings in the language, there must be some strings that have probability less than $\epsilon$. In fact, all but finitely many strings will have probability less than $\epsilon$. This leads to the inconvenient fact that the probability of some long grammatical strings will be less than the probability of short ungrammatical ones. Therefore it is clear that we can not simply reduce grammaticality to a particular probability bound.

Returning to the IIL, rather than assuming that the teacher is antagonistic, it seems natural to identify a proper subset as typical or helpful example sequences and require the learner to succeed only on these. It turns out to be difficult to construct a non-trivial model of non-adversarial learning [Goldman and Mathias, 1996]. A more realistic approach is to assume that the data has a probabilistic (random) dimension to it. There is much current interest in probabilistic models of language [Bod et al., 2003]. We remain neutral as to whether linguistic competence itself should be modeled probabilistically, or categorically as a grammar, with probabilities incorporated into the performance component. Here we are concerned with probabilistic properties of the input data and the learning process, rather than the target that is acquired.

If we move to a probabilistic learning paradigm, then the problem of negative evidence largely disappears. The most basic form of probabilistic learning is Maximum Likelihood Estimation (MLE), where we select the model (or set of parameters for a model) that makes the data most likely. When a fixed set of data $D$ (which here corresponds to a sequence of grammatical sentences) is given, the learner chooses an element, from a restricted set of models, that maximises the probability of the data, given that model (this probability value is the likelihood of the model). The MLE approach has an important effect. The smaller the set of strings that the model generates, while still including the data, the higher is its likelihood for that data. To take a trivial example, suppose that there are 5 types of sentences that we could observe, and we see only three of them. A model that assigns a probability of $1/3$ to each of the three types that we encounter, and zero probability to the two unseen types, will have higher likelihood than one which gives $1/5$ to each of the 5 types. This example illustrates the obvious fact that we do not need explicit negative data to learn that some types do not occur (a point developed more compellingly and more thoroughly in, inter alia, [Abney, 1996;
When we are concerned with cases, as in language acquisition, where there are an unbounded or infinite number of sentence types, it is important to limit the class of models that we can select from. There are many closely related techniques for doing this (like Bayesian model selection and Minimum Description Length), where these techniques enjoy different levels of theoretical support. They all share a common insight. We need to consider not just the likelihood of the model given the data, but we must also take into account the model’s size and complexity. Larger and more complex models have to be justified by additional empirical coverage [Goldsmith, 2001].

In statistical modeling it is standard to regard the data as independently and identically distributed. This is the IID assumption. It entails that for language acquisition there is a fixed distribution over sentences, and each sentence is chosen randomly from this distribution, with no dependency on the previous example. This claim is clearly false. The distribution of examples does change over time. The relative probabilities of hearing “Good Morning” and “Good Night” depend on the time of day, and there are numerous important inter-sentential dependencies, such as question answer pairs in dialogue.

Many linguists find the IID objectionable for these reasons. In fact, we can defend the IID as an idealization that approximates the facts over large quantities of data. All we need is for the law of large numbers to hold so that the frequency of occurrence of a string will converge to its expected value rapidly. If this is the case, then the effect of the local dependencies among sentences in discourse will be eliminated as the size of the data sample increases. This view of the IID offers a much weaker understanding of the independence conditions than the claim that the sentences of a distribution are generated in full independence of each other. It is a view that applies to a large class of stochastic processes.

Moreover if we can prove learnability under the IID assumption, then we can prove learnability under any other reasonable set of assumptions concerning the distributions of the data as well. Therefore, if we are modeling the acquisition of syntax (i.e. intra-sentential structure), then it is reasonable to neglect the role of inter-sentential dependencies (at least initially). We assume then that there is a fixed distribution. For each string we have a probability. The distribution is just the set of probabilities for all strings in a data set, more accurately, a function that assigns a probability to each string in the set.

To avoid confusion we note that in this chapter we use the word distribution in two entirely different senses. In this section a distribution is a probability distribution over the set of all strings, a function $D$ from $\Sigma^* \rightarrow [0,1]$, such that the sum over all string of $D$ is equal to 1. In later sections we use distribution in the linguistic sense to refer to the set of environments in which a string can occur.

There are a number of standard models of probabilistic learning that are used in machine learning. The best known of these is the PAC-learning paradigm [Valiant, 1984], where ‘PAC’ stands for Probably and Approximately Correct. The paradigm recognises the fact that if data is selected randomly, then success in learning is
random. On occasion the random data that you receive will be inadequate for learning. Unlike the case in IIL, in the PAC framework the learner is not required to learn the target language exactly, but to converge to it probabilistically. This aspect of the paradigm seems particularly well-suited to the task of language learning, but some of its other features rule it out as an appropriate framework for modeling acquisition.

PAC models study learning from labeled data in which each data point is marked for membership or non-membership in the target language. The problem here is, of course, the fact that few, if any, sentences in the PLD are explicitly marked for grammaticality.

A second difficulty is that PAC results rely on the assumption that learning must be (uniformly) possible for all probability distributions over the data. On this assumption, although there is a single fixed distribution, it could be any one in the set of possible distributions. This property of PAC-learning entails that no information can be extracted from the actual probability values assigned to the strings of a language. Any language can receive any probability distribution, and so the primary informational burden of the data is concentrated in the labeling of the strings. The actual human learning context inverts this state of affairs. The data arrives unlabeled, and the primary source of the information that supports learning is the probability distribution that is assigned to the observed strings of the PLD. Therefore, despite its importance in learning theory and the elegance of its formal results, the classical version of PAC-learning has no direct application to the acquisition task. However PAC’s convergence measure will be a useful element of a more realistic model.

If we consider further the properties of learnability in the PAC paradigm, we encounter additional problems. A class is PAC learnable if and only if it has a finite VC-dimension, where its VC-dimension is a combinatorial property of the class (see [Lappin and Shieber, 2007] and [Clark and Lappin, 2010b], Chapter 5 for characterizations of VC-dimension and discussions of its significance for language learning in the PAC framework). A finite class of languages has finite VC-dimension, and so one way of achieving PAC learnability is to impose a cardinality bound on the target class. So, for example, we might limit the target class to the set of all context-sensitive languages whose description length, when written down, is less than some constant $n$, the class $CS_n$. The class of all context-sensitive languages $CS$ has infinite VC-dimension, but we can consider it as the union of a gradually increasing set of classes, $CS = \bigcup_n CS_n$. On the basis of this property of PAC-learning one might be tempted to argue along the following lines for a strong learning bias in language acquisition. As $CS$ has infinite VC-dimension it is not learnable. Therefore the class of languages must be restricted to a member of the set of $CS_n$s for some $n$. It follows that language learners must have prior knowledge of the bound $n$ in order to restrict the hypothesis space for grammar induction to the set of $CS_n$s.

This argument is unsound. In fact a standard result of computational learning theory shows that the learner does not need to know the cardinality bound of the
target class. [Haussler et al., 1991]. As the amount of available data increases, the learner can gradually expand the set of hypotheses that he/she considers. If the target is in the class $CS_n$, then the learner will start to consider hypotheses of size $n$ when he/she has access to a sufficiently large amount of data. The size of the hypotheses that he/she constructs grows in proportion to the amount of data he/she observes. A prior cardinality restriction on the hypothesis space is unnecessary.

This point becomes clear when we replace $CS$ with the class of finite languages represented as a list, $FIN$. A trivial rote learning algorithm can converge on this class by memorising each observed example for any of its elements. This procedure will learn every element of $FIN$ without requiring prior information on the upper bound for the size of a target language, though $FIN$ has unbounded VC-dimension.

More appropriate learning models yield positive results that show that large classes of languages can be learned, if we restrict the distribution for a language in a reasonable way. One influential line of work looks at the learnability of distributions. On this approach what is learned is not the language itself, but rather the distribution of examples (i.e., a stochastic language model).

[Angluin, 1988] and [Chater and Vitányi, 2007] extend [Horning, 1969]'s early work on probabilistic grammatical inference. Their results show that, if we set aside issues of computational complexity, and restrict the set of distributions appropriately, then it is possible to learn classes of grammars that are large enough to include the set of natural languages as a subclass.

As [Angluin, 1988] says

These results suggest the presence of probabilistic data largely compensates for the absence of negative data.

[Angluin, 1988] also considers the learnability of languages under a stochastic version of IIL. She shows, somewhat surprisingly, that Gold's negative results remain in force even in this revised framework. Specifically, she demonstrates that any presentation on which an IIL learner fails can be converted into a special distribution under which a stochastic learner will also not succeed. This result clearly indicates the importance of selecting a realistic set of distributions under which learning is expected. If we require learning even when a distribution is perverse and designed to sabotage acquisition, then we end up with a stochastic learning paradigm that is as implausible as IIL.

The negative results that we derive from either the IIL paradigm or from PAC-learning suffer from an additional important flaw. They do not give us any guide to the class of representations that we should use for the target class, nor do they offer insight into the sort of algorithms that can learn such representations. This is not surprising. Although IIL was originally proposed as a formal model of language acquisition, it quickly became apparent that the framework applies more generally to the task of learning any collection of infinitely many objects. The inductive inference community focuses on learnability of sets of numbers,
rather than on sets of strings. Similarly PAC-learning is relevant to every domain of supervised learning. Since these frameworks are not designed specifically for language acquisition, it is to be expected that they have very limited relevance to the construction of a language learning model.

5 COMPUTATIONAL COMPLEXITY AND EFFICIENCY IN LANGUAGE ACQUISITION

An important constraint on the learner that we have not yet considered is computational complexity. The child learner has limited computational resources and time (a few years) with which to learn his/her language. These conditions impose serious restrictions on the algorithms that the learner can use. These restrictions apply not just to language acquisition, but to other cognitive processes. The Tractable Cognition Thesis [van Rooij, 2008] is uncontroversial.

Human cognitive capacities are constrained by the fact that humans are finite systems with limited resources for computation.

However, it is not obvious which measure of complexity provides the most appropriate standard for assessing tractability in human computation. Putting aside for a moment the problem of how to formulate the tractability thesis precisely for language acquisition, its consequences are clear. An algorithm that violates this thesis should be rejected as empirically unsound. An inefficient algorithm corresponds to a processing method that a child cannot use, as it requires the ability to perform unrealistic amounts of computation.

It is standard in both computer science and cognitive science to characterise efficient computation as a procedure in which the amount of processing required increases relatively slowly in relation to the growth of an input for a given task. A procedure is generally regarded as tractable if it is bounded by a polynomial function on the size of its input, for the worst processing case. This condition expresses the requirement that computation grow slowly in proportion to the expansion of data, so that it is possible to solve large problems within reasonable limits of time. If the amount of processing that an algorithm \( A \) performs grows very rapidly, by an exponential function on the size of the data, then as the input expands it quickly becomes impossible for \( A \) to compute a result.

Therefore, we can rule out the possibility that child learners use procedures of exponential complexity. Any theory that requires such a procedure for learning is false, and we can set it aside.\(^8\)

We consider the tractability condition to be the most important requirement for a viable computational model of language acquisition to satisfy. The problems

\(^8\)There are a number of technical problems to do with formalising the idea of efficient computation in this context. For instance, the number data samples that the learner is exposed to increases, and the length of each sample is potentially unbounded. There is no point to restricting the quantity of data that we use at each step in the algorithm, unless we also limit the total size of the data set, and the length of each sample in it.
involved in efficient construction of a target representation of a language are more substantial than those posed by achieving access to adequate amounts of data. Efficiency of learning is a very hard problem, and it arises in all learning models, whether or not negative evidence is available.

The computational complexity of learning problems emerges with the least powerful formalisms in the Chomsky hierarchy, the regular languages, and so the more powerful formalisms, like the class of context free (or context sensitive) grammars also suffer from them. These difficulties concern properties of target representations, rather than the language classes as such. It is possible to circumvent some of them by switching to alternative representations which have more tractable learning properties. We will explore this issue in the next section.

There are a number of negative results concerning computational complexity of learning that we will address. Before we do so, we need to register a caveat. All of these results rest on an assumption that a certain class of problem is intrinsically hard to solve. These assumptions, including the famous \( P \neq NP \) thesis, are generally held to be true. The results also rely on additional, more obscure presuppositions (such as factoring Blum integers etc.). But these assumptions are not, themselves, proven results, and so we cannot exclude the possibility that efficient algorithms can be devised for at least some of the problems now generally regarded as intractable, although this seems highly unlikely.

The most significant negative complexity results [Gold, 1978; Angluin and Kharitonov, 1991; Abe and Warmuth, 1992; Kearns and Valiant, 1994] show that hard problems can be embedded in the hidden structure of a representation. In particular the results given in [Kearns and Valiant, 1994] indicate that cryptographically hard problems arise in learning even very simple automata. They entail that the complexity of learning representations is as difficult as code cracking. This suggests that the framework within which these results are obtained does not adequately model human learning. It should distinguish between the supportive environment in which child learners acquire grammar, and the adversarial nature of the code-breaking task. The codes are designed to maximize the difficulty of decryption, while natural languages facilitate acquisition and transmission.

Parametric theories of UG encounter the same complexity issues that other learning models do. Assuming that the hypothesis space of possible grammars is finite does not address the learnability issue. In fact, the proofs of the major negative complexity of learning results proceed by defining a series of finitely parameterised sets of grammars, and demonstrating that they are difficult to learn. Therefore, Principles and Parameters (P&P) based models do not solve the complexity problem at the core of the language acquisition task. Some finite hypothesis spaces are efficiently learnable, while others are not. The view that UG consists of a rich set of innate, language specific learning biases that render acquisition tractable contributes nothing of substance to resolving the learning complexity problem, unless a detailed learning model is specified for which efficient learning can be shown. To date, no such model has been offered.
Alexander Clark and Shalom Lappin

Figure 1. Two clustering problems. On the left the three clusters are well separated and the problem is easy, and on the right they are not, and the problem is hard.

It is important to recognize that the computational hardness of a class of problems hard does not entail that every problem in the class is intractable. It implies only that there are some sets of problems that are hard, and so we cannot construct an algorithm that will solve every problem in the class uniformly. To take a simple example, suppose that the task is clustering. The items that we are presented with are points in a two dimensional plane, and the “language” corresponds to several roughly circular regions. The learning task is to construct a set of clusters of the data where each cluster includes all and only the points with a particular property. Formally this task is computationally hard, since the clusters may contain substantial overlap. If this is the case, then there may be no alternative to trying every possible clustering of the data. However if the clusters are well-separated, the learning task is easy, and it is one that humans perform very well.

There are provably correct algorithms for identifying clusters that are well-separated, and humans can do this simply by looking at the data on the left of Figure 5. It is easy to draw a circle around each of the three clusters in this diagram. Conversely, when the data are not separated, as in the example on the right of Figure 5, then it is hard to pick out the correct three clusters.

We can represent this difference in hardness by defining a separability parameter. If the centers are well-separated, then the value of the separability parameter will be high, but if they are not, then its value will be low. The parameter allows us to stratify the class of clusters into problems which are easy and those which are hard. Clearly, we do not need to attribute knowledge of this parameter, as a learning prior, to the learner. If the clusters are separated, then the learner will exploit this fact to perform the clustering task, and if they are not, he/she will not succeed in identifying the clusters. From a learnability point of view we could define a class of “learnable clusterings” which are those that are separable. We can prove that an algorithm could learn all of the elements of this class, without incorporating a separability parameter into the algorithm’s design.

The analogy between clustering and language learning is clear. Acquiring even simple language representations may be hard in general. However, there might be parameters that divide easy learning problems from hard ones. Stratifying
learning tasks in this way permits us to use such parameters to identify the class of efficiently learnable languages, and to examine the extent to which natural languages form a subset of this class.

6 EFFICIENT LEARNING

In fact there are some efficient algorithms for learning classes of representations. [Angluin and Kharitonov, 1991] shows that there is an important distinction between representations with hidden structure, and those whose structure is more readily discernible from data. [Angluin, 1987] shows that the class of regular languages can be learned using the class of deterministic finite state automata, when there is a reasonably helpful learning paradigm, but the class of non-deterministic automata is not learnable [Angluin and Kharitonov, 1991]. In practice DFAs are quite easy to learn from positive data alone, if this data is not designed to make the learner fail. Subsequent work has established that we can learn DFAs from stochastic data alone, with a helpful distribution on the data set.

If we look at the progress that has been made for induction of DFAs, we see the following stages. First, a simple algorithm is given that can learn a restricted class from positive data alone, within a version of the Gold paradigm [Angluin, 1982]. Next, a more complex algorithm is specified that uses queries or some form of negative evidence to learn a larger set, in this case the entire class of regular languages [Angluin, 1987]. Finally, stochastic evidence is substituted for negative data [Carrasco and Oncina, 1999]. This sequence suggests that the core issues in learning concern efficient inference from probabilistic data and assumptions. When these are solved, we will be able to model grammar induction from stochastic evidence as a tractable process. The pattern of progress that we have just described for learning theoretic inference of representation classes is now being followed in the modeling of context free grammar induction.

An important question that remains open is whether we will be able to apply the techniques for efficient learning to representation classes that are better able to accommodate natural languages than DFSA or CFGs. There has been progress towards this goal in recent years, and we will briefly summarize some of this work.

We can gain insight into efficient learnability by looking at the approaches that have been successful for induction of regular languages. These approaches do not learn just any finite state automaton, but they acquire a finite state automaton that is uniquely determined by the language. For any regular language $L$ there is a unique minimal DFA that generates it.\footnote{It is possible to relabel the states, but the structure of the automaton is uniquely determined.}

In this case, the minimal DFAs, are restricted to only one, and the uniqueness of the device facilitates its learnability. Moreover, they are learnable because the representational primitives of the automaton, its states, correspond to well defined properties of the target language which can be identified from the data. These states are in one-to-one correspondence to what are called the residual languages.
of the language. Given a language \(L\) and a string \(u\), the residual language for \(u\) of \(L\), written \(u^{-1}(L)\) is defined as \(\{v\mid uv \in L\}\). This is just the set of those suffixes of \(u\) that form a grammatical string when appended to \(u\). A well known result, the Myhill-Nerode theorem, establishes that the set of residual languages is finite if and only if the language is regular. In the minimal DFA, each state will generate exactly one of these residual languages.

This DFA has a very particular status. We will call it an objective finite automaton. It has the property that the structure of the automaton, though hidden in some sense, is based directly on well defined observable properties of the language that it generates.

Can we specify an analogous objective Context Free Grammar with similar learnability properties? There is a class of Deterministic CFGs, but these have the weaker property that the trees which they generate are traversed from left to right. This condition renders an element of the parsing process deterministic, but it does not secure the learnability result that we need.

To get this result we will pursue a connection with the theory of distributional learning, which is closely associated with the work of Zellig Harris [Harris, 1954], and has also been studied extensively by other structuralist linguists [Wells, 1947; Bar-Hillel, 1950]. This theory was originally taken to provide discovery procedures for producing the grammar of a language, but it was soon recognized that its techniques could be used to model elements of language acquisition.

The basic concept of distributional learning is, naturally enough, that of a distribution. We define a context to be a sentence with a hole in it, or, equivalently, as a pair of strings \((l,r)\) where \(l\) represents the string to the left of the hole, and \(r\) represents the one to the right. The distribution of a string \(u\) is just the set of contexts in which it can be substituted for the hole to produce a grammatical sentence, and so \(CL(u) = \{(l,r)\mid lur \in L\}\). Distributional approaches to learning and grammar were studied extensively in the 1950s. One of the clearest expositions is [Bar-Hillel, 1950], which is largely concerned with the special case where \(u\) is a single word. In this instance we are learning only a set of lexical categories.

Joshua Greenberg was another proponent of distributional learning. [Chomsky, 1959] lucidly paraphrases Greenberg’s strategy as “let us say that two units A and B are substitutable_1 if there are expressions X and Y such that XAY and XBY are sentences of L; substitutable_2 if whenever XAY is a sentence of L then so is XBY and whenever XBY is a sentence of L so is XAY (i.e. A and B are completely mutually substitutable). These are the simplest and most basic notions.”

In these terms \(u\) is “substitutable_1” with \(v\) when \(CL(u) \cap CL(v)\) is non empty and \(u\) is “substitutable_2” with \(v\) when \(CL(u) = CL(v)\). The latter relation is now called syntactic congruence, and it is easily seen to be an equivalence relation. The equivalence classes for this relation are the congruence classes, expressed as \([u]_L = \{v\mid CL(u) = CL(v)\}\).

It is natural to try to construct an objective context free grammar by requiring that the non-terminals of the grammar correspond to these congruence classes, and this approach has yielded the first significant context free grammatical inference
result, presented in [Clark and Eyraud, 2007]. Interestingly, the class of CFG languages that this result shows to be learnable is one for which, in Chomsky’s terms, one form of substitutability implies the other: a language is substitutable if whenever $A$ and $B$ are substitutable$^1$, then they are substitutable$^2$. This class was precisely defined by Myhill in 1950 [Myhill, 1950], which raises the question of why this elementary result was only demonstrated 50 years after the class was first defined. The delay cannot be plausibly attributed to the technical difficulty in the proof of the result in [Clark and Eyraud, 2007], as this proof is constructed on direct analogy with the proofs given in [Angluin, 1982].

Rather the difficulty lies in the fact that linguistic theory has been focused on identifying the constituent syntactic structure of a language, which corresponds to the strong generative capacity of a grammar. This structure cannot be uniquely recovered from the PLD without additional constraints on learning. This is because two CFGs may be equivalent in their weak generative power (ie. they generate the same set of strings), but differ in their strong generative capacity (they assign distinct structures to at least some of these strings). Therefore, a learner cannot distinguish between weakly equivalent grammars on the basis of the observed evidence.

In order to achieve the learnability result given in [Clark and Eyraud, 2007] it is necessary to abandon the idea that grammar induction consists in identifying the correct constituent structure of the language. Instead learning is characterized in terms of recovering the distributional structure of the language. This structure is rich enough to describe the ways in which the primitive units of the language combine to form larger units, and so to specify its syntax, but the resulting grammar, and the parse trees that it produces, do not correspond to the traditional constituents of linguistic theory. This may seem to be a defect of the learning model. In fact it isn’t. The constituent structure posited in a particular theory of grammar is itself a theoretical construct invoked to identify the set of grammatical sentences of the language, as speakers represent them. If we can capture these facts through an alternative representation that is provably learnable, then we have demonstrated the viability of the syntactic structures that this grammar employs.

We have passed over an important question here. We must show that a learnable grammar is rich enough to support semantic interpretation. We will shortly take up this issue in outline.

In the end, the basic representational assumption of the simple distributional approach is flawed. From a distributional point of view congruence classes give the most fine-grained partitioning of strings into classes that we could devise. Any two strings in a congruence class are fully interchangeable in all contexts, and this condition is rarely, if ever, satisfied. Therefore, a learning algorithm which infers a grammar through identification of these classes will generate representations with large numbers of non-terminals that have very narrow string coverage.

The grammar will also be formally inadequate for capturing the full range of weak generative phenomenon in natural language, because at least some languages
contain mildly context sensitive syntactic structures [Shieber, 1985].

Finally, distributional CFGs do not offer an adequate formal basis for semantic interpretation, as neither their tree structures nor their category labels provide the elements of a suitable syntax-semantics interface.

These three considerations indicate that we need a more abstract representation which preserves the learnability properties of the congruence formalism. Our challenge, then, is to combine two putatively incompatible properties: deep, abstract syntactic concepts, and observable, objective structure. It was precisely the apparent conflict between these two requirements that first led Chomsky to discard simple Markov (n-gram) models and adopt linguistic nativism in the form of a strong set of grammar specific learning biases.

In fact there is no intrinsic conflict between the demands of abstract structure on one hand, and categories easily identifiable from the data on the other. [Clark, 2009] specifies a rich distributional framework that is sufficiently powerful to represent the more abstract general concepts required for natural language syntax, and he demonstrates that this formalism has encouraging learnability properties. It is based on a Syntactic Concept Lattice.

The representational primitives of the formalism correspond to sets of strings, but the full congruence of distributional CFGs is replaced by partial sharing of contexts. This weaker condition still generates a very large number of possible categorial primitives, but, by moving to a context-sensitive formalism, we can compute grammars efficiently with these primitives [Clark, 2010]. We refer to these representations as Distributional Lattice Grammars (DLG), and they have two properties that are important for our discussion of language acquisition.

First, the formalism escapes the limitations that we have noted for simple congruence based approaches. DLGs can represent non-deterministic and inherently ambiguous languages such as

\[ \{a^n b^a c^m | n, m \geq 0\} \cup \{a^m b^a c^n | n, m \geq 0\} \]

It can encode some non-context free languages (such as a variant of the MIX or Bach language), but it cannot represent all context free languages. The examples of context-free languages that the formalism cannot express are artificial, and they do not correspond to syntactic phenomena that are attested in natural languages.

It is important to recognize that our objective here is not to represent the full set of context free grammars, but to model the class of natural languages. It is not a flaw of the DLG framework that it is not able to express some CFGs, if these do not represent natural languages. In fact, this may be taken as a success of the paradigm [Przedziecki, 2005].

Second, DLGs can be efficiently learned from the data. The current formal results are inadequate in a number of respects. (i) they assume the existence of a membership oracle. The learner is allowed to ask an informant whether a given sentence is grammatical or not. As we discussed above, we consider this to be a reasonable assumption, as long as such queries are restricted in a way that renders them equivalent to indirect negative (stochastic) evidence. (ii) The
learnability result is not yet sharp enough. Efficiency is demonstrated for each step in the learning procedure, rather than for the entire process. (iii) Although the formalism exhibits the partial structural completeness that the congruence-based models have, the labels of its parse trees have the rich algebraic structure of a residuated lattice.\footnote{In some circumstances, the derived structural descriptions will not be trees, but non-tree directed acyclic graphs. This will generally be the case when the language is not context-free.}

The operations in the lattice include the residuation operators $\div$ and $\backslash$, and the partial order in the lattice allows us to define labeled parse trees, where the labels are “maximal” in the lattice. Ambiguous sentences can therefore be assigned sets of different representations, each of which can support a different interpretation. The theory of categorial grammar tells us how we can do this, and Categorial Grammars are based on the same algebraic structure [Lambek, 1958].

The theory of DLGs is still in its infancy, but for the first time we appear to have a learning paradigm that is provably correct, can encode a sufficiently large class of languages, and can produce representations that are rich enough to support semantic interpretation.

The existence of probabilistic data, which we can use as indirect negative evidence, allows us to control for over-generalisation. DLGs provide a very rich framework which can encode the sorts of problems that give rise to the negative results on learning that we have cited. We should not be surprised, then, to find that uniform learning of an entire class in this framework may be hard. So it will certainly be possible to construct combinations of distributions and examples where the learning problem is difficult. But it is crucial to distinguish the assumptions that we make about the learner from those that we adopt for the environment. We can assume that the environment for language learning is generally benign, but we do not need to attribute knowledge of this fact to the learner.

In the context of the argument from the poverty of the stimulus, we are interested in identifying the minimal initial information which we must assume that the learner has in order to account for acquisition. We are making the following claim for DLGs. In order for acquisition of DLGs to proceed we need to hypothesize a bias for paying attention to the relation between substrings and their contexts, and an ability to construct concept lattices [Ganter and Wille, 1997]. The representational formalism and the learning algorithm both follow naturally from these assumptions. Additionally we need to posit a robust mechanism for dealing with noise and sparsity of data. Our second claim is that these mechanisms are adequate for representing a large amount of natural language.

We acknowledge that these claims require substantial empirical support, which has yet to be delivered. We do know that there are a wide range of efficient algorithms for the inference of large classes of context free languages, where these were not available as recently as ten years ago. The exact limits of the approach to learning that we are suggesting have not yet been fully explored. However, the results that we have briefly described here give some reason to think that language acquisition is computationally possible on the basis a set of minimal
learning biases. The extent to which these biases are truly domain-general is a subject for future discussion.

7 MACHINE LEARNING AND GRAMMAR INDUCTION: SOME EMPIRICAL RESULTS

In the previous sections we have considered the problem of efficient learnability for the class of natural languages from the perspective of formal learning theory. This has involved exploring mathematical properties of learning for different sorts of representation types, under specified conditions of data, time, and computational complexity. In recent years there has been a considerable amount of experimental work on grammar induction from large corpora. This research is of a largely heuristic kind, and it has yielded some interesting results. In this section we will briefly review some of these experiments and discuss their implications for language acquisition.

7.1 Grammar Induction through Supervised Learning

In supervised learning the corpus on which a learning algorithm $A$ is trained is annotated with the parse structures that are instances of the sort of representations which $A$ is intended to learn. $A$ is tested on an unannotated set of examples disjoint from its training set. It is evaluated against the annotated version of the test set, which provides the gold standard for assessing its performance.

$A$’s parse representations for a test set $TS$ are scored in two dimensions. Its recall for $TS$ is the percentage of parse representations from the gold standard annotation of $TS$ that $A$ returns. $A$’s precision is the percentage of the parse structures that it returns for $TS$ which are in the gold standard. These percentages can be combined as a weighted mean to give $A$’s $F_1$-score.

The Penn Treebank [Marcus, 1993] is a corpus of text from the Wall Street Journal that has been hand annotated for lexical part of speech (POS) class for its words, and syntactic constituent structure for its sentences. A Probabilistic Context Free Grammar (PCFG) is a context-free grammar whose rules are assigned a probability value in which the probability of the sequence of symbols $C_1 \ldots C_k$ on the right side of each rule is conditioned on the occurrence of the non-terminal symbol $C_0$ on the left side, which immediately dominates it in the parse structure. So $P(C_0 \rightarrow C_1 \ldots C_k) = P(C_1 \ldots C_k|C_0))$.

---

11 For a more detailed discussion of this applied research in grammar induction see [Clark and Lappin, 2010a].

12 Devising reasonable evaluation methods for natural language processing systems in general, and for grammar induction procedures in particular raises difficult issues. For a discussion of these see [Resnik and Lin, 2010] and [Clark and Lappin, 2010a].

13 Recall, precision, and F-measure were first developed as metrics for evaluating information retrieval and information extraction systems. See [Grishman, 2010] and [Jurafsky and Martin, 2009] on their application within NLP.
For every non-terminal $C$ in a PCFG, the probabilities for the rules $C \rightarrow \alpha$ sum to 1. The probability of a derivation of a sequence $\alpha$ from $C$ is the product of the rules applied in the derivation. The probability that the grammar assigns to a string $s$ in a corpus is the sum of the probabilities that the grammar assigns to the derivations for $s$. The distribution $D_G$ that a PCFG specifies for a language $L$ is the set of probability values that the grammar assigns to the strings in $L$. If the grammar is consistent, then $\sum_{s \in T^*} D_G(s) = 1$, where $T^*$ is the set of strings generated from $T$, the set of the grammar’s terminal symbols.

The probability values of the rules of a PCFG are its parameters. These can be estimated from a parse annotated corpus by Maximum Likelihood Estimation (MLE) (although more reliable techniques for probability estimation are available).

\begin{equation}
\frac{c(C_0 \rightarrow C_1 \ldots C_k)}{c(C_0 \rightarrow \gamma)}
\end{equation}

where $c(R)$ = the number of occurrences of a rule $R$ in the annotated corpus.

The performance of a PCFG as a supervised grammar learning procedure improves significantly when it is supplemented by lexical head dependencies. In a Lexicalized Probabilistic Context Free Grammar (LPCFG), the probability of the sequence of symbols on the right side of a CFG rule depends on the pair $\langle C_0, H_0 \rangle$. $C_0$ is the symbol that immediately dominates the sequence (the left hand side of the rule), and $H_0$ is the lexical head of the constituent that this symbol encodes, and which the sequence instantiates.

Collins [1999; 2003] constructs a LPCFG that achieves an F-score of approximately 88% for a WSJ test set. [Charniak and Johnson, 2005] improve on this result with a LPCFG that arrives at an F-score of approximately 91%. This level of performance represents the current state of the art for supervised grammar induction.

Research on supervised learning has made significant progress in the development of accurate parsers for particular domains of text and discourse. However, this work has limited relevance to human language acquisition. The PLD to which children are exposed is not annotated for morphological segmentation, POS classes, or constituent structure. Even if we grant that some negative evidence is contained in the PLD and plays a role in grammar induction, it is not plausible to construe language acquisition as a supervised learning task of the kind described here.

### 7.2 Unsupervised Grammar Induction

In unsupervised learning the algorithm is trained on a corpus that is not annotated with the structures or features that it is intended to produce for the test set. It must identify its target values on the basis of distributional properties and clustering patterns in the raw training data. There has been considerable success in unsupervised morphological analysis across a variety of languages [Goldsmith, 2001; Goldsmith, 2010; Schone and Jurafsky, 2001]. Reliable unsupervised POS taggers have also been developed [Schütze, 1995; Clark, 2003].
Early experiments on unsupervised parsing did not yield promising results [Carroll and Charniak, 1992]. More recent work has produced systems that are starting to converge on the performance of supervised grammar induction. [Klein and Manning, 2004] (K&M) present an unsupervised parser that combines a constituent structure induction procedure with a head dependency learning method.\footnote{See [Bod, 2006; Bod, 2007a; Bod, 2007b; Bod, 2009] for an alternative, largely non-statistical, method of unsupervised parsing.}

K&M’s constituent structure induction procedure determines probabilities for all subsequences of POS tagged elements in an input string, where each subsequence is taken as a potential constituent for a parse tree. The procedure invokes a binary branching requirement on all non-terminal elements of the tree. K&M use an Expectation Maximization (EM) algorithm to select the parse with the highest probability value. Their procedure identifies (unlabeled) constituents through the distributional co-occurrence of POS sequences in the same contexts in a corpus. It partially characterizes phrase structure by the condition that sister phrases do not have (non-empty) intersections. Binary branching and the non-overlap requirement are learning biases of the model which the procedure defines.

K&M’s unsupervised learning procedure for lexicalized head-dependency grammars assigns probabilities to possible dependency relations in a sentence $S$. It estimates the likelihood for every word $w_i$ in $S$ that $w_i$ is a head for all of the subsequences of words to its left and to its right, taken as its syntactic arguments or adjuncts. The method computes the likelihood of these alternative dependency relations by evaluating the contexts in which each head occurs. A context consists of the words (word classes) that are immediately adjacent to it on either side. This procedure also imposes a binary branching condition on dependency relations as a learning bias.

K&M combine their dependency and constituent structure grammar systems into an integrated model that computes the score for a constituent tree structure as the product of the values assigned to its terminal elements by the dependency and constituency structure models. This method employs both constituent and head dependency distributional patterns to predict binary constituent parse structure. The method achieves an F-score of 77.6% when it applies to text annotated with Penn Treebank POS tagging, and an F-score of 72.9% when this test set is marked by [Schütze, 1995]’s unsupervised tagger. The latter case is a more robust instance of unsupervised grammar induction in that the POS tagging on which the learning procedure depends is itself the result of unsupervised word class identification.

7.3 Machine Learning and Language Acquisition

[Fong and Berwick, 2008] (F&B) argue that supervised parsers, like Collins’ LPCFG, do not acquire syntactic knowledge of the sort that characterizes the linguistic competence of native speakers. They run several experiments with variants of Collins’ grammar. Their results contain incorrect probabilities for wh-questions, putatively problematic parses for PP attachment cases, and (what they claim to
be) some puzzling effects when non-grammatical word order samples are inserted in the data.

Some of the effects that F&B obtain are due to the very limited amount of training data that they employ, and the peculiarities of these samples. It might well be the case that if Collins’ LPCFG were trained on a large and suitably annotated subset of the CHILDES child language corpus [MacWhinney, 1995], it would yield more appropriate results for the sorts of cases that F&B consider.

But even if their criticisms of Collins’ parser are accepted, they do not undermine the relevance of machine learning to language acquisition. As we noted in Section 7.1, supervised learning is not an appropriate model for human learning, because the PLD available to children is not annotated with target parse structures. Work in unsupervised grammar induction offers more interesting insights into the sorts of linguistic representations that can be acquired from comparatively raw linguistic data through weak bias learning procedures. In order to properly evaluate the significance of this heuristic work for human language acquisition, it is necessary to train and to test machine learning algorithms on the sort of data found in the PLD.

Unsupervised grammar induction is a more difficult task than supervised parsing, and so we might expect F&B’s criticisms to apply with even greater force to work in this area. In fact, recent experimental research in unsupervised learning, such as K&M’s parsing procedure, indicates that it is possible to achieve accuracy approaching the level of supervised systems. Of course, these results do not show that human language acquisition actually employs these unsupervised algorithms. However, they do provide initial evidence suggesting that weak bias learning methods may well be sufficient to account for language learning. If this is the case, then positing strong biases, rich learning priors, and language specific learning mechanisms requires substantial psychological or neural developmental motivation. The APS does not, in itself, support these devices.

8 CONCLUSIONS AND FUTURE RESEARCH

We have considered the ways in which computational learning theory can contribute insights into language acquisition. We have seen that while formal learning models cannot replace empirically motivated psycholinguistic theories, they can provide important information on the learnability properties of different classes of grammatical representations. However, the usefulness of such models depends on the extent to which their basic assumptions approximate the facts of the human acquisition process.

We looked at two classical learning paradigms, IIL and PAC learning. Each of these has been the source of negative results that linguists have cited in support of the APS. When we examine these results closely we find that they do not, in fact, motivate a strong domain specific bias view of language acquisition. The results generally depend on assumptions that are implausible when applied to acquisition.
In some cases, they have been inaccurately interpreted, and, on a precise reading, it becomes clear that they do not entail linguistic nativism.

We observed that the main challenge in developing a tractable algorithm for grammar induction is to constrain the computational complexity involved in inferring a sufficiently rich class of grammatical representations from the PLD. We looked at recent work on probabilistic learning models based on a distributional view of syntax. This line of research has made significant progress in demonstrating the efficient learnability of grammar classes that are beginning to approach the level of expressiveness needed to accommodate natural languages.

A central element in the success of this work is the restriction of the set of possible distributions to those that facilitate learning in a way that corresponds to the PLD to which human learners are exposed. A second important feature is that it characterizes representational classes that are not elements of the Chomsky hierarchy, but run orthogonally to it. A third significant aspect of this work is that although the primitives of the grammars in the learnable classes that it specifies are sufficiently abstract to express interesting syntactic categories and relations, they can be easily identified from the data.

We then considered recent experiments in unsupervised grammar induction from large corpora, where the learning algorithms are of a largely heuristic nature. The results are encouraging, as the unsupervised parsers are beginning to approach the performance of supervised systems of syntactic analysis.

Both the formal and the experimental work on efficient unsupervised grammar induction are in their initial stages of development. Future research in both areas will need to refine the grammar formalisms used in order to provide a fuller and more accurate representation of the syntactic properties of sentences across a larger variety of languages. It is also important to explore the psychological credibility of the learning procedures that successful grammar induction systems employ. This is a rich vein of research that holds out the prospect of a rigorously formulated and well motivated computational account of learning in a central human cognitive domain.

ACKNOWLEDGEMENTS

We are grateful to Richard Sproat for his careful reading of an earlier draft of this chapter, and for his invaluable suggestions for correction. Of course we bear sole responsibility for the content of the chapter.

BIBLIOGRAPHY


LINGUISTICS FROM AN EVOLUTIONARY POINT OF VIEW

James R. Hurford

1 LINGUISTICS AND EVOLUTION

Beginning linguistics students are sometimes treated to an array of mock “theories” about the evolution of language, including the “Bow-wow” theory, the “Ding-dong” theory and others with equally silly and dismissive names. The 1886 ban on the subject (along with proposals for a universal language) by the Société Linguistique de Paris is well known, and also sometimes thrown up as a proscription that should be reimposed. Research into the evolution of language never really died, though its serious contributors, such as C.F. Hockett [1960] and Philip Lieberman [1984], were tiny in number. In the past twenty years, the resurrection of the subject has accelerated dramatically. The resurgence can be attributed to a general increase in multidisciplinary research, and to impressive empirical advances in relevant fields such as genetics, psychology of language, ethology (especially primatology), computer modelling, linguistics (especially language typology and some formal modelling) and neuroscience.

Linguistics has traditionally been isolated from evolutionary considerations. Saussure’s [1916] emphasis on the primacy of synchronic descriptions coloured all of mainstream 20th century linguistics. The core of generative grammar is synchronic work. Moreover, the emphasis in generative theory on the discovery of abstract formal principles governing the shape a language can take tends to isolate the study of language from neighbouring disciplines. The prevailing assumption within this dominant paradigm has been that the principles to be discovered are peculiar to language alone [Chomsky, 1965; 1975; 1981]. If regularities are observed that can be accounted for in terms of more general human behaviour, or even animal behaviour, such as memory limitations, limitations on the physiology of the output device (vocal or manual), or constraints on processing complexity, these have tended to be sidelined as not within the domain of linguistics proper, which is taken to be whatever is special to language alone. There is more than a whiff of Platonism in much 20th century theorizing about language. Of course, as linguistics is a large field, there have been dissenting voices (e.g. [Bybee, 1985; 1994; Givón, 1979; 1990]), emphasizing the integration of the study of language structure with the study of human and animal behaviour generally, and taking a
more favourable attitude to explanations in terms of function (as opposed to an
appeal to deep-seated abstract principles not necessarily motivated by function).

Historical linguists, though working with diachrony, have almost universally
taken a uniformitarian stance, postulating that reconstructed proto-forms of lan-
guages are no different in kind from modern languages. In a uniformitarian view
of this kind, change in language cycles around through states of language that are
all of the same recognizably modern type. This is consistent with the standard
教学 in linguistics that there are no primitive languages. Thus the idea of
languages evolving from earlier types different from the types observed today does
not get an airing. The exception is Creole studies, where it is often acknowledged
that these newly emerging languages are in some senses simpler than languages
with long histories.

The isolation of mainstream linguistics from evolutionary considerations is puz-
zling in light of Chomsky’s emphatic and influential re-location of linguistics within
psychology and ultimately biology. Human language is a product of human minds
and bodies, and these in turn are products of evolution. Chomsky and his fellow-
thinkers do not deny that the human language capacity has evolved; rather, the
argument is that the course of this evolution has not been significantly affected
by natural selection. Whatever it was that gave humans this impressive capacity,
setting us off very distinctively from other species, cannot (the argument goes)
be attributed to incremental pressure to mould a system well adapted to commu-
nication. These dominant views were challenged influentially in 1990 by Pinker
and Bloom, under the eloquent title “Natural Language and Natural Selection”.
Pinker and Bloom took their cue from mainstream generative grammar, whose
methodology they accepted, and in which the tenor of the day was still that hu-
mans are born with an innate richly structured cognitive subsystem accounting
for the rich complex structures of languages so easily acquired against the odds
by all non-pathological children. They likened the complex structure of a human
language to the complex structure of the human eye. Both, they argued, are
the products of natural selection, working gradually. The challenge was to find
some selective rationale for each separate component of the assumed many-faceted
innate language acquisition device.

The fifteen years following Pinker and Bloom’s article witnessed a spectacular
reversal of the central theme in generative theorizing. Rather than the human
language faculty being innately rich in detailed structure, a ‘Minimalist Program’
emerged [Chomsky, 1995]. From the viewpoint of language evolution, the most
important manifestation of this movement was seen in an article co-authored by
Hauser, Chomsky and Fitch [2002]. In this article, they usefully distinguished
between the human language faculty in a broad sense (FLB) and the human lan-
guage faculty in a narrow sense (FLN). FLN, that which is distinctive of human
language, when compared to animal communication and to non-linguistic cogni-
tion, may consist, at most, of a capacity for recursion. That’s all, and maybe FLN
is even empty. In this latter case (FLN is null), what makes humans capable of
language may be just the capacity to apply recursion to their communication sys-
tems; animals may conceivably be able to do some recursive computation (maybe in their navigation), but they don't use it in their communication.

The suggestion that the human language faculty in a narrow sense (FLN) is minimal is attractive to biologists and evolutionary theorists because there is less to account for. We don't have to find special selective rationales for a whole set of apparently arbitrary principles of an assumed innate complex template for language structure (‘universal grammar’, UG), peculiar to humans. Nevertheless, it remains the case that human phenotypic communicative behaviour is vastly more complex than anything in the non-human animal world. Scientific methodology, in linguistics as in biology, dictates that we postulate as little in our explanatory story as necessary. Somehow, we have to find plausibly little evolutionary changes that generated, perhaps in a cascade of subsequent changes, the massive difference we see today. And if we view language broadly, addressing the human faculty of language in the broad sense (FLB), much of the evolutionary basis for language can be sought in animal behaviour and human non-linguistic cognition.

The two major contenders for crucial evolutionary changes leading to modern language-using humans have been (1) a capacity for complex syntax, and (2) a capacity to learn tens of thousands of arbitrary symbols. The assertion that human syntax is complex is impressionistic, since it is not backed up by any quantitative metric, but the impression is surely nevertheless correct. In section 6 below on syntax, the most complex examples of syntax in non-humans will be briefly described. With ingenuity, it may be possible to reduce the syntactic complexity of languages to the interaction of recursive operations with somewhat complex lexical structure. Both complex syntax and vocabulary are acquired impressively fast and despite a poverty of the stimulus. Children produce complex sentences that they have never heard before, and acquire new lexical items on the basis of as few as one exposure in context. A third contender for a crucial difference between humans and non-humans is social trust, a factor that I will mention further in section 3, on pragmatics.

A modern child is born into a society with a rich historically developed language, and internalizes most of this historical product in less than a decade. The ability to do this is a biological genetic endowment, which must have evolved, though we don't know in detail how it happened or how long it took. Being a matter of biological evolution, it was relatively slow, possibly taking millions of years (how many millions depending on how far back you start counting). Contrasting with this slow biological evolution of the human language faculty is the historically-cultural evolution of particular languages. The very first communicative codes used by biologically modern humans were presumably extremely simple, without the elaborate structure we see in modern languages. The pre-historical evolution of languages in communities of biologically modern humans was subject to the same range of pressures as are considered today by historical linguists studying the much more recent history of languages. Analogously, the laws of physics acting in the formation of the earth over four billion years ago were the same as the laws of physics acting today, but the states of the earth then and now are very different.
The pressures on language include economy of effort on the part of the speaker, balanced by a need to get ones meaning across clearly and distinctly, learnability, and usefulness in the normal arena of use (which itself evolves). Child creators of modern Creole languages are in a privileged position compared to the earliest humans who began to use language. Our remote ancestors had no model to learn from. There had to be invention, in some informal sense of that term. The early stages in the evolution of modern languages by cultural learning over successive generations would have been very slow at the start, some time between 200,000 and 100,000 years ago. It probably speeded up exponentially over the centuries. Certainly the ancient classical languages we know of today, less than 10,000 years old, look completely modern. To summarize, there are two senses of “evolution of language”. One is the relatively slow biological evolution of humans up to a language-ready stage; the other is the historico-cultural evolution of particular languages. A possibility that now seems increasingly plausible is that there has been a certain amount of gene-language coevolution. Given some cultural development of shared symbolic communication systems, the use of which conferred advantage both on groups and individuals, genes would have been selected enabling more efficient use of such systems. Such biological adaptations to an incipient linguistic culture would have favoured faster processing and increased memory storage for mappings between forms and their meanings.

Research into the evolution of language is unlikely to provide answers to the most commonly asked, and most naïve, factual questions. The most common, and most naïve, at least from laypeople, is “What was the first language?”. Linguists are rightly dismissive of this naïve question, as the techniques of historical reconstruction lose their power after at most 10,000 years. A minority of researchers (e.g. [Ruhlen, 1994]) claim to be able to reconstruct at least a few lexical items of “Proto-World”, the putative mother language of all modern languages. This work is very widely rejected by linguists, especially historical linguists, who argue that the statistical effects of merely coincidental change are not properly considered, and that the sampling methods on which the reconstructions are based are unsound. Less naïve questions, such as, for example, “Did Homo neanderthalensis have a language capacity comparable to Homo sapiens?”, or “At what stage in human pre-history did subordinate clauses first appear?”, are not now answerable, and may never be answerable.

So what kinds of questions do researchers in the evolution of language address? The key, I believe, is to take a cue from the dictum of the evolutionary biologist Dobzhansky [1973], who wrote “Nothing in biology makes sense except in the light of evolution”. In parallel, I claim, nothing about language makes sense except in the light of evolution. Linguists, for the most part firmly embedded in a synchronic paradigm, tell us in great detail what individual languages are like, and generalize, with the help of developmental psychologists, to what the innate human language faculty is like. These are descriptions of the current historico-cultural and biological states of affairs. Evolutionary linguistics adds an explanatory dimension. Of both kinds of state of affairs, the biological and the historico-cultural, we pose
the additional question, “And how did things get to be that way?”. Chomsky can be credited with adding the dimension of explanatory goals, as opposed to merely descriptive goals, to linguistics. The Chomskyan type of answer to “Why are languages the way they are?” is an appeal to innate dispositions in the language-learning child. This presupposes an existing complex language, to which the child is exposed. At the historico-cultural level, evolutionary linguistics adds the question “And how did existing complex languages get to be the way they are?”. At the biological level, the relevant additional question is “And how did the human species get to be the language-ready way it is?”. Obviously, these are contentful (if difficult) questions; they lead us away from the inward-looking study of language on its own, to a wider perspective of language in the context of perception, cognition and social arrangements, in non-human animals as well as in humans. The inward-looking theories of more conventional linguists must eventually be compatible with the wider perspective afforded by taking evolution, both biological and cultural, into account.

A language, conceived broadly (i.e. taking an FLB viewpoint) is a bridge between meanings and sounds (or manual gestures), and the meanings and sounds are themselves parts of the system, the end supports of the bridge, to pursue the metaphor. It will be convenient here to discuss this bridging system in terms of the usual compartments posited in linguistic theory, namely pragmatics, semantics, syntax, phonology and phonetics, as shown in Figure 1.

Figure 1. A language system is a bi-directional bridge between meanings and sounds. Linguistics naturally carves this bridge into the structurally different components identified here.

Viewing linguistic structure from an evolutionary point of view, one asks of each separate structural component of a language system “How did it get to be that way?” This leads one to consider possible evolutionary antecedents to the subsystem in question, which in turn leads to the recognition that, from an evo-
utionary point to view, some modification of the traditionally defined boundaries between parts of a language system is appropriate. Thus the separate sections that follow will each begin with a definition of the relevant domain (e.g. pragmatics, phonetics) which is convenient from an evolutionary perspective.

The various subsystems of an overall language system are of varying antiquity and provenance. Some aspects of human linguistic behaviour are extremely ancient, shared with many mammals as part of a common biological heritage. Other aspects are also biologically determined, but special to humans, and are therefore more recent evolutionary developments. Finally, much human linguistic behaviour is learned, and culturally transmitted across generations; such parts of a language have evolved historico-culturally. For some conceptual clarity, in Figure 2 I give an extremely schematic view of the varying antiquity and provenance of subcomponents of a language system. This diagram should be taken with great caution; it is meant to be suggestive only, and there is no way in which such a diagram can be properly scaled. In what common quantal units, for instance, can one ‘measure’ the relative ‘sizes’ of the semantic and phonetic components of a language? It is clearly not possible. Yet I hope that this diagram will at least serve as a mnemonic for a general message, namely that the more peripheral parts of a language system, those dealing directly with meanings (pragmatics and semantics) and with sounds (phonetics) are the more ancient, and more rooted in our common biological heritage with other mammals. The inner parts of a language system, those having more to do with the formal organization and distribution of morphemes, words, phrases, clauses and sentences (i.e. morphosyntax) and of syllables and phonetic segments (i.e. phonology) have substantial learned components; thus the body of the syntax and phonology of a language, which must be learned by a child, has evolved historico-culturally, though it is still enabled by specific relevant biological capacities. As languages have grown culturally, there has also been a degree of biological co-evolution, adapting the organism to cope with the new complexities. In Figure 2, these new culturally-driven biological adaptations are represented by blocks labelled ‘NEW BIO’.

2 SEMANTICS FROM AN EVOLUTIONARY POINT OF VIEW

Semantics is usually, within linguistics, distinguished from pragmatics by not involving language-users in communicative situations. Traditionally, semantics has been defined as involving the truth-conditions of sentences and the denotations of words and phrases, considered out of context. This does not mean that linguistics ignores communication between people; this is dealt with under the heading of pragmatics. From an evolutionary point of view it is useful to keep half of the traditional core of semantics, namely the notion of a relation to entities in a world outside language, populated by objects, their static and dynamic properties, relations between objects, events involving such objects, and so on. Thus semantics is viewed as essentially extensional, involving a relation to an outside world. But if we are interested in the evolution of language as a psychobiological phenomenon,
we cannot contemplate a direct relation between elements of language (words, sentences) and the world. Rather, this relation has to be mediated by minds. This idea is encapsulated in Ogden and Richards’ [1923] “triangle of signification” in which it is emphasized that the relation between language and the world is indirect, being mediated by mental entities such as concepts and thoughts. Words and phrases express concepts; the relation of denotation can be reserved for the relation between concepts (mental entities) and things, properties, events and so on in the world.

Now, I assume that some non-linguistic creatures, such as apes and human babies, can have thoughts and concepts. That is, I reject the view that concepts can only be possessed by language-possessing creatures. Some thoughts and concepts pre-existed language. These were private in the creatures that possessed them. With the advent of language, public conventional labels got attached to these private concepts. The attachment of public labels to formerly private concepts had some effect on them. Concepts became socially standardized. But this is to rush ahead. For the moment, suffice it to say that the definition of semantics adopted here is the relationship between concepts (alias thoughts) and entities in the world, as it exists in both linguistic and non-linguistic creatures.

What kinds of concepts and thoughts are non-human animals capable of? It is safe to say that relatively “higher” animals, such as mammals and birds, form mental categories of the types of objects and situations which are relevant to their daily lives, such as different types of food, different food sources and types of predators. They discriminate systematically between such things. This is the evolutionary link to the denotations in an outside world of expressions in human languages. First the concepts of things, events and situations in the world existed
in pre-humans. These were private concepts, about which the creatures did not communicate among themselves. Later, humans became able and willing to attach public labels to these pre-existing concepts, and to use the labels for communication. As a terminological matter, it may be useful to speak of ‘proto-concepts’ in pre-linguistic minds, and to reserve the unmodified term ‘concept’ for those mental entities linked to expressions in a public mode of communication.

Probably some, but not all, of the mental categories non-humans form are learned from experience. There are more or less strong innate dispositions to acquire certain specific categories, and these dispositions are linked to the evolved sensory apparatus of the animal. In some cases, like that of the famous vervet monkey alarm calls ([Seyfarth and Cheney, 1982] — and the similarly specific alarm calls of many other mammalian and avian species) the private concept is linked to a public signal. In these cases, the linkage itself between the concept (e.g. LEOPARD) and the appropriate call (e.g. a ‘bark’) is innately specified to a high degree. The particular behaviour just grows in each animal in a uniform way determined by its genetically driven developmental program.

It is important to make the reservation that a non-human animal has little or no awareness of, or possibility of voluntary control of, its mental categories. A dog cannot, we assume, monitor its mental states to anything like the degree to which adult humans can monitor theirs. Recently claims have been made that non-human animals show some evidence of metacognition, that is, an ability to know their own internal states. For example, an animal trained to make certain categorical distinctions in the lab, can seem to be aware of its own uncertainty in borderline cases [Smith et al., 1995; 1997; 2003a; 2003b; Smith and Washburn, 2005].

It is often said that non-human animals live exclusively in the ‘here and now’. This difference from humans is a matter of degree, and not a categorical difference. For one thing, the concepts of ‘here’ and ‘now’ are extremely flexible. The place referred to as ‘here’ on one occasion may be a dot on a printed page, or it may be this whole universe, as opposed to some alternative or parallel universe, or some region of intermediate size between these extremes. Similarly, ‘now’ can mean this instant, or today, or this century. There is clear evidence that some non-human animals can keep in mind representations of things or events that are not immediately present to their senses. Experiments in ‘object-permanence’ show that a dog, for example, can be aware of the presence of an object hidden behind a screen for up to about five minutes after it last saw it (and without being able to smell it) [Gagnon and Doré 1992; Watson et al., 2001]. A gorilla could recall a sequence of events that it has been shown for up to fifteen minutes after seeing them [Schwartz et al., 2005; 2004]. A chimpanzee who saw food being hidden one day remembered where it had been hidden the next day [Menzel, 2005]. These animals, then, can remember things and events that are not just in the ‘here and now’. Certainly, human abilities far outstrip these non-human performances, but it is a matter of degree. Closely related to this research is work on episodic memory. It has been claimed that only humans have memories for specific events
that have happened to them. Research on several species begins to undermine this suggested absolute difference between humans and non-humans. Scrub jays, for instance, have been shown to remember where they hid food, what kind of food they hid (i.e. perishable or non-perishable) and how long ago they hid it [Clayton et al., 2001; 2003; Clayton and Dickinson, 1998]. Again, such performances are not in the same impressive league as human memories for past events, but the difference is a matter of degree.

It is clear that many animals make plans. Some non-human animals show an ability to distinguish memories of past events from their plans for future behaviour, thus demonstrating an incipient mental distinction between representation of the past from representation of the future. In an experimental situation, rats showed balanced memory limitations for places in a maze that they had already visited and places that they had not yet visited, but presumably planned to [Cook et al., 1983]. This shows the overlap of mechanisms for retrospective memory of past events and prospective memory for planned actions.

Animals attend to individual objects and make categorical judgements about them. The mechanisms of attention, for example involving gaze orientation for visual attention, are distinct from the mechanisms for recognition of objects as belonging to particular categories. An animal’s attention is drawn to an object by some simple noticeable change, such as movement or flickering, and subsequently different neural streams feed information from the perceived object into brain areas for categorical recognition, e.g. as food, or as prey. This separation of neural mechanisms, it has been suggested [Hurford, 2003a; 2003b], parallels the logical distinction between a predicate and its argument, where the argument is taken to be an individual variable, not an individual constant. For example, it is suggested that a reasonable schematic representation for what happens in an animal’s brain when it spots a wriggling object and then recognizes it as a snake is the formula $\text{SNAKE}(x)$. The variable $x$ stands for the bare object, with no categorical information bound to it; the term SNAKE here stands for the categorical predicate-like judgement that the animal makes about the object. More generally, many non-humans animals represent the world in basic ways compatible with the logical schema $\text{PREDICATE}(x)$. More thought needs to be given to how complex scenes are represented, and on the overall plausibility of this idea as a pre-human platform for the subsequent evolution of semantic representations in a form something like the formulae of predicate logic. For suggestions, see [Hurford, 2007].

Some remarkable laboratory animals have shown an ability to master second-order judgements, that is to apply predicates to predicates. The late African grey parrot, Alex, could correctly report on the colour, or shape, or material, of an object he was shown [Pepperberg, 2000]. On being shown a red ball, for instance, he could report that is was red, or a ball, whichever superordinate category (colour or shape) he had been asked about. This is the crux of the demonstration that he managed second-order judgements. He was asked, for instance, “What colour is this?” , and he would answer “Red”. Note that this requires that he knows that
red is a colour. To judge that an object is red is to make a first-order judgement, about a property of the object. To know that red is a colour is to know a second order fact, predicating COLOUR of the predicate RED. Alex, now sadly dead, was trained in a lab. He could generalize across tasks, so there is no question of his simply memorizing rote answers to specific questions. Just what call there might be in a wild state of nature for this advanced ability is not obvious. This illustrates a general point that captive animals often exhibit latent abilities for which they have no apparent need in a wild situation. We will see further examples below. The implication is that the mental abilities of animals, no doubt including humans, are sensitive to their environmental conditions, including social conditions.

Non-human animals do have what can reasonably called thoughts, primarily about the immediate world of perception and action. Human thought, with the aid of language, can be more subtle and complex. When public symbolic labels are attached to previously private (proto-)concepts, their boundaries tend to become sharpened in individual minds and standardized across the community. Much research with children and adults demonstrates that learning distinct labels for different regions in a conceptual space makes discrimination within that space easier [Balaban and Waxman, 1992; Xu, 2002; Booth and Waxman, 2002; Katz, 1963; Goldstone, 1994; 1998].

After the emergence of symbols for relatively basic concepts (‘words’), humans at some point began to string them together to encode more complex messages. More complex thoughts could now be held in memory for longer with the aid of ‘out loud’ rehearsal of the public sequences of words. We are all familiar with how ‘thinking out loud’ helps us to manage our thoughts. Chomsky tends to the view that this is the primary function of language. It certainly is one function, but the interpersonal communicative function preceded the internal thought-managing function, because the forms that we think in when we think out loud are just those of the language we have learned for communication in our social group. English speakers use English as an instrument for complex thought, Mandarin speakers use Mandarin for the same purpose. The combinatoriality of syntax makes some thoughts accessible which were previously unthinkable. Think, to the extent that you can, of the square root of minus one, or even just minus one. These were concepts inconceivable before combinatorial language.

3 PRAGMATICS FROM AN EVOLUTIONARY POINT OF VIEW

We, unlike apes, feel free to give each other potentially useful information, and we believe the information given to us. Apes, even domesticated ones such as Kanzi [Savage-Rumbaugh, 1986; 1990; 1999], are notably ungenerous in their communication, though they have learned to be trusting of their human foster parents. By contrast, in the wild, life is much more competitive, and it is unknown for non-human animals to inform each other about the state of the outside world by learned symbols. Some species have evolved small innate systems for such purposes as alerting conspecifics to the presence of predators or food. Vervet monkeys
and honeybees are the best known examples, but there are many others. All such systems in non-human nature are extremely limited in the scope of what they can ‘refer’ to (e.g. just three types of predator, or the direction and distance of food), and do not need to be learned from experience by the animals concerned.

Non-human animals do, however, communicate. All higher species communicate in some way or other. Here communication is defined as behaving in a way that affects the consequent behaviour of others, other than by straightforwardly causal physical manipulation of their bodies. The most basic, and very widespread, type of communication is purely dyadic, just designed to bring about a response from the receiver of the signal. Courtship behaviour is a central example. The wooing animal has evolved a characteristic way of behaving (e.g. strutting, singing, chest-puffing, distended sexual organs), and wooed animals have evolved complementary ways of responding to such signals. Threat displays such as teeth-baring or piloerection, and submissive displays such as cowering and rolling over are further examples. Such behaviour enhances the survival or reproduction chances of the participants and is largely instinctive.

We can see the evolutionary link to human linguistic behaviour in Austin’s [1962] phrase ‘doing things’. Animals do things to each other in their communication. Humans also use words to do things to each other. A human greeting such as ‘Hello’ is functionally parallel to a dog’s tail-wagging; it is a preliminary move toward subsequent friendly interaction. Of course, human greeting is under voluntary control, whereas the greeting behaviour of dogs is involuntary. Another difference is that human greetings in language are learned arbitrary signals. For each language, you have to learn a different conventional greeting word or phrase. But the functional connection to animal behaviour remains. Most communication in language is not purely dyadic like a ‘Hello’ greeting. Almost all linguistic communication is referential, in the sense of being about something other than the speaker or hearer. But the ‘doing-things-to-each-other’ aspect of communication is always present. Why else would we speak? In linguistic pragmatics, this aspect of language is captured by the term ‘illocution’. The illocution of an utterance is what is done, typically to the other person, in making that utterance. For instance, my uttering “the door is open” can be used to invite you, to dismiss you, to warn you of danger, or to get you to close the door, depending on the mutually understood context.

Mutual understanding of the purposes of communication is omnipresent in human linguistic behaviour. When someone says something, we assume it is said for a reason, and we try to divine the speaker’s goal [Sperber and Wilson, 1986]. Sometimes the process of figuring out a speaker’s actual intent can be quite circuitous, as in my example of ‘the door is open’. We humans do this with the benefit of a well developed theory of mind. We know the range of possible psychological states that a speaker may be in, and we can guess quite well what the speaker in a given situation knows and does not know about that situation. Much human discourse is consequently oblique. A typical exchange might be “There’s no milk”, followed by “It’s Sunday”. Such oblique communication works because the interlocutors
understand each other’s possible motives and current knowledge. Non-human animals also display some very basic understanding of the moods and knowledge of others. A chimpanzee, for example, can tell the difference between very similar physical actions, according to whether they are aggressive (e.g. teasing) or unsuccessfully cooperative (e.g. fumbling) [Call et al., 2004]. Other experiments seem to show that a chimpanzee can know whether another, dominant, chimpanzee has seen, and thus knows about, a particular food item [Call and Tomasello, 2005; Hare et al., 2000; 2001]. Humans far outstrip non-humans in this ‘mind-reading’ behaviour, but the difference is a matter of degree.

The evolutionary move from merely dyadic communication, involving only the sender and the recipient, to triadic signalling, where messages are also about some other object or event in the world, is facilitated by a capacity for joint attention. In simple cases, when humans converse about some object in the context of their encounter, both parties attend to the object. Uttering “Pass me that cup” assumes in some sense that both speaker and hearer can attend to the cup in question. Some non-human animals are adept at following the gaze of others, thus bringing about joint attention to the same object [Bräuer et al., 2005; Call et al., 1998].

If for some reason language is not available, as when people don’t speak the same language, humans use pointing to achieve joint attention. Although they are physically capable of pointing, apes in the wild never use any form of pointing to draw attention to objects [Leavens, 2004; Leavens and Hopkins, 1998]. By contrast, human toddlers often voluntarily point to objects, apparently merely to draw attention to them for the purpose of sharing an interest with an adult. In captivity, apes have learned to point to things, but almost exclusively as a means of requesting objects. Compared to human babies, apes are mercenary or instrumental in their dealings with humans and with other apes. Any pointing is motivated by directly selfish ends. Children are not like that.

This raises a central puzzle in the evolution of language. Why should any creature voluntarily share information with another? Information can be valuable, and a selfish disposition advocates that one should keep valuable information to oneself. Various theories developed in biology begin to unravel this puzzle. Passing information to ones close kin (e.g. offspring or siblings) can enhance the fitness of individuals with whom one shares genes, and thus the sharing of information is expected to be adaptive between close kin, by a process known as kin selection [Hamilton, 1964]. Many non-human animals act altruistically toward close kin, and even humans have been shown to share information more with close kin than with unrelated individuals. This establishes a parallel between humans and non-humans. But humans willingly share information, and more generally act altruistically, with non-kin. Here the theory of reciprocal altruism [Trivers, 1971] can make some headway. Within a social group, as theoretical models (computational and mathematical) models show, “tit-for-tat” behaviour is adaptive. This is, a strategy of acting altruistically toward another individual is advantageous if there is some reasonable assurance that the altruism will be reciprocated at some future time. There is evidence for such reciprocal altruism in some non-human species,
as when chimpanzees form alliances for mutual defence, and in some food-sharing activity [de Waal, 1989]. Reciprocal altruism is much better developed in human communities. We are more disposed to communicate cooperatively with people in our own social group than with outsiders, and within-group cooperation is typically reciprocated. A further motivating factor in human signalling of valuable information is that it brings prestige to the communicator, and is thus adaptive [Dessalles, 1998].

In short, human communication in language shares with animal communication the doing-things-to-each-other feature; many non-human species have limited instinctive unlearned systems for alerting others to things crucial to their survival such as predators or food; non-human animals show hints of what it takes to figure out the communicative intentions of others, such as gaze-following and a rudimentary theory of mind, but in the wild they do not apply these abilities in learned systems for communicating referentially about a wide range of external objects and events. The difference lies largely in the degree of natural cooperation that is built into the genes and the societies of humans and non-humans. We humans (believe it or not) are the species most disposed to act altruistically and cooperatively with members of our own social group.

4 PHONETICS FROM AN EVOLUTIONARY POINT OF VIEW

Phonetics is conveniently defined from an evolutionary viewpoint as the hardware of speech production and perception. Although human language exists in both oral and manual modalities, it does not seem (on the evidence so far) that human manual dexterity is specially adapted for signing, or that our vision is specially adapted for interpreting manual signs. On the other hand, the output machinery for speech, namely the whole human vocal tract, is clearly adapted, probably rather recently, for speech. As for the input stream, there is less agreement about whether human hearing is specially adapted for speech processing. I will discuss the question of human hearing first.

Mammalian hearing, to the cochlea, is rather uniform. In experimental situations, it can be shown that chinchillas have similar categorical perception of voice onset time (e.g. the difference between a [b] and a [p]) as humans. Tamarin monkeys make the same discriminatory judgements of rhythm in different languages (e.g. the rhythmic difference between Dutch and Japanese) as human babies [Tincoff et al., 2005]. Chimpanzees perceive the differences between simple syllables (e.g. [ba], [ga], [da]) in the same way as humans [Kojima et al., 1989]. And chimpanzees can do vocal tract normalization, that is they can recognize the ‘same sound’ spoken by different speakers [Kojima and Kiritani, 1989]. Opposing such evidence proposed against any special adaptation for speech in human hearing, the point has been made that normal speech perception by humans involves putting all such separate abilities together very fast in extracting complex hierarchical meanings from the stream of speech, and there is no evidence that non-human animals can manage that [Pinker and Jackendoff, 2005]. The bonobo Kanzi can
follow simple spoken instructions such as “put the coffee in the milk”, so evidently he can pick individual words out of the stream of speech [Savage-Rumbaugh et al., 1993]. On the other hand, it has also been shown that chimpanzees’ auditory working memory is impoverished as compared to humans [Hashiya and Kojima, 2001].

The issue of whether human hearing is specially adapted for speech is distinct from the issue of whether humans have distinct mechanisms for processing speech sounds and other environmental sounds (such as the sound of wind blowing or rocks falling). Humans do have mechanisms for speech processing that are separate from their mechanisms for processing other sounds [Liberman and Mattingley, 1989]. At the periphery of the system there is no difference, but at some point in the processing system there is a filter that directs speech sounds to brain regions specialized for speech processing, not surprisingly largely in the left hemisphere. But this dual-system arrangement is not special to humans. Many animals, including primates, have at least partly separated brain mechanisms for processing the calls of conspecifics and other environmental noises [Zoloth et al., 1979; Heffner and Heffner, 1984; 1986; Ghazanfar and Hauser, 2001; Hauser and Andersson, 1994; Ghazanfar et al., 2001]. Within humans, the slogan “speech is special” applies, because of this separation between speech sounds and other sounds. But it does not follow that humans are special in this regard, because as we have seen, many primates also have distinct mechanisms from processing the communicative sounds of their species.

Coming now to human speech production, there is no doubt that specialized adaptations have occurred in our species fairly recently. All of our speech apparatus has been exapted from other functions. The tongue and teeth originally evolved for eating, the lungs for breathing, and the glottis (vocal cords) for keeping water out of the lungs and bracing the chest at times of exertion.

The most widely discussed adaptation for speech is the lowering of the larynx. In all other mammals the normal position of the larynx is close up behind where the nasal passage joins the oral passage, just behind the velum. This is also the position of the larynx in newborn human infants, which allows them to breathe and suckle at the same time. During the first half year of life the human larynx lowers to near its later adult position. In this way ontogeny recapitulates phylogeny, as the adult human larynx has lowered in our evolution from apes. A lowered larynx creates a two-chamber supraglottal vocal tract. The rear chamber, the pharynx, and the front chamber, the mouth, can be narrowed or broadened complementarily. As a result, the vibrating air column used in vowel production can either pass first through a narrow tube and later through a wider tube (giving an [a] vowel), or first through a wider tube and then through a narrower tube (giving an [i] vowel). This flexibility of the upper vocal tract make possible a range of different vowel qualities that apes cannot produce. It seems reasonable that this is an adaptation allowing for a greater variety of spoken signals [Lieberman, 1984].

The story of larynx lowering is slightly more complicated. Some non-humans can lower their larynx dynamically, as dogs do momentarily when barking, and male
deer do in a more stately fashion when roaring [Fitch and Reby, 2001] But such animals nevertheless do not have the two-chamber vocal tract that makes possible the range of human vowel qualities. Although the selective value of a lowered larynx was largely for greater versatility in vowel production, the further slight lowering of the larynx in human males at puberty is also probably an adaptation for sexual selection. The difference between adult male and female voices is by far the most prominent case of sexual dimorphism in language.

Another recent adaptation is voluntary control of vocalization and breathing. Other apes have little or no voluntary control over their vocalizations. There are no significant connections from cortex to the larynx. Ape cries are spontaneous and automatic. Humans also have good control over their breathing. During speech, an outbreath may last up to thirty seconds, with the air being released in a very slow and controlled manner. Comparisons of the skeletal channels for the nerves that work the muscles involved in breathing shows a recent expansion in humans, suggesting an adaptation for greater control of breathing [MacLarnon and Hewitt, 1999]. A similar claim for the hole in the base of the skull through which the nerves controlling the tongue pass has not been substantiated [Kay et al., 1999; DeGusta et al., 1999], though there is little doubt that humans have finer control over the configurations of their tongues than other apes. Human speech production is exquisitely orchestrated, and the human vocal tract and the cerebral machinery controlling it are undoubtedly recent adaptations since divergence from our last common ancestor with the chimpanzees about six million years ago.

5 PHONOLOGY FROM AN EVOLUTIONARY POINT OF VIEW

Phonology is defined as the patterns of speech in languages. Languages organize their sound patterns within the possibilities afforded by the auditory and vocal apparatus. The physical apparatus is pretty much universal, give or take some individual variation not reflected in the organization of particular languages. The raw material of phonological organization is given by the mobility of the jaws and velum, the great flexibility of the lips and tongue and the several possible states of the glottis while air passes through it. The range of possible noises that can be made using these instruments is vast. Imagine a sequence of random uncoordinated impulses to these articulators. The product would be nothing like speech. Speech is to such random vocal noise as ballet is to the uncoordinated staggering and falling of a drunkard. Speech in languages is disciplined into repertoires of precisely specified and tightly controlled conventional moves. Acquiring perfect pronunciation in a language requires early exposure and practice. People starting a new language after the age of about eight rarely achieve perfect pronunciation.

The vocal articulators are like an orchestra [Browman and Goldstein, 1992]. During tuning, each instrument acts independently of the others, and the result is cacophony. For example, the lips can open and close at any time, the vibration of the vocal cords can be switched on or off at any time, the tongue can move between any of its possible configurations at any pace and the velum can be raised
or lowered at any time. Nothing in the inborn physical apparatus dictates that any of these actions be coordinated. All spoken languages, however, are structured in terms of basic building blocks, namely phonetic segments and syllables, which are produced by strict coordination of the actions of the various articulators. Without such coordination, speech sounds as they are commonly understood, and for which the International Phonetic Alphabet has symbols, do not exist.

The basic training of the speech apparatus to produce these discrete speech sounds occurs during a child’s development. Human infants, unlike the young of other apes, spontaneously babble, exercising their vocal apparatus at first in random ways but progressing toward sequences which more recognizably consist of speech sounds organized into syllables of consonant-vowel (CV) structure. A basic alternation between consonants and vowels makes each individual sound easier to recognize as a self-standing unit. The CV structure is found in all languages. Some languages have developed more complex syllable structures, with short clusters of consonants and more complex vowels (e.g. diphthongs), but any tendency toward such complexity is at the cost of easy speech perception. The auditory feedback received by the babbling infant helps it to map its motor movements onto acoustic patterns. The disposition to babble is thus adaptive in a social group that already benefits from communication in speech. It seems likely that a capacity for finer tuning of the articulators and more precise coordination of their interaction evolved biologically as the benefits of well articulated speech emerged. This would have been a case of gene-culture (more specifically gene-language) co-evolution.

We analyze languages as having inventories of phonemes just because these units are re-used over and over in many different words. Given a massive vocabulary of tens of thousands of words, it is costly for each separate word form to be phonetically sui generis, memorized holistically. In every language there is a handful of expressive forms that resist representation as a sequence of the normal phonemes of the language. Examples include: the expression of disgust conventionally, but inaccurately, spelled as ‘Ugh’; the alveolar click used to express disapproval (with ‘tsk’ as an attempted spelling); the bilabial affricate used to respond to cold (spelled ‘brrr’), and so on. Such expressions are not composed of phonemes in regular use elsewhere in the language. This type of expression is perhaps an evolutionary remnant of a pre-phonological stage when speech was limited and not organized around a versatile inventory of re-usable phonemes. But once large vocabularies became available it was not practical to organize the bulk of the word store in this way. The process by which re-usable phonological units get crystallized out of a mass of inchoate vocalizations has been modelled computationally [Zuidema and de Boer, 2009; de Boer and Zuidema, 2010]. The competing adaptive pressures leading to the emergence of small inventories of systematically re-usable segments are ease of articulation and mutual distinctiveness of words from each other. This evolutionary process can be seen as an instance of self-organization of a system in the environment provided by the phonetic apparatus and given the twin pressures just mentioned.

Self-organization an also be seen in the evolution of vowel inventories. Mod-
elling vowels is relatively straightforward as the continuous articulatory and acoustic spaces that they occupy are well understood, with only three main dimensions that do most of the work. Languages differ in the number of their vowel phonemes, from as few as two to over a dozen as in English. In the statistical distribution of the size of vowel inventories, the most common size is five vowels, roughly [i], [e], [a], [o], [u], as in Spanish. Systems with fewer than five vowels and with more than five vowels are decreasingly common in languages as the number differs from five. However many vowels a language has, they tend to be arranged symmetrically around the vowel space, this making maximum use of the space. The evolution by self-organization of vowels from randomly distributed beginnings has been simulated computationally. The model captures well the distribution of different numbers of vowels across languages. The model can be interpreted as mimicking the ancient processes by which well-organized vowels systems emerged in the earliest languages. The joint adaptive pressures causing this emergence are ease of articulation and mutual distinctiveness of each vowel from all the others. It is these same pressures that maintain vowel systems in extant languages in roughly the same symmetrical states over the course of their histories.

The studies surveyed above account quite successfully for the gross features of the phonological organization of all languages, namely their basic CV structure, their basis in sets of consonant and vowel phonemes, and the typical distribution of vowels in the acoustic/articulatory space. Modelling has not yet progressed to the fine detail of the ways in which adjacent sounds in a language affect each other, though this is a pervasive aspect of phonological organization. But we can nevertheless see an evolutionary dimension in such phonological effects. Natural phonetic influences which are felt by all speakers, modifying the canonical form of a phoneme, can become conventionalized, so that a synchronic phonological rule describes the regular effect. For instance, it is natural for a canonically voiced phoneme to be devoiced in anticipation of a following pause (as pauses are voiceless). In German, this devoicing has become institutionalized and extended to all word-final canonically voiced phonemes. We can see the modern synchronic rule as the trace of more optional processes earlier in the history of the language. Many synchronic phonological rules are the lasting after-effects of earlier historical sound changes in a language.

6 SYNTAX FROM AN EVOLUTIONARY POINT OF VIEW

As mentioned in the introduction, the two major contenders for crucial evolutionary changes leading to modern language-using humans have been (1) a capacity for complex syntax, and (2) a capacity to learn tens of thousands of arbitrary symbols. The former, the capacity for syntax, has always been regarded as the most challenging and theoretically interesting. A large memory for symbols was regarded as less interesting. The exciting focus of linguistic theory was syntax. Humans obviously have a unique capacity for syntax. From the early days of generative grammar in the 1950s until the mid-90s, it was assumed that this capacity
was complex, comprising up to half a dozen interacting principles. These principles were assumed to be innate, not needing to be learned, and arbitrary, not motivated by functional factors. A child learning a complex human language was assumed to receive substantial help from inborn knowledge of the abstract ways in which languages work. Here ‘abstract’ means that the innate principles were held to deal in terms of generalizations over syntactic categories (such as noun, noun phrase, verb, verb phrase), and general constraints on operations on the hierarchical tree structures of sentences (for example, an element could not ‘move’ over certain specified constituents). Discovering the set of these inborn principles, and the manner of their interaction, was the central goal of generative syntactic theory. Theorists in the generative paradigm became known as ‘formalists’. Outside this paradigm, the ‘functionalists’ objected to the generativists’ emphasis on abstraction and their lack of concern for functional explanations of the properties of language.

In the mid-1990s a major revision of generative syntactic theory appeared in the form of Chomsky’s ‘Minimalist Program’ [Chomsky, 1995]. Here the number of innate principles was in theory reduced to just one. It was suggested that perhaps the only distinctive feature of the human syntactic capacity is a capacity for recursively combining words and the phrases they compose [Hauser et al., 2002]. The central operation of syntax was ‘Merge’. Since even simple operations, if applied recursively, can lead to impressive complex structures (and, sociologically, because of old habits), the discussions of adherents to the Minimalist Program continued to have a highly abstract flavour. It became clear, however, that there was evolving convergence, from many camps, on the simple idea that what is distinctive about the human syntactic capacity is just semantically compositional combinatoriality. Various generative, but non-Chomskyan, theoretical frameworks, such as Head-driven Phrase Structure Grammar (HPSG) [Pollard and Sag, 1987; 1994; Levine and Meurers, 2006] and Construction Grammar [Fillmore and Kay, 1993; Fillmore et al., 2003; Goldberg, 1995; 2006; Croft, 2001], had already been pointing in this direction for several decades. From an evolutionary point of view, the reduced complexity of the syntactic apparatus innately programmed to develop in the child was welcome, as it simplified the likely course of human evolution. The evolution of one trait is less challenging to explain than the evolution of several mutually influencing traits. Biologists interested in human evolution welcomed this theoretical development in linguistics. Nevertheless, even with this simplification, it was still thought that there had been in human evolution a qualitative leap from non-syntactic ‘protolanguage’ to fully combinatorial language. No continuity was seen between unstructured stringing together of words and the more complex morphosyntactic systems seen in modern languages. In box diagrams of the architecture of language, separate boxes for ‘lexicon’ and ‘syntax’ were assumed. At a certain level of granularity this is acceptable. Analogously, any sensible description of human anatomy identifies separate organs. The challenge to evolutionary theory is to explain, for example, how the mammalian eye could have evolved from a non-eye, or a backbone from a non-backbone. How, without a biologically
implausible saltation, could human syntax have evolved out of non-syntax?

Recently, a way of approaching this question has emerged, mainly under the banner of Construction Grammar, and with support from much research in child language development (e.g. [Bates and Goodman, 1997]). It is suggested that there is a ‘lexicon-syntax continuum’. The lexicon can contain items of varying complexity, from simple words to whole memorized sentences (or perhaps even the whole memorized Koran). Many conventional idioms and proverbs are stored as wholes, rather than being productively generated. All such stored items are known as ‘constructions’; a word is a construction, a whole memorized sentence is a construction. Crucially, constructions may also vary in flexibility or abstractness. A certain idiom, for example, may not be completely rigidly specified, but may appear in different permutations. The idiom *kick the bucket* can be modified for past or non-past tense, so we can have both *kicks the bucket* and *kicked the bucket*. The idiom is stored as a whole, but with a variable slot for specification of tense. Somewhat more flexible are ‘syntactic idioms’ such as *take advantage of*. In this construction, the verb *take* and the noun *advantage* are in a constant verb-object syntactic relationship, and this can interact with other constructions, as in the passive *Advantage was taken of John* or *John was taken advantage of*. Putting it briefly, humans have evolved a capacity for storing building blocks of various sizes and of varying degrees of flexibility, and a capacity for combining them with others. The first building blocks ever used were small and inflexible. Later, somewhat larger and/or more flexible units were invented and re-used if they proved useful. The initial step from non-combining to combining is still an inevitable qualitative shift, but it did not immediately give rise to an explosion into the extreme productivity of modern languages. In a story of how complex human languages emerged, it is plausible that the very earliest combinations were of the simplest and least abstract items, like the *Me Tarzan* and *You Jane* of the movies. The evolution from that early stage to modern complex languages was a gradual co-evolutionary process, involving cultural invention and re-use of progressively more complex and flexible stored forms, accompanied by biological expansion of the available mental storage space and speeding-up of the possible online combinatorial processes.

Thinking of syntax from an evolutionary point of view prompts a revision of a central tenet of generative theory, namely the relationship between competence and performance.

Competence is the specifically grammatical knowledge in a speaker’s head that allows him to produce and interpret complex sentences. Competence in a language is always there in the speaker’s head, whether it is being used or not. It is thus ‘timeless’. Performance, on the other hand, is the actual processes, located in time and space, of production and interpretation. Some such distinction is indispensable, like the distinction between a computer program and its running at different times, with different inputs and outputs. The focus of generative theory has always been on competence, with performance factors such as limitations on memory and speed of processing being relegated to the status of distracting noise. A central
example is the case of centre-embedded clauses, which I will explain with some examples. Separately, all the following expressions are grammatical sentences: The mouse the cat caught died; The cat the dog chased escaped; The dog the man kicked howled; The man I saw laughed. They can be acceptably combined to a certain extent, so that The mouse the cat the dog chased caught died can, with suitable context and intonation, just about be understood. Nevertheless this last sentence clearly puts strain on the language processor. Further combinations, such as The mouse the cat the dog the man I saw kicked chased caught died are impossible to process without paper and pencil, or concentration on the written form. Generative theory has always held that since such complex examples are formed by exactly the same rules as simpler examples, they must be within a speaker’s competence, though it is admitted that they are outside the limits of his performance. In short, generative theory has resisted any quantitative or numerical element in competence. Quantitative limitations belong to a theory of performance, not to a theory of competence. This is a coherent and understandable theoretical position. But from an evolutionary point of view, it is not possible to see how a capacity to acquire competence in a language can ever have been separate from a capacity for production and interpretation of the objects defined by that competence. Twin interdependent capacities, for internal representation of the permissible regularities of a language (competence), and for putting that knowledge to use on specific occasions (performance) must have co-evolved. Competence without a capacity for performance would have had no impact on the world, so no evolutionary advantage, and complex performance could not happen without a complex internal program to guide it. Hurford [2011] develops this idea with a construct of ‘competence-plus’, a package of rule-like representations combined with numerical limits on their application, for instance limits in depth of embedding. In the evolution of mankind, there was parallel linked growth of the possible complexity of internal representations of the regularities of a language and the quantitative limits on what could be produced or interpreted.

Larger and more flexible constructions can be advantageous to their users, both speakers and hearers, allowing more versatile and less clumsy communication. Complex syntax is especially advantageous when it is compositionally linked to semantics, that is when the meaning of a complex expression is a function of the meanings of the parts and the way these parts are put together. Complex syntax is easier to process when one is able to interpret it as meaningful. Human parsing of complex sentences is a process of deciphering the sequence of words into a representation of some propositional content, plus some indication of the pragmatic intent of the speaker. Parsing uses clues from the words themselves, from markers of grammatical structure in the sentence, and from the whole situational context of the utterance. Long paradoxical or nonsensical strings of words are less easy to parse than meaningful ones of the same length. The evolutionary driver of the modern human capacity for complex syntax was surely the semantic carrying power of complex sentences.

In the history of modern languages we see a process of ‘grammaticalization’
Unless contaminated by contact with other languages, there is a tendency, in all languages, for common patterns in discourse to become entrenched, or conventionally fixed, not just as possible ways of expressing certain meanings, but as required ways of expressing meanings. In English, for instance, every non-imperative sentence must have a subject. A string such as just came today, with no overt subject, is not a full or proper sentence, although it could be perfectly understandable in an appropriate discourse. In other languages, this is not a requirement, as for example e venuto oggi is grammatical in Italian. But even in languages like Italian, so-called null-subject languages, there is a grammatical indication of an understood subject, in the form of the agreement of the verb. It is widely held that grammatical subjects in languages are fossilized topics. Some form of topic-comment structure is universal in languages. All languages have a way of marking the expression denoting the thing that is being talked about (the topic of a sentence), as opposed to what is being said about it (the comment, or focus). Some languages do not mark the subjects of sentences at all, and some only mark them optionally with few specific markers. In English and many other languages, verbal agreement singles out the subject of a sentence. In many languages the relationship between agreement inflections on verbs and subject pronouns is transparent, suggesting a diachronic process of grammaticalization of subjects from pronouns in an overtly topic-comment structure, as in That guy, he’s crazy [Givón, 1976].

I have singled out the grammatical role of subject because it is a centrally grammatical notion, as opposed to a semantic notion like agent (the ‘doer of the action’) or a pragmatic notion like topic (what a speaker assumes is shared information). The grammatical role of subject has emerged, by grammaticalization from the non-grammatical discourse-structural function of topic repeatedly and independently in the histories of many languages. Many other widespread aspects of the grammars of modern languages, such as their typical inventories of syntactic categories (‘parts of speech’) have also arisen through grammaticalization. Heine and Kuteva [2007] survey a wide range of languages and give many examples of particular parts of speech and grammatical markers arising historically from other word-classes. For instance, prepositions commonly arise from bodypart nouns, such as back; auxiliary verbs arise from main verbs (e.g. have); relative clause markers often derive from demonstratives (e.g. English that) or from question words (e.g. which). The process of grammaticalization is overwhelmingly unidirectional, and so it is plausible to reconstruct earlier stages of human languages as lacking the grammatical features produced historically by grammaticalization. Indeed, the few modern cases which come close to genuine language creation de novo, such as Nicaraguan Sign Language, show such bare featureless properties in their early stages, with no or few grammatical markers. Very quickly, however, grammaticalization processes kick in and languages soon develop grammatical structure characteristic of modern languages. The grammatical complexity of modern languages is a historical product, enabled, to be sure, by a biologically endowed capacity to manage such complex systems with facility.
Some degree of syntactic complexity exists in nature without any compositional syntax. The songs of some whales [Payne and McVay, 1971] and many songbirds (see, e.g., Todt and Hultsch, 1996; 1998) are hierarchically structured into what can naturally be called phrases, but these ‘phrases’ make no meaningful contribution to the overall meaning of the complex song. As far as we know, an entire complex birdsong functions either as an invitation to mate or as a warning to keep away from the singer’s territory. Birdsong is syntactically complex, but carries no meaning that is a function of the meanings of its constituent notes and phrases. Indeed, the notes and phrases have no meanings. It has been suggested, by thinkers as illustrious as Rousseau, Darwin and Otto Jespersen, that pre-humans possessed some capacity for such syntactically complex song before it became a vehicle for the expression of messages composed from the meanings of the parts. This is possible, but a problem with the story is that we find no such complex syntactic behaviour in species closely related to humans, in particular in apes and almost all primates, with the possible exception of gibbons.

Some monkeys and apes do string a few meaningful elements together to make sequences that are also meaningful, but the meanings of the whole strings are apparently not a function of the meanings of the parts. For instance, a species of monkey observed in the wild by Arnold and Zuberbühler [2006] has two alarm calls, one for eagles and one for leopards. A combination of these two calls seems to function as a summons for, or comment on, unpanicky movement of the group to another location. It is not clear that the meaning of the two-element combination (roughly ‘all move’) is a function of the meanings of the two parts (roughly ‘eagle’ and ‘leopard’). Truly semantically compositional syntax occurs only in humans, and humans have taken it to a high order of complexity.

Summarizing the evolutionary view of language structure, the human language capacity, especially the capacity for massive storage of constructions large and small, with greater or lesser flexibility and combinability, and the facility for recursively combining constructions fast during speech production, and disentangling them fast during speech perception, were selected because of the advantages of carrying propositional information. Speakers capable of greater fluency benefitted individually, by gaining prestige. Groups containing such speakers, and hearers capable of understanding them, prospered because of the advantages of communicating informatively. The complex structures of individual languages evolved historically over many millennia through such processes as the self-organization we have seen in phonology and grammaticalization in syntax. An evolutionary approach to the language faculty and to languages asks ‘How did they get to be that way?’ I hope to have shown that there are some answers well worth considering.

BIBLIOGRAPHY


Linguistics from an Evolutionary Point of View


LINGUISTICS AND GENDER STUDIES

Sally McConnell-Ginet

INTRODUCTION

Gender studies is an interdisciplinary field of inquiry that draws on philosophy, anthropology, political science, history, sociology, psychology, biology, science studies, literary and cultural studies, queer studies, and many other areas, including linguistics. Its subject matter centers on gender: the sociocultural, political, and ideological dimensions of sexual categorization and sexual relations. As this formulation suggests, matters of sexuality are important for thinking about gender, but it is also important to think about sexuality on its own terms. Sexuality is a broad cover term I use to include sexual orientation, sexual practices, and sexual desires. When I speak of gender studies, I have in mind gender and sexuality studies.

How are gendered and sexualized individual and social identities constructed and changed? How are biology and social life intertwined in these processes? What kinds of diversity do we find among those who claim or are ascribed the same gender or the same sexual orientation? How do gender and sexuality interact with class, race, religion, economic status, age, nationality, and other dimensions of social difference? How do they enter into social and power relations? How do they relate to cultural ideologies and values, to default assumptions about ourselves and others? What moral issues are raised? Such questions are central to gender studies. Their answers involve many complex factors far beyond the realm of linguistics, but there are nearly always also significant matters of language and its use involved.

Gender studies raises many challenging questions about language and about linguistic inquiry.

- Which identities does language index and how?
- Where do linguistic diversity and language change fit in a model of language as part of universal human biological endowment?
- How well can speech acts be understood in individualistic terms?
- How can available linguistic resources affect individual cognition and joint social pursuits?
• How are linguistic forms invested with semantic content? With other kinds of communicative significance?

• How and why do substantive conflicts of interest sometimes play out as ‘mere’ disputes over meaning?

While not organized around these questions, this chapter explores some of the research on language, gender, and sexuality that have stimulated my own thinking about them.

1 INDEXING IDENTITIES

When linguists first began thinking about gender issues in the early 1970s, the focus was on what was then called women’s studies. Linguists began exploring two main topics. One was so-called ‘women’s language’ — ways of speaking supposed to be distinctively ‘feminine’, patterns that indicated or indexed identity as a woman. The other was ways of speaking about women — e.g., what were described as ‘euphemisms’ like lady and ‘dysphemisms’ like broad. American linguist Robin Lakoff [1975] offered this division into speaking by and speaking of and made a number of impressionistic claims that others sought to test more systematically.1 Of course, neither Lakoff nor others who began these investigations thought that gender was only about women, but the impetus for much early gender studies work in linguistics came from second-wave American feminism. Emphasis thus was on the impact on women of social norms for their speech and of the ways others spoke of them — and ‘women’ were mostly white middle-class Americans.

Heterosexuality was also widely presumed in these studies — sexual minorities were mostly ignored. Yet Lakoff did suggest that gay men (and male academics!) might speak ‘women’s language’ to index not a feminine identity but a rejection of the power-seeking stance she associated with (straight non-academic) men. And by the 1980s there began to be investigations of so-called ‘gayspeak’, linguistic practices that (supposedly) indexed gay (male) identities, as well as of labels applied to sexual and gender minorities.2

Both gender and sexual identities were often taken as given, a matter of what someone unproblematically ‘is’. Although not put this way, gender identities were often thought of as somehow flowing automatically from genetic, genital, and hormonal facts, with linguistic studies aiming to uncover speakers’ (inevitable) indexing of their identities. Lots of testosterone, lots of swearing; lots of estrogen, lots of high-pitched giggling. Of course no one put things in such simplistic terms, but a picture much like this was lurking in the background, I think, even when there

---

1See, e.g., papers in [Thorne & Henley, 1975; McConnell-Ginet et al., 1980; Thorne et al., 1983; Philips et al., 1987; Coates & Cameron 1988]. For more recent discussion of Lakoff’s ideas, see [Lakoff, 2004].

2Chesebro [1981] is an early collection; Leap [1996] is the first monograph on gay male speech. Cameron & Kulick [2003, ch. 4], is a useful albeit strongly critical review of much of the earlier work indexing sexual identities; it draws on [Kulick, 2000], which is even more critical.
was dutiful talk of ‘socialization’ and ‘acculturation’ into ‘sex roles’. Boys and girls were expected by others and themselves to grow into men and women respectively, and adults in the children’s environment provided them ‘models’ towards which they ‘naturally’ aimed, adjusting their speech to be ‘gender-appropriate’. Evaluations of absent friends’ clothes, gossiping girls; replays of yesterday’s big game, guys on the scene. And on and on, from matters of syntax (women speak more ‘correctly’) to intonational patterns (women speak more ‘dynamically’), conversational organization (men rudely interrupt, women politely take turns) to speech acts (men authoritatively assert, women hesitantly seek confirmation). Never mind that real people often deviated considerably from these generalizations (even in the mid-20th century but more and more as gender arrangements in Western industrial societies changed during the last few decades), the sex-difference bandwagon attracted many.

The implicit idea seemed to be that just as kids ‘automatically’ acquire the language spoken by caretakers and others with whom they interact regularly as they mature — you’ll speak Burmese if you’re growing up in a local family near Mandalay, Finnish if you’re a local kid near Pori — so girls acquire women’s ‘ways of talking’ and boys acquire men’s. Rather than different languages, the (better) analogy often used was (more or less) mutually comprehensible regional dialects or ‘accents’ — English spoken in Casper, Wyoming differs from English spoken in Te Anau, New Zealand, and which variety a speaker uses indexes their geographical origins, early contacts. And of course we can cut things finer: in any region there is variation linked to socioeconomic status, which often connects to neighborhoods and other social groupings. The explanation might again be contact, only of course frequency of contact is involved rather than simple contact per se.

If there are just essentially automatic processes linked to frequency of what one hears, however, there would seem to be something of a puzzle for understanding supposed gender differences. After all, children often/usually grow up with lots of input from adults and other children of both sexes. Eckert [1990] pointed out quite explicitly that the geographic and social class models of linguistic variation really cannot be expected to work for ‘genderlects’. Women and girls in general don’t live together: they are distributed across regions and across socioeconomic groups. Any transregional, transclass sex-linked ways of speaking must arise in somewhat different ways from regional or class-based varieties.

It is also important that part of gender is about heterosexual assumptions: most kids are reared to expect that they will eventually partner with someone of the

---

3The sex-indefinite they, frowned on by those who educated me, was used widely until about two and a half centuries ago, when grammarian Ann Fisher [1745] declared that its plurality rendered it unsuitable in singular contexts and decreed that he could and should be used in singular contexts that might involve women or girls. Boon-Tieken van Ostade [2000] discusses Fisher’s influence; Boon-Tieken van Ostade [1992] offers reasons for not attributing this proposal to Kirkby as most earlier writers have done. Good writers did not immediately stop using singular they (Jane Austen, Lord Byron, Charles Dickens, George Eliot, to name a few), but gradually school-teachers and editors succeeded in getting it out of most published writing and ‘careful speech’. Many people have continued using it in speech, and its frequency is increasing in edited writing. I will discuss attempts to use language more ‘inclusively’ below.
other sex. They are not tuning out the other sex. One of the more plausible attempts to account for supposed ‘genderlects’ was first offered by anthropologists Daniel Maltz and Ruth Borker [1982] and later popularized by Deborah Tannen [1990] and further extended by less linguistically sophisticated writers like John Gray [1992]. The idea was that girls and boys grow up in somewhat different ‘subcultures’ and that it is these different experiences that lead them to speak in somewhat different ways as women and men.4 Certainly single-sex peer groups can play an important role in forming linguistic habits and expectations, but they are almost always only part of the story.

This picture showed its deficiencies as a comprehensive account of how language might index gender identities once it was widely acknowledged that gender groups might not be monolithic. Not only is it very different growing up female in Madagascar5 rather than in Southern California and in the early 21st century rather than in the mid-20th century, but even at the same time in the same region, socioeconomic class, and ethnic group there can be significant differences among girls/women and among boys/men. Such diversity has led to investigation of the distinct linguistic manifestations of different femininities and of different masculinities rather than of some single femininity or single masculinity.

And understanding how differences in adult sexual orientation might be indexed by features of speech provides an even greater challenge initially. Neither adult models nor peer subcultures whose ways of talking children might acquire through exposure would readily explain how a boy who might eventually ‘come out’ as a gay man could become fluent in ‘gayspeak’. Smyth and Rogers [2008] argue that there are vocal features that create a way (or ways?) of speaking that English listeners’ interpret as indexing a gay male identity, which, they suggest, share much with women’s vocal features. But they don’t really want to claim that boys who will become gay men ‘naturally’ model their speech on that of women rather than men nor that they are imitating adult gay men in their environment. Rather, they suggest, sounding gay links to gender non-conformity and social affiliation with girls and women.

4The further claim was that just as people from different regions or ethnic groups are sometimes unaware of some kinds of differences in communicative practices and thus misinterpret one another as insightfully discussed by John Gumperz [1982] and colleagues in their work on cross-cultural (mis)communication, so men and women ‘miscommunicate’ because of genderlectal differences of which they are ignorant, given the different ‘cultures’ in which they grew up. Maltz & Borker, who were students of Gumperz, explicitly draw on his model. But, as noted above, gender difference is very unlike regional, class, or ethnic difference. People expect (often wrongly) that those of the other sex will act differently from them, including speak differently: they are far more likely to attribute difference where there is sameness than the converse. And, of course, people typically observe and think about people of the other sex, whereas this is by no means always true with other kinds of social difference. As Cameron [2007] succinctly and wittily details, it is comforting but ultimately highly problematic for people to interpret gender conflict as miscommunication.

5Ochs Keenan [1976] is a widely cited account of gender differentiation among Malagasy speakers in Madagascar, which documented Malagasy-speaking women engaged in the ‘direct’ and sometimes confrontational styles often attributed to men by those investigating Anglo-American gendered styles, and men preferring indirection and avoiding direct clashes.
Given the considerable evidence that everyone manipulates voice and other aspects of linguistic style, Smyth and Rogers do not subscribe to the idea that there might be aspects of one’s genetic endowment that both dispose someone to seek same-sex partners and affect the voice in some of the ways heard as ‘gay’. No one denies that phonetic properties of speech are affected by the physical make-up of the speaker. Children’s voices can usually be identified as such as can voices of the very elderly, and, by and large, women’s and men’s voices can be reliably distinguished. In males at puberty, vocal cords thicken and lengthen, leading to lower (average) pitch ranges for male compared to female voices. Interestingly, however, girls and boys tend to sound different even before there is any anatomical reason for vocal differences. There is also considerable evidence that pitch ranges and average pitches for women and for men overlap considerably and that pitch and other features of voice quality vary cross-culturally and, for an individual, situationally. Sexual arousal also affects voice quality in various ways yet, as Cameron and Kulick [2003] point out, citing the hilarious scene from the film “When Harry Met Sally” where Sally vocally fakes an orgasm while sitting at a restaurant table, those vocal effects can be intentionally manipulated. Similarly, Hall [1995] is a fascinating account of some people who perform as heterosexually attractive women on a phone sex line: off the job they present themselves very differently, some being lesbians and one a bisexual male. They are able to do this because certain ways of talking (breathy voice, e.g.) are interpreted, ‘heard’, as indexing female heterosexual desirability (and attraction to the male caller). This could be ‘natural’ meaning (potentially) transmuted into ‘non-natural’ meaning, which can then be exploited for various purposes. In the case of the so-called ‘gay voice’, there is no evidence of any biological or ‘natural’ basis: what is clear is that some men whose voices are heard as gay identify as straight whereas many self-identified gay men are heard as straight.

Nonetheless, the biological compulsion, ‘essentialist’, story of gender and sexual identities and the speech styles indexing them continues to loom large in popular thinking. Our genetic endowment does indeed constrain the ways we can speak and perform other kinds of actions. We are born disposed to be certain kinds of people and not others. But there is considerable evidence that there is more diversity in ways of speaking (and in all sorts of other kinds of characteristics and activities) within each sex than between the sexes. And we also know that people of the same sexual orientation vary enormously. Everyone operates within a range of genetically determined possibilities: their experience determines which of these possibilities get realized. Experience includes not only material circumstances, treatment by others, and ideologies influential in the environment, but also people’s own active shaping of their lives, their interpretation of possibilities for themselves.
and others, and their strategic choices and interpretation of responses to those choices.

Gender labels are indeed ascribed early in life, accepted as applying by young children, and reascribed by strangers throughout life. In this way, gender identities are more like racial or ethnic identities than like identities centering on sexual orientation — it’s practically impossible to escape them in some form or other. For almost everyone bodily sex is experienced not as chosen but as given, innate; trans women and men are not an exception to this as they typically take their gender identities to be innate but at odds with the sex class to which they were assigned at birth. For many people, sexual orientation is also experienced as fixed, although this is apparently more true of men than women. Still sexual orientation, which may well be unknown to others and is seldom in play at all before adolescence and often not until much later, is by no means always interactionally salient whereas gender categorization is always readily accessible to interactants and very often made salient.

What exactly is made of the gender categorization, how it meshes with other aspects of the identities a person develops, is, of course, extremely variable. Sexual identities are even less homogenous. Indeed, sexual orientation is not in all times and places even considered an identity component: as Foucault [1981] notes, until the late 19th century engaging in homosexual activity was just that, something one did like taking long showers or going to bars. It was often frowned on but it did not automatically get you classified as a certain kind of person, it did not confer an identity. Being straight/heterosexual is still not generally considered an ‘identity’ — it’s often seen as just being a ‘regular’/‘normal’ person. Not surprisingly, many gay men and lesbians protest the assumption that their sexual orientation in itself makes them distinctive ‘kinds’ of people.

The sexual division creating gender classes, however, always seems to constitute an identity component. Sex-class is often seen as the most fundamental or basic difference among people. Indeed, as Helen Haste [1994] observes, sexual difference stands proxy for all sorts of other differences, serving as difference par excellence (consider the French expression, vive la différence). Most languages incorporate sexual distinctions in terminology for referring to, addressing, and describing people. Some languages go even further with markers indexing sex of conversational participants that occur in almost every utterance — Japanese, e.g., is often described as such a language (though the picture is much more complicated than language textbooks or popular understanding might indicate.). Nonetheless, the

---

9Some people feel they do not belong in the sexed body into which they were born and undergo hormonal and sometimes surgical treatment in order to move into the other sex category. Whether or not they would feel this way if there were not such pervasive gender ideologies that try to constrain people’s actions, attitudes, and aspirations, we may never know. Bornstein [1995] is a fascinating exploration of gender by a trans woman (a MtoF transsexual), and Bornstein [1998] is a thought-provoking student resource. In a videotaped interview, Kate Bornstein recounts the speech instruction she received as she was transitioning. Imitating the breathy and swoopy speech her instructors modeled for her, she says “But I didn’t want to be that kind of woman.”

idea that gender identity somehow floats free of other components of identity and of situations and social practice as suggested by rubrics like ‘women’s language’ is deeply problematic. Identities are, as we shall see, primarily indexed at somewhat more local levels, where gender is intertwined with other identity components. And identity-indexing is just the tip of the sociocultural iceberg that linguists concerned with gender and sexuality studies have begun to explore.

2 SOCIAL MEANING

It is hard to understand how identities might be indexed at all if we only think of language in the individualistic and static structural terms provided by most varieties of theoretical linguistics. It is not, e.g., that different social identities somehow result in different grammars being represented in individual minds. Whatever is going on does not seem to involve basic linguistic structures but rather patterns of language use as well as language ideologies — ‘ladies’ don’t swear, ‘macho’ guys don’t apologize. And actual usage patterns as well as ideologies keep shifting: analysts interested in the social significance of language have to aim at a moving target, which is methodologically very challenging.

Language does indeed provide many resources that speakers can exploit to index their own and others’ identities, but this is only one aspect of what is sometimes called social meaning. Social meaning is involved when features of linguistic communications trigger inferences about communicators and their attitudes and relationships, both long-term and at the moment of communication. Social meanings are generally not part of either the literal truth-conditional content of the expression uttered or, in some cases, even of what the speaker means, what is implicated (or of explicatures as in relevance theoretic approaches). Although social meanings can be linguistically encoded and highly conventionalized — e.g., honorifics in languages like Japanese\textsuperscript{11} — they certainly need not be.

Social meanings generated by my face-to-face utterance of Nice day, isn’t it could include, depending on just how I say it (and also my appearance and other features of the situation in which I utter it): I am an English-speaking American woman of European ancestry, over sixty, well-educated and relatively affluent, friendly, interested in interacting with you, etc., etc. Or maybe rather than friendly you interpret me as pushy or phony, rather than well-educated and relatively affluent as affected or stuck-up. Even unwarranted inferences could be considered part of social meaning if indeed they are regularly made. If I’ve drunk a little too much wine, you may infer that fact from certain phonetic features of my utterance and you may also then detect the influence on my vowels and speech tempo of a childhood in the southern US. And so on.

\textsuperscript{11}Potts & Kawahara [2004] draw on the two-dimensional semantics outlined in [Potts, 2004] for an account of honorifics that does make them similar to conventional implicatures. Although this is an interesting suggestion and honorifics, unlike some other carriers of social meaning, certainly are governed by linguistic conventions, I don’t think this approach gets at what is distinctive about them.
Notice that there seem to be many possible social meanings and none of them need be ones I intend to convey though some may be. Others, however, I might hotly protest and others might just embarrass me. At the same time all seem to have me (including my attitudes toward my interlocutors and others) as topic. Social meanings situate me in a social landscape, sometimes in neighborhoods from which I want to dissociate myself.

I have sometimes tried to avoid speaking of social ‘meaning’ because both the kind of significance involved and its role in communication are very different from what those working in semantics and pragmatics usually include under ‘meaning’. The phrase is so well established within sociolinguistics, however, that avoiding it tends to confuse rather than clarify. What is lacking in the literature is sustained discussion of what it takes for an assignment of social meaning to some linguistic feature to be ‘warranted’. This does not mean that social meanings are simply drawn from thin air. Podesva [2008] notes that investigators have used ethnographic data, historical evidence, and experimental techniques to show how social meaning gets associated with single features or even whole varieties. He himself considers the role of discourse in creating social meaning, arguing that variants occur “where their meanings are indexed in interaction… interactional stances give social meaning to linguistic features, and in turn linguistic features help to establish interactional stances… commonalities in the discourse contexts in which given variants occur … [are] their social meanings.” But, useful as Podesva’s discussion is, how it is to be applied to assessing claims that variants carry social meanings like ‘femininity’ or ‘gayness’ is not clear. And what is the role of hearers’ perceptions?

Kulick [2000], e.g., is highly critical of analysts who dub certain linguistic practices as carrying the social meaning of gayness when those features are neither restricted to gay speakers nor characteristic of all (or even, arguably, most) gay speakers. And similar objections were earlier raised to claims about ‘women’s language’ (by me and many others). But we certainly can say that an adult male lisp in American English carries gay identity as a social meaning in that hearers regularly use it to infer gayness (even though they may often be mistaken) and speakers can successfully use it (in many contexts) to signal (perhaps simulate) gay identity.12

The linguistic system itself often includes formal alternates whose main difference seems to be indexical — i.e., indicative of who is involved in the linguistic interaction, their attitudes, their understanding of their situation. Bertrand Russell’s famous ‘emotional conjugation’ — ‘I am firm; you are obstinate; he is a pig-headed fool’13 — nicely illustrates indexical possibilities of choice among words that seem to be (more or less) informationally equivalent. Relevant for gender and

---

12 In *Naked*, comedian David Sedaris talks about encountering all the other (ultimately) gay boys at the speech therapist’s office in elementary school, sent there to get their lisps corrected.

sexuality are sets like lady/woman/broad/... or gay/homosexual/fairy/.... Pronouns are often indexically significant for gender and sexuality (as well of course as for their primary marking of relation to the speech situation): what Brown & Gilman [1960] dubbed the T/V distinction for 2<sup>nd</sup> p(erson) pronouns in most Indo-European languages (e.g. French tu and vous), 3<sup>rd</sup> p pronouns in contemporary English (we’ll discuss them more later), 1<sup>st</sup> p in such unrelated languages as Japanese and Burmese. Personal names, kin terms, and other forms for address and reference are richly indexical, indicating personal and family histories as well as relationships and frequently generating inferences connected to gender and sexuality. And of course morphosyntactic choices may be indexically relevant. “I ain’t seen nobody” vs “I haven’t seen anybody” is heard as indicating education and class levels, but can also index relative formality of the speech situation or a speaker’s stance toward classroom prescriptions, which in turn may connect to gender ideologies.

Much indexical work is accomplished, however, below the level of content-bearing linguistic form: variationist sociolinguistic research as pioneered by William Labov [1966; 1972a; 1972b] has emphasized linguistic units like vowels or consonants or affixes that have a range of variant pronunciations. Labov formulated the notion of the linguistic variable, any linguistic unit with alternative realizations or variants that are equivalent (at least for certain purposes) in their contribution to the informational content of what is said. So, for example, the verbal suffix –ing can be pronounced with an alveolar nasal, indicated as a ‘dropped g’ by the written form –in’, or with a velar nasal, the pronunciation associated with the standard spelling. One can then look at frequencies of occurrences of different variants in some corpus and investigate systematically not only possible indexing of identities but other social significance that might attach to a variant, the social meaning of a variant. It becomes harder when one looks at lexical or syntactic variation or, even further removed from phonological variation, code-switching from one distinct language to another (or from quite different varieties of the same language — e.g., a local ‘dialect’ and a national standard), to claim that alternates are completely synonymous and also to identify sites of potential occurrence of a variant. Nonetheless, the notion of a linguistic variable has been very useful for allowing sociolinguists to systematize their investigations of linguistic variation. Over the past four decades variationist sociolinguistics has made enormous strides in clarifying some of the structured regularity of what earlier had

---

14Lavendera 1978 made these points but alternative methodologies are thin on the ground. It can often be useful and not too misleading to treat alternates as rough equivalents. What is perhaps most difficult is arriving at a workable notion of ‘possible occurrence’ for units larger than speech sounds. What generally happens in considering the social significance of a tag (The war in Afghanistan is awful, isn’t it) is that investigators simply look at number of tags compared to something like overall amount of speech (how many tags per 1000 words, e.g.).

Code-switching — alternating between distinct languages that are potentially mutually incomprehensible — is even less amenable than variation within a single language to standard Labovian variationist methodology, but it too can be studied systematically. In multilingual settings code-switching may happen without conscious intent, and, at the same time, carry considerable communicative significance.
Sally McConnell-Ginet seemed just ‘noise’ (‘free variation’ as the Bloomfieldians called it) in language production and in relating variation to linguistic change, especially but not only sound change.

Labov distinguished three kinds of variable, which he called INDICATORS, MARKERS, and STEREOTYPES. Indicators correlate with demographic characteristics — e.g., sex, socioeconomic status, religion — but language users do not seem to attach any social significance to these correlations. That is, their use does not seem to vary stylistically (for Labov, style was just a matter of the amount of attention being paid to speech) or in any way by social situation or discourse purposes. Indicators are devoid of social meaning beyond simple indexing (and speakers seem unaware of the indexing or at least they do not exploit it in any way). In contrast, social markers clearly carry social significance and show stylistic variation, perhaps even being ‘corrected’ on occasion though basically they seem to operate below the level of conscious attention. Full-blown stereotypes are the topic of overt discussion and are often stigmatized. But these distinctions do not get us as far as is needed in thinking about the social meaning of variation. What is more problematic, they assume that the sociolinguistically relevant categories are pre-formed and static. Eckert [2000], however, argues that social categories are shaped and often transformed in the course of ongoing social practice, and the path from indicator to marker to stereotype goes through social practice and cultural ideologies.

Speaker sex does often correlate with frequency of particular variants. Research in a number of western urban settings shows women using more standard grammar yet also leading in sound change in progress, using more ‘vernacular’ variants (i.e., less ‘standard’ pronunciations), a situation that Labov [1990] dubs the ‘gender paradox’. Grammatical variants may more often be stereotyped, whereas sound change in progress is far more likely to be a marker (at the earliest stages, perhaps only an indicator) and not so clearly ‘noticed’. Such observations, however, don’t take us far. Of course there is no gender paradox at all once one recognizes that these variants can have quite different social significance, play a different role in social life. Just because use of a variable correlates, say, with female speaker sex, we cannot conclude that its social significance is femaleness or femininity — linguistic indexing of a large-scale demographic category is almost always INDIRECT, 15 me-

15Ochs [1991] introduced the terminology INDIRECT INDEXING in discussing the fact that rather few features of talk directly indicate gender class — i.e., simply mean femaleness or femininity, maleness or masculinity. What is more typical, she suggested, is that there will be ways of talking that indicate politeness and being polite will figure in gender norms. The idea without the terminology is already present in [Lakoff, 1973; 1975], which argued that certain ways of talking indicated powerlessness or unwillingness to assume responsibility and that through this association those ways of talking were part of what she called ‘women’s language’. And, though McConnell-Ginet 1978, 1983 does not endorse the idea of ‘women’s language’, I did argue there that certain prosodic patterns are associated with strategies and stances that may be more important for women than for men, given social settings in which male dominance and certain traditional gendered divisions of labor figure prominently. Both Lakoff and I were in fact proposing indirect indexing of gendered identities, though neither of us had a well developed view of the complexity of identity construction.
diated through other kinds of more specific (and often also localized) significance or meaning attached to variants. It might be resistance to institutional authority, claims to local authenticity and knowledge, friendliness or personal warmth, education or religiosity — and it might be some combination of these. Identities are internally complex, and there is a wide range of femininities and masculinities and of sexual identities.

The study of the social meaning of linguistic variation has developed considerably since its roots in early Labovian correlational sociolinguistics. Sociolinguistics at the end of the 20th and beginning of the 21st centuries moved beyond the earlier focus on correlations with static preexisting demographic categories to an emphasis on the activities and practices that are critical not just for initial social category definition but for sustaining and transforming categories. The work of my coauthor, Penelope Eckert, and her students (e.g., Sarah Benor, Kathryn Campbell-Kibler, and Robert Podesva) is exemplary in this respect, emphasizing complex and sometimes competing significance attached to use of particular linguistic variables. The shift is away from indexing of macro-level identities like being a woman or being gay and toward much more specific associations with speakers’ discursive positioning and their use of an array of linguistic resources to create their styles.

Eckert [2008] develops the idea that the meanings of sociolinguistic variables “are not precise or fixed but rather constitute a field of potential meanings — an indexical field, or constellation of ideologically related meanings, any one of which can be activated in the situated use of the variable.” (453) So, e.g., release of a final /t/ in American English can be thought to indicate relatively enduring traits — being articulate, educated, elegant, prissy — or to indicate more fleeting stances — being formal, careful, emphatic, angry. Of course even momentary stances can figure in identities if someone positions themself or is positioned by others as habitually assuming that stance — as characteristically careful or angry, for example. This final /t/ release variable figures in what Eckert calls PERSOA STYLE, ways of talking that index a social type, a position in the social landscape. Of course /t/ release can (indirectly) index quite distinct social types: she mentions school teacher, British, nerd girl, gay diva, Yeshiva boy. What particular meanings attach to the variable and which identity they help index depends on the setting in which the variable is realized, including other features of the speaker’s style.

Persona style involves, of course, far more than alternative ways of saying ‘the same thing’: not only can the kinds of things one talks about be important in constructing particular social types (the global economic crisis, the kids’ upcoming exams, a home renovation project, prospects for the Chicago Cubs [American

16Bucholtz and Hall [2005] cite [Rauniomaa, 2003] for the termSTANCE ACCRETION to indicate this possibility of what might seem fundamentally situated properties becoming persisting traits that contribute to an individual’s identity.

17For /t/ release in the speech of nerd girls, see [Bucholtz, 1999; Podesva, 2006; 2007] is source for gay diva; [Benor, 2004] for Yeshiva boy.
football] or Arsenal [English] this season) but also what one wears, one’s leisure activities, the music one listens to, and even what one eats and drinks.\textsuperscript{18} And a single individual may construct very different social types in different contexts. Podesva [2006; 2007; 2008] looked at the speech of a gay man, Heath, as he interacted in a medical clinic with patients as a ‘competent and caring physician’ and as he interacted with close gay friends at a backyard barbecue as ‘gay diva’. There was a higher incidence of /t/-release from Heath as physician in the clinic, indicating his educational level and his carefulness, Podesva suggests. But the actual phonetic manifestations of /t/ release when it did occur in the barbecue setting involved longer and stronger bursts than those in the clinic: Podesva suggests they were part of Heath’s construction in a social setting with good friends of a flamboyant ‘gay diva’ persona, a parodic ‘prissy’ effect.

Where do such meanings come from? They often arise from associations with the speech of certain people engaged in certain social practices: attitudes not just towards those people but also towards those practices are centrally important in establishing social meanings. Sometimes there seem to be ‘natural’ meanings that are further developed and consciously exploited. For example, being emotionally engaged in an interaction can increase prosodic dynamism, including extent and frequency of pitch shifts. Prosodic dynamism can then be interpreted as indicating emotional engagement and can be avoided or exaggerated, depending on a speaker’s particular interests and goals, sometimes at the level of conscious attention, often below. Here too other meanings get attached: e.g., Delph-Janiurek [1999] suggests that men (at least in the British university lecturer population he studied) who sound lively and engaged are often heard as gay, or at least as not conforming to expected hegemonic masculine norms (not surprisingly, those who avoid such liveliness are highly likely to be heard as boring).

Social meanings are slippery indeed. I can, e.g., be led to use a slightly rising upward final rather than a falling one when my aim is to invite my interlocutor to comment on what I have just said or to make some other contribution to our ongoing exchange. My interlocutor (or third-party observers) may interpret me as insecure, unable or at least unwilling to voice my views with authority. It is true, of course, that I have passed up an opportunity to show myself as confident and authoritative. But that does not mean I lack confidence and fear asserting myself authoritatively: I really may be trying to draw my addressee into the conversation. Nonetheless others may mark me down as wishy-washy. And, of course, social meanings often attach to features not of individual communicative acts but of communicative histories, of patterns that emerge over many utterances. Frequency can matter. So pronouncing the affix –ing as if it were –in does not in itself ‘mean’ informality — it is a certain pattern or frequency of such pronunciations (together with other speech features) that may have that social meaning.\textsuperscript{19} Comparisons

\textsuperscript{18} Real men don’t eat quiche quipped a humorous 1982 book by Bruce Feirstein about masculine stereotypes, and many a bartender continues to assume that the white wine must be for me, the red for my husband, whereas our preferences tend to go the other way.

\textsuperscript{19} And of course this variable is associated with many other meanings; see [Campbell-Kibler,
with other utterances are often relevant. A lively prosody, relatively high pitch, and ‘smiley’ voice (with formant frequencies raised because of the shortened vocal tract smiles produce) can trigger inferences of phoniness if that style is adopted in a phone conversation with a disliked work superior and dropped immediately after for a face-to-face exchange with a friend. And of course more specific situated discourse history can also matter for what significance attaches to particular ways of speaking.

In our joint work, Penelope Eckert and I have argued for the central importance of studying language, gender, and sexuality in local face-to-face communities of practice. (See [Eckert & McConnell-Ginet, 1992a; 1992b; 1995; 1999; 2003; 2007].) A community of practice (CofP) is a group jointly engaged in social practice around some project or projects (perhaps as loose as enjoying oneself): choirs, lacrosse teams, workplace groups, families, academic departments, classrooms, urban gangs, church youth groups. ‘Jointly’ is key: communities of practice involve orientation towards other members. Arguably, people engaging regularly in social practices with others to whom they hold themselves mutually accountable — within CsofP — give rise to much social meaning. Social meaning also arises, however, in encounters across CsofP and in relations between CsofP and larger institutions: indeed CsofP are central in connecting individuals to institutions and to dominant cultural ideologies.

Podesva [2008] makes a couple of points about the social meanings of linguistic variants and their contribution to socially meaningful styles that are particularly important for thinking about social significance that goes beyond particular local CsofP. Particular variables, he proposes, may have “a kernel of similarity . . . across communities.” He proposes that fortition — e.g., pronunciation of this and that with initial stops rather than fricatives, i.e., as dis and dat — may have something like ‘toughness’ as its core meaning, even though that meaning is elaborated and enriched somewhat differently in different local CsofP that exploit this particular variable. Podesva also emphasizes that variants do not ‘mean’ in isolation but work together with one another and with other components of interaction to produce stylistic meaning, social meaning at a higher level. This emergence of higher-level meaning from assembling meanings of components can be thought of as a kind of ‘compositionality’, perhaps not rule-governed in quite the ways formal semanticists study but importantly depending on some kind of interactions among meanings of smaller units.

In this kind of ‘compositionality’ and in being rooted in social practice, with especially close connections to local CsofP, social meaning is very like what I’ll call (CONTENT) MEANING, the meaning that is the subject matter of linguistic utterances, the basic message, and what scholars of semantics and pragmatics emphasize. I will sometimes drop the qualifier ‘content’ and speak simply of meaning where it is clear what I have in mind. Content meanings, both expression meaning and utterance/speaker meaning, play a central role in both the social and the intellectual dimensions of life. One cannot think about gender and sexuality for
long without confronting this kind of meaning, sometimes overtly and sometimes more subtly.

3 CONTENT MEANINGS (AND THEIR ‘BAGGAGE’) MATTER

We have already noted above that content meanings may play a role in constructing identities. But beyond that, content, both explicit and implicit, is fundamental to individual and collective thought and action. Before discussing content meanings let me just note that there is a kind of content that hearers often infer from what’s said that is not part of expression or of speaker meaning.

3.1 Conceptual baggage

Content is often, as I suggested in [McConnell-Ginet, 2008], accompanied by what I dubbed there CONCEPTUAL BAGGAGE (CB). CB involves background assumptions that, though not part of what expressions or their users mean nor presupposed conventionally or conversationally, are nonetheless made salient and accessible by use of those expressions. Like social meanings, CB can trigger inferences that the speaker does not intend and might sincerely disavow. We might think of it as (potential) ‘hearer meaning’.

Consider this variant of a story told decades ago by feminists seeking to raise ‘consciousness’ of the relative invisibility of women in certain contexts.

A young boy and his father were driving on a mountain road when the car skidded and went plummeting into a ravine. Another car came on the scene and rushed to the rescue. They found the father dead but the boy, though unconscious and bleeding profusely, was still breathing. He was rushed to the nearest hospital and prepared for emergency surgery, the hospital’s chief of surgery having been alerted. The surgeon walked in and gasped “Oh no, I can’t operate — that’s my son.” How could this be?

Very few listeners immediately thought of the possibility that the surgeon was the boy’s mother, a phenomenon that some said showed that surgeon was ‘semantically’ male. Of course, no one would argue that the term surgeon is inapplicable to women nor even that its use conversationally implies maleness, but it could still be that that word itself tends to obscure women from view. My diagnosis is that surgeon often is laden with conceptual baggage that includes strongly male stereotypes, which can make it hard for interpreters to remember the possibility of female referents.

Of course there are also other possibilities, consistent with the surgeon’s being father of the injured boy: perhaps the boy has multiple fathers, two gay men or a stepfather or an adoptive father as well as a biological father. Neglect of these possibilities arises, I think, from conceptual baggage attached to father, which triggers inferences about a child’s having exactly two parents, one male and one female. (One might say that it is the definite his in his father that triggers the uniqueness assumption but consider that his brother would not trigger assumptions of exactly
one brother.) It was Celia Kitzinger’s study of referring terms in calls to a medical practice that first led me to the idea of CB attached to particular words or expressions. Kitzinger [2005] looked at how people who called on behalf of someone else identified the potential patient and the subsequent effects on discourse of the referential form used. Identifying the person about whom one was calling as my husband or my child triggered all kinds of inferences that were not drawn if the person was identified as my roommate or my friend: e.g., willingness to get the person in for an appointment, knowledge of their medical history, responsibility for their ongoing welfare. A woman who called about my child was assumed to have a duty to be with an at-home sick child. These assumptions manifested themselves in what was said in response to the initial caller.

What is of interest here, I think, is that the words themselves seem potentially to trigger certain inferences that cannot be thought of as what the words or their utterers mean. They bring with them CB, which is often heavily laced with gender and sexual ideologies. Such ideologies can also affect communication in other ways, as we will see. What I am calling CB, however, covers communicatively significant inferences that do not arise from any of the standard kinds of ‘meaning’ familiar to philosophers of language. Some linguists have included what I call CB under such rubrics as ‘frame meaning’, ‘connotation’, and other categories. Such approaches are, I think, correct in recognizing the quasi-conventional character of the inferences triggered and their relative independence from speaker’s intentions. But unlike conventionally encoded content, CB is not ‘meant’ by those speaking literally — and they can at least try to disavow it.

3.2 ‘Sexist’ language

Much attention has been paid to cases in which expression and speaker meaning can generate inferences that seem to work against the interests of women or of sexual minorities. American feminists in the early 1970s spoke of sexist language and included under that heading derogatory and hypersexualized labels like bitch, broad, slut, belittling labels for adult women like girl, chick, doll, frequent lexical asymmetries (cp the verbal mothering with fathering or the existence in American English of cleaning lady, unmatched by anything like garbage gentleman), asymmetries in forms of address and reference (titles indicating marital status for women but not for men. e.g.), and much more. A major focus was on so-called generic masculines — he with sex-indefinite antecedents or man to designate humanity.

Many linguists and philosophers of language were somewhat skeptical about the significance of such discussions: the language, they (we) often said, is not the problem, but the uses people make of it. Above, I’ve used words like derogatory and belittling as descriptors of certain labels that some speakers apply (or did apply — as mentioned earlier, language use keeps changing) to women in general. But how can such evaluative terms apply to ‘mere words’? After all, it is not the words as such but the social practices of using them that might derogate or
Even supposing these words are derogatory, skeptics continued, there are also derogatory words that apply to men. What about *prick*, *bastard*, or *asshole*? Indeed English does have plenty of resources for speaking negatively of particular men. Unlike those applying to women, however, they are not routinely used of men in general but are reserved for speaking of specific men. But is this a feature of their conventional meaning? Such a question turns out to be far harder to answer than one might think. Not only does language keep changing so that new meaning conventions arise (as surveying successive editions of good dictionaries makes clear). Even at a given point in time, it can be challenging to try to assess just what meaning(s) is (are) ‘conventionally’ attached to particular expressions. And in some sense it doesn’t matter if the sexism evinced is conventionally encoded in particular expressions. What ultimately seems significant is not individual words but wider discursive practices, patterns of usage that ‘naturalize’ certain sexist attitudes and actions.

In thinking about sexist and homophobic language, I have come to view lexical meaning conventions somewhat differently than I once did. Such ‘conventions’ can shift readily, can vary significantly in different CsofP, and can be contested when the content in question connects to matters in the world that are themselves contested. Even where the dispute is over the range of application of a word, its denotation, rather than over the attitudes conveyed by choosing a particular word to do a referential job (e.g., referring to an adult woman as *girl*), the question of what is conventionally attached to the word can be murky. Is it a feature of the conventional meaning of *marriage* that it can only designate unions of heterosexual couples? If it were then *gay marriage* ought to be seen as incoherent, in the realm of *square circle*, which is not the position taken by most opponents of gay marriage, including most who want to have the exclusion of same-sex couples incorporated into legal definitions of *marriage*. Similarly, is there a linguistic convention that application of *girl* to adult women is condescending? What is clear is that, although particular lexical items may indeed seem problematic, the real issue is discursive patterns of use and the ways of thinking and acting such patterns facilitate.

Consider the convention of using *he* for mixed-sex or sex-indefinite antecedents, now wobbling but by no means dead. Although many people do still learn to follow this convention, it is far less straightforward than it might seem. Similarly
although *man* (used without an article) can designate humanity in general, its use often has the effect of conflating humanity with the male portion thereof, treating femaleness as a marked condition and maleness as the default norm. There is empirical evidence that many people do not reliably interpret such usages inclusively (though some certainly do take them to include women). For example, Martyna 1980 reported women less likely to apply for a job if the job-description used *he-man* language than if the same job were described in more inclusive terms. Putatively generic uses do not always work, probably in part because they are tainted by the masculine uses of the same forms that are so common.

But as Black & Coward [1981] made clear, even language that is by usual criteria conventionally inclusive — words like *villager* or *pioneer* that clearly may designate female referents — can in some discursive contexts be used so that potential female referents disappear (Treichler & Frank [1989] dub such uses FALSE GENERICS).

1. We awoke in the morning to find that the villagers had paddled away in their canoes, leaving us behind with the women and children.

2. The pioneers and their wives were often ill-equipped to deal with the bone-chilling winds and the heavy snows they experienced once winter began.

Linguistic expressions are never themselves the whole story. They are interpreted in discursive contexts. In (1) and (2), the immediate sentential contexts make it clear that semantic sex-inclusivity has been ignored in favor of male-exclusivity. Female-exclusive interpretations of semantically inclusive expressions do occur but far less frequently. An example I heard a few years ago was from a woman who had been among the very first commercial airline flight attendants. In talking to an NPR interviewer, she said

3. In those days, people stopped flying when they got married.

Historical knowledge plus the rest of her interview made it clear that she was only speaking of the few women working for the airlines, not of other people.

Here it is speaker meaning rather than conceptual baggage that imports gender significance that is lacking in the conventional meaning of the ‘false generic’ expressions used. What the speaker ‘meant’ when uttering “the villagers” is what could have been more explicitly expressed by uttering “the adult male villagers”: in Gricean terms, the speaker CONVERSATIONALLY IMPLICATES that only the adult men are to be included in “the villagers.” In contrast, the inference triggered by the word *surgeon* when someone utters “the surgeon” that the person so designated is male is not part of what the speaker meant — it is not implicated but simply comes along with the word. The word very often activates assumptions about the maleness of ‘prototypical’ surgeons, but the speaker could not have said “the male surgeon” and meant the same thing. The point of the story is that the surgeon in question is not male but the story shows that there is a strongly operative and to avoid generic masculines and also the potential clumsiness of repeated disjoined pronouns (*she or he, her or him, herself or himself*).
widespread default assumption that surgeons are male, one that operates at some level below overt attention even for those who know full well that some surgeons are female.

This is very different from the case where “the villagers” are differentiated from “the women and children.” There need be no default assumption that villagers are generally adult men; rather, the speaker manages to make clear that it is only adult men who are covered by this utterance of “the villagers.” It’s a matter of a contextual restriction of the scope of “villagers”: the speaker’s meaning intentions are what do the work. Of course, the hearer is helped to recognize those intentions by having access to background ideologies that attach far greater value and significance to adult men than to women and children, but the hearer may well strongly disavow those assumptions while nonetheless comprehending what the speaker means. In contrast, the speaker demonstrates complicitity in those assumptions by ‘meaning’ to restrict “the villagers” to the adult men in the village. And even hearers disposed to object to such a restriction may nonetheless find themselves tacitly endorsing it when they do not explicitly object, one way in which such ideologies are sustained.

Conceptual baggage and implicatures are two different ways in which gender and sexual ideologies expand/amplify/distort content meaning beyond what linguistic expressions encode. In the next section we will see that background assumptions about gender and sexuality can affect communicative significance in yet a third way via their impact on the hearer’s assessment of the speaker. And we will also see some of the mechanisms through which linguistic practices contribute to establishing and sustaining such ideologies.

4 DISCOURSE

Work on language, gender, and sexuality makes clear that the full significance of content meaning is not exhausted by the Gricean trio of expression, utterance, and speaker meaning. I have added conceptual baggage to the mix and have also considered the important role gender and sexual ideologies play in supporting speaker meaning, and I have also suggested that hearers’ access of those ideologies in order to comprehend what speakers mean by their contributions to ongoing discourse may play some role in reproducing and sustaining sometimes highly problematic ideologies.

But the discursive dimensions of content meaning go further. I want to note three important aspects of what might be called DISCURSIVE MEANING: EFFECTIVENESS of individual speech acts, MEANINGFUL PATTERNS of linguistic practice that transcend particular utterances, and MULTIVOCALITY. Effectiveness is tied to background gender and sexual ideologies in a way that differs both from conceptual baggage and from ideologically driven conversational implicature: effectiveness has to do with the hearer’s assessment of the speaker’s capacities, dispositions, and authority and involves both UPTAKE and UPDATING. Meaningful patterns across different utterances play a critical role in communicating and sustaining ideologi-
cal assumptions, even assumptions that many would find problematic if considered explicitly. Multivocality is about ways in which language users individually and collectively appropriate and reshape others’ contributions, often with ironic keying and ambivalence.

4.1 Effectiveness

Effectiveness is a matter of how a particular speech act is received, what changes it effects in the ongoing course of affairs, including the conversational record of what interactants are mutually assuming as well as attitudes and actions that might result. If I issue a directive to you, do you undertake to comply, take on the indicated obligation? If I make an assertion, do you add its content to your belief set or at least strengthen your confidence in the proposition I convey? If I make a joke, do you laugh? If I ask a question, do you take on the job of trying to answer it? If I make a suggestion, do you consider it seriously? And so on.

In trying to “do things with words,” some are more consistently successful than others. Success depends on many factors but the relation of the interlocutors, including their views of one another, is extremely important. The response or uptake to an utterance is critical; immediate and long-term effects on others must be considered. Grice [1957] initially defined what it was for speaker S to mean that p to hearer H in terms of S’s intending to produce in H the belief that p — i.e., to produce a certain kind of response. Similarly, Austin in his [1960] account of illocutionary acts called for a certain kind of ‘uptake’ on the hearer’s part: in order for an assertion to come off successfully, the hearer need not come to believe what the speaker asserts but must at a minimum recognize that the speaker is trying to bring it about that the hearer so believe.

Both Grice and Austin made success in meaning or in performing illocutionary acts depend on others than the speaker. Many critics noted that one could not ‘intend’ to do what one did not believe one could do and yet I might well mean that p to you even if I have no expectation that you could be led to believe p. So far as it goes, this criticism has some merit, but it does not really come to grips with the fact that semantic ‘success’ is not a purely individualistic matter: others play an important role. Although both the Gricean and the Austinian account do have problems, they are not just confused as some have assumed. A speaker whose attempts to mean are impeded by others — e.g., by others who simply refuse to believe that she means what she says when she rejects a man’s sexual advances — is semantically handicapped, illocutionarily constrained. I made this point in [McConnell-Ginet, 1989], building on a related observation made in an unpublished talk I’d heard philosopher Sara Ketchum make in the early 1980s.

Philosophers Jennifer Hornsby and Rae Langton make a similar point when they argue (in, e.g., [Hornsby & Langton, 1998]; and elsewhere) that the notion of ‘speech’ that is involved in advocacy of ‘freedom of speech’ is illocutionary. They suggest that speakers’ freedom of speech is curtailed if they cannot get their communicative intentions recognized, cannot fully perform certain illocutionary acts
because they cannot get hearers to recognize certain illocutionary intents. This point is developed in the context of arguing that pornography tends to produce a communicative climate in which women’s attempts to refuse men’s sexual advances are simply not recognized as such because it is assumed women’s putative refusals are just sham and that the women saying ‘no’ are really communicating ‘yes but you should keep on insisting’.

But the fundamental point is independent of whether pornography really does have the effects Hornsby and Langton posit. Communicative acts are not purely individualistic — at a minimum their success requires recognition by the hearer. And what seems clear is that recognition may be impeded by gender and sexual ideologies that the hearer draws on in assessing the communicative intentions of particular speakers. Without appropriate uptake from the hearer — recognition of the speaker’s illocutionary intent — the speech act is doomed.

Of course, effectiveness requires more. When I make a sincere assertion I want you to believe what I have said, a directive is issued with an interest in the hearer’s doing what it enjoins. Formal theories of discourse usually incorporate these desired outcomes in their models of the impact of utterances on developing conversation: e.g., an assertion adds the propositional content asserted to the conversational record, a directive places the requested hearer action on the hearer’s “to-do” list, etc. (See, e.g., [Roberts, 2004] for discussion.) In other words, the hearer’s updating the conversational record or context as intended is fundamental to the speaker’s illocutionary goals. Dynamic semantics does not assign propositional content to declarative sentences but rather their “context-change potential.” Assuming a possible worlds account, a declarative sentence has the potential to narrow down the set of worlds still considered ‘live options’ (thinking propositionally, the assertion can add to the stock of propositions taken for granted) in the discourse context. Whether this potential is actually realized when a sentence is uttered depends on social factors operative in the particular discourse, among which the hearer’s relation to the speaker, including his views of her interests and abilities, may often play a major role.

It is sometimes said that the desired outcomes involving, e.g., hearers’ beliefs or intentions to act are not part of what a speaker means or of the illocutionary force of her utterance but are rather further perlocutionary effects of the same kind as impressing or frightening or surprising. But the kinds of outcomes we are discussing are far more intimately tied to meaning as is evident from their role in dynamic semantics: speakers aim (or at least hope) to update the discursive context in certain standard ways by their utterances. Their capacity to effect such updating, however, is socially constrained.

So both uptake — recognition of what the speaker means — and appropriate updating — making the indicated changes in the discourse context — are essential for what I will call FULL SEMANTIC EFFECTIVENESS. It is not only gender and sexuality that can limit semantic effectiveness, of course, but it is in the context of thinking about issues like women’s difficulties in refusing some men’s sexual advances (and also in some other communicative attempts) that I have seen the
inadequacy of treating meaning without reference to hearers’ uptake and updating. As noted above utterance effectiveness is not limited to semantic effectiveness: the hearer might offer the required uptake (comprehension) and make the appropriate contextual updates yet further (perlocutionary) effects desirable to the speaker might not ensue. A speaker offers a suggestion in a business meeting and it is indeed adopted but the higher-ups adopting it attribute it to a male colleague: the speaker is not credited. Apportioning credit (or blame) for discursive moves requires keeping accurate tabs on who has said what and folks are not always good at that, especially when what was said (including the uptake and update) might not have been expected from that particular speaker. And of course further aims can be frustrated in many other ways as achieving them usually involves speaker’s and hearer’s assessments of one another’s immediate situations and longer-term interests and goals as well as of factors independent of discourse participants. But the first essential step is semantic effectiveness, both uptake and updating. Securing the intended uptake is crucial for illocutionary agency — one must be understood — but if the intended updating does not occur, the illocutionary act is seriously deficient.

4.2 Meaningful discursive patterns

Theoretical linguistics has made a big leap to go from static semantic accounts of word and sentence meaning supplemented by pragmatic explanations of speaker and utterance meaning to dynamic accounts of meaning as developing in discourse. But things are still pretty much focused on the effects of single utterances in a developing exchange of utterances, a discourse in the sense of a sequence of utterance. In assessing meaning, discourse analytic work among formal semanticists does not look at statistical patterning across utterances. Of course, pragmatic theorists in assessing what a speaker means by a particular utterance do consider what that speaker and others might say in analogous situations in order to assess the significance of the utterance at hand. But in those cases the interest is in ascribing meaning to a single speaker’s utterance.

In contrast, other approaches to discourse analysis, especially critical discourse analysis, do consider the significance of patterns at more global levels. What is meant by discourse in such work is often a far broader and more fully social notion than that of a sequence of utterances. Analysts often speak of discourses ‘of gender’, ‘of sexuality’, ‘of patriotism’, ‘of domesticity’, ‘of poverty’, and so on, where socially recognizable ways of talking and thinking about particular matters are highlighted. There are also empirical studies looking at statistical patterning across utterances, some of which provide support for theorists’ claims about social and cultural ‘discourses’.

As with phonetic variation, where it seems to be frequencies that matter, we also find frequencies significant in thinking about content. For example, media studies

\[\text{22See Valian, 1998 for examples and the discussion in McConnell-Ginet, 19xx (to appear in McConnell-Ginet, forthcoming).}\]
looking at modes of reference to women and to men in otherwise parallel contexts have found some striking patterns, among others that news accounts, including those of women in leadership positions, much more frequently mention women’s marital and parental status and their physical appearance (bodies and clothing) than they do men’s [Gill, 2007]. These same differences have also been (recently) found in letters of recommendation for medical faculty and, in addition, the men were far more often described by recommenders as brilliant, the women as hard-working — and these differences emerged from examining letters recommending successful applicants [Trix & Psenka, 2003]. Syntax text examples are not only populated more often by men but male referents are more likely found as subjects, female in object position [Macaulay & Brice, 1997].

Agency is more often attributed to men than women except in contexts where there is an issue of something like sexual assault: Ehrlich [2001] found the male defendant not only himself downplaying his agency (and responsibility) but the panel hearing the case against him failing to challenge his characterizations of himself as at the mercy of outside forces while at the same time panel members kept attributing agency (and responsibility) to the two females who were complaining of his continuing to make sexual advances in the absence of their consent. In these and many other cases, evidence of persisting gender ideology and assumptions are found in patterns of linguistic usage rather than in specific utterances or texts.

Assumptions about heterosexuality provide another example. Kitzinger [2002] reports that on arriving at the dining room of the hotel at which they were staying, she and her female partner were asked “will anyone else be joining you”? The question itself might seem routine but she stationed herself and observed numerous mixed-sex couples entering who were not asked that same question. That the same-sex pair was seen as needing augmentation is clear only by contrasting what was said to them with what was — or was not, in this case — said to mixed-sex pairs.

These are just a few examples where what matters, the significance, of something said does not reside in any particular utterance (or even the conjunction of utterances in a particular situated discourse) but in larger-scale patterning. Often these patterns not only reflect but help produce ideological positions, positions that may seldom be explicitly articulated as such in situated discourses but which emerge at more global levels to constitute discourses of gender and sexuality.

4.3 Multi-vocality

As mentioned above, discursive meaning is often modeled as an expanding set of propositions or, alternatively, a contracting set of possible worlds, those in which the propositions in the common ground are all true. Although no one denies that even individuals often have sets of beliefs that are inconsistent (usually not recognized to be so) or unstable (with rapid vacillation between one and another assessment of some position, especially when matters are complex), formal semanticists and philosophers of language typically make the simplifying assumption
that the common ground is a set of mutually compatible propositions that grows in a well-behaved way. Ignoring any contestation among different discourse participants or even ambivalence and confusion on the part of individuals is a useful and appropriate strategy for certain purposes. But such an idealization misses much of the complex character of the gender and sexual dimensions of discursive meaning, especially in an era where change and conflict are so prominent.

Ehrlich and King [1994] address some of the ways in which feminist-inspired lexical innovations have been transformed as they have been adopted and used more widely. So, for example, words like chairperson or chair, introduced to dislodge the problematic masculine generic chairman are often used to designate women only, while chairman is retained to designate men. And there are many other instances that show unwillingness to lose the gender distinction. Or consider the social title Ms, which was introduced as a substitute for the Mrs/ Miss pair that, parallel to Mr, would not mark marital status. Many, however, just added it to the other two titles for women: some, e.g., use Ms for divorced women, Miss for never married women, and Mrs for those currently married (or widowed), and there are a variety of other ways in which the original project of removing marital distinctions from women’s titles has gotten derailed. And terms like date rape have been rendered far less useful than they could be because anti-feminist critics have misrepresented their scope, claiming that feminists view as rape any heterosexual encounter that is unsatisfactory for the woman in any way. Words like feminist and feminism get loaded with CB of opponents’ stereotypes of feminists as man-hating crazies and feminism as a movement to rid the world of playful and pleasurable heterosexual relationships.23

Multivocality of two kinds is involved here. One is the cooperative/collaborative joining of voices that gets new ways of talking (and thinking) launched. It was a collective effort that got people to start thinking about how familiar masculine generics contributed to ways of thinking in which men were seen as the canonical exemplars of humanity, women being a ‘special interest’ group. The other kind of multivocality arises from contestation/disagreement from a responding/reacting set of voices, which can reshape how people interpret whatever is being responded to. Overt anti-feminist backlash discourses exemplify this as do the debates over ‘defining’ marriage (and marriage) discussed in [McConnell-Ginet, 2006].

Finally, I want briefly to note multivocality of a third kind, a more complex and conflicted mingling of alternative viewpoints. These are mixed messages, which can come in varied forms. Gill [2007, 261] discusses a variety of kinds of media aimed at girls and young women in which “achieving desirability in a heterosexual context is (re)presented as something to be understood as being done for yourself and not in order to please a man.” You are enjoined to choose freely but at the same time you are instructed in some detail as to which choices men will like (often including what consumer products will produce the “desired” effects). Gill speaks

23They also get loaded with a less distorted discourse history of racism in the women’s movement in mid-20th century America. See the discussion in [Eckert & McConnell-Ginet, 2003, 228-231].
of a “postfeminist sensibility” that invokes feminist-inspired visions of women’s autonomy and empowerment while still offering up familiar scenarios of happiness for a woman depending on her being chosen by “Mr Right.” Similarly, overt old-fashioned misogyny can be presented as laughable, humorous, offered with an ironic touch. But as is often the case with irony and with humor, ambivalence may be just below the surface, an attraction towards the viewpoints that are overtly being caricatured or mocked. The point is that many of these discourses are susceptible to multiple (and perhaps shifting) interpretations: there is no real resolution of the conflicting voices.

5 CONCLUSION

Language interacts with gender and sexuality in many different ways, only a few of which I have touched on here. As someone attracted to linguistics from a background in mathematics and the philosophy of language, I felt initially like a tourist in gender studies. After all, I really had no acquaintance with the kind of anthropological, sociological, and political analyses that informed the field. I dutifully acquainted myself with sociolinguistic and ethnomethodological work, but much of what I read did not seem very relevant to the kinds of questions that concerned me as a linguist. Even discourse analysis, which more directly addressed matters of meaning, seemed foreign, given the relative distance from the kinds of formal models of semantics and the insights into meaning offered by analytic philosophy that I found illuminating and which informed my own research.

But as I became more familiar with work of feminist scholars and queer theorists, I found myself beginning to think about language, especially content meaning, in some new ways. What are the relations between linguistic change and social conflict and change? How do linguistic resources affect our thinking and how does our thinking effect changes in linguistic resources? How might power and privilege affect what people are able to do with language? Of course there are no simple answers to such questions, but the growingly sophisticated body of work arising from thinking about connections of language to gender and sexuality, together with ongoing work in linguistic semantics and pragmatics as well as in philosophy of language, suggests some potentially productive strategies for addressing them.

Gender studies challenge linguists and linguistics to consider language from dynamic, interactional, and social perspectives. This does not mean that the individualistic cognitive focus characteristic of much theoretical linguistics has no place. What it does mean is that such an asocial and usually static approach cannot suffice.

ACKNOWLEDGEMENTS

Thanks to Ruth Kempson, who kept encouraging me to write this chapter, and to Susan Ehrlich, who gave me some very helpful comments on an earlier version.
BIBLIOGRAPHY


LINGUISTICS AND ANTHROPOLOGY

William O. Beeman

Anthropology and linguistics share a common intellectual origin in 19th Century scholarship. The impetus that prompted the earliest archaeologists to look for civilizational origins in Greece, early folklorists to look for the origins of culture in folktales and common memory, and the first armchair cultural anthropologist to look for the origins of human customs through comparison of groups of human beings also prompted the earliest linguistic inquiries.

There was considerable overlap in these processes. The “discovery” of Sanskrit by the British civil servant and intellectual, Sir William Jones in the late 18th Century set the stage for intensive work in comparative historical linguistics that continues to the present day. Jacob Grimm was not only a pioneering folklorist, but the pivotal figure in 19th Century linguistics through his discovery of regularities in consonantal shifts between different branches of Indo-European languages over historical time. His formulation, called today Grimm’s Law was not only the basis for modern linguistics, but also one of the formative concepts leading to 20th Century structuralism, particularly as elaborated in the work of Ferdinand de Saussure [1959], perhaps the most influential linguist in the 20th Century. The scholarly tradition that followed developments in historical linguistics in the Old World and Europe generally led to the development of formal linguistics as taught today in most university departments of linguistics.

Starting in the 20th Century, anthropological linguistics began to develop along somewhat different lines than formal linguistics. Anthropological linguistics today generally views language through a cultural and behavioral lens rather than through a formal, cognitive lens. Anthropological linguistics definitely concerns itself with the formal properties of phonetics, phonemics, morphemics and syntax, as well as the cognitive skills that are required for linguistic communication. However, its central questions lie in how language is used in the social and cultural life of people in different societies. It is also concerned with the broad question of how language evolved as part of the repertoire of human biological skills and behavioral adaptation.

LINGUISTIC ANTHROPOLOGY IN AMERICA–EARLY ROOTS

Many of the concerns of linguistic anthropology are shared by scholars throughout the world in varieties of disciplines ranging from philology and literature to psychology and cognitive science. However it is almost exclusively in North America that linguistics is included as part of the basic training of all anthropologists.
Because anthropology is a “four field” discipline in North America, encompassing cultural anthropology, biological anthropology and archaeology as well as linguistics, this also broadens the concerns of anthropological linguistics to interface with these other sub disciplines.

There is a special historical emphasis in North America as well on American Indian languages. This may stem in part from the early history of anthropology as a discipline, which focused heavily on indigenous North American peoples from the nineteenth century onward.

Intellectual interest in Native American languages predates anthropology itself, dating from the very earliest colonizing efforts in North America. Roger Williams, founder of Rhode Island, compiled a small dictionary of Narragansett [Williams, 1827; 1643, 1]. In the nineteenth century, this continuing U.S. Governmental responsibility for tribal peoples led to the writing of a large number of studies by the Bureau of American Ethnology on tribal groups throughout the Americas, including many grammars, dictionaries and compilations of folkloric material in original languages.

Linguistics was arguably introduced into the formal study of anthropology by Franz Boas, who exercised enormous influence on the discipline through his own work and that of his students. Boas was interested in linguistics for a number of reasons. First, as a result of his early work in the Arctic, he made attempts to learn Inuit, and found it an exceptionally subtle and complex language. Later this insight was incorporated into his anti-evolutionary theoretical perspective: historical particularism. He separated out the concepts of race, language and culture maintaining that they were independent of each other [Boas, 1940]. He maintained that any human was capable of learning any language, and assimilating any cultural tradition. He pointed out that different societies might have some aspects of their culture that were highly developed, and others that were simple relative to other world societies. Thus the idea that a society might be “primitive” in all ways — linguistically, culturally and biologically because they were evolutionarily backward was rejected. Each society was seen by Boas to develop independently according to its own particular adaptive pattern to its physical and social environment. Language too was seen as reflective of this general adaptive pattern. Boas’ views formed the basis for the doctrine of linguistic relativism, later elaborated upon by his students, whereby no human language can be seen as superior to any other in terms of its ability to meet human needs [Boas et al., 1966].

Boas’ second reason for considering linguistics important for the study of anthropology had to do with his feeling that linguistic study was able to provide deep insight into the workings of the human mind without the need for judgments on the part of informants. By eliciting data from native speakers, a linguist could build a model for the functioning of language of which the speaker him or herself was unaware. This avoided the “secondary rationalizations” that cultural anthropologists had to deal with in eliciting information from informants about politics, religion, economics, kinship and other social institutions. As ephemeral and programmatic as these ideas concerning language were, they would set the agenda for
anthropological linguistics for the balance of the century as they were elaborated by Boas’s students, particularly Edward Sapir, and Sapir’s student, Benjamin Lee Whorf [Mathiot, 1979].

1920-1950–SAPIR, WHORF AND MALINOWSKI

The most famous linguistic anthropologist to study with Boas was Edward Sapir. Although Sapir did not concern himself exclusively with linguistic research, it constituted the bulk of his work, and remains the body of his anthropological research for which he is the most well-known [Sapir and Mandelbaum, 1949].

Sapir’s interest in language was wide-ranging. He was fascinated by both psychological and cultural aspects of language functioning. The newly emerging concept of the “phoneme” was of special interest to him, and his seminal paper “The Psychological Reality of the Phoneme” [Sapir, 1933, 247-265], is an unsurpassed study showing that the phoneme is not just a theoretical fiction created by linguistic analysts, but represents a cognitive construct that is so strong that it leads individuals to assert the existence of sounds that are not present, and deny the existence of sounds that are present. In another paper, “A Study in Phonetic Symbolism” [Sapir, 1929, 225-239], he investigates the relationship between pure sounds and peoples’ semantic associations with them. Taking nonsense syllables, Sapir was able to show that people associate high vowels with small sensory phenomena and low vowels with large phenomena. Only recently have acoustic phoneticians returned to this problem in investigating the psycho-acoustic abilities of individuals to judge the length of the vocal tract of other speakers based solely on the sound of their voices.

Sapir also did pioneering work in language and gender, historical linguistics, psycholinguistics and in the study of a number of native American languages. However, he is best known for his contributions to what later became known as the Whorfian Hypothesis, also known as the Sapir-Whorf Hypothesis. Sapir maintained that language was “the symbolic guide to culture.” In several seminal articles, the most important of which may be ”The Grammarian and his Language” [Sapir, 1924, 149-155], he develops the theme that language serves as a filter through which the world is constructed for purposes of communication.

This work was carried forward by Sapir’s student Benjamin Lee Whorf, who devoted much of his research to the study of Hopi. Whorf took Sapir’s notion of language’s interpenetration with culture to a much stronger formulation. Whorf’s writings can be interpreted as concluding that language is deterministic of thought. Grammatical structures were seen not just as tools for describing the world, they were seen as templates for thought itself [Whorf, 1956; Whorf, 1956, 87-101]. To be sure, Whorf’s views on this matter became stronger throughout his life, and are the most extreme in his posthumous writings. The actual formulation of the Sapir-Whorf hypothesis was not undertaken by either Sapir or Whorf, but rather by one of Whorf’s students, Harry Hoijer [1954].

Aside from their views on language and thought, Sapir and Whorf were both
exemplary representatives of the dominant activity in American anthropological linguistics during the period from 1920-1960: descriptive studies of native American languages. This work focused largely on studies in phonology and morphology. Studies of syntactic structures and semantics were perfunctory during this period.

During this same period in England a parallel interest in linguistics in anthropology was developed from an unexpected source: the well-known social anthropologist Bronislaw Malinowski. Malinowski’s work in the Trobriand Islands was becoming well known. In his study, *Coral Gardens and their Magic* [Malinowski, 1935], Malinowski includes an extensive essay on language as an introduction to the second volume of the work. In this he addresses the problem of translation, taking as his principal problem the issue of the translation of magical formulas.

Magic formulas cannot really be translated, he maintains. They have no comprehensible semantic content. They do, however, accomplish cultural work within Trobriand society. They are therefore functionally situated. In order to “translate” such material, the ethnographer must provide a complete explanatory contextualization for the material. Otherwise it can make no sense. This functional theory of linguistics engendered a small, but active British school of linguistic anthropology, sometimes called the “London School” [Langendoen, 1968] whose principal exponent was the linguist J.R. Firth [Firth and Firth, 1986; 1964; Firth and Palmer, 1968], and later Edwin Ardener [Ardener, 1972, 125-132; Association of Social Anthropologists of the Commonwealth, and University of Sussex, 1971, 1-14].

**1950-1970—A PERIOD OF TRANSITION**

In the late 1950s and 1960s a number of linguistics and linguistically oriented cultural anthropologists collaborated on a linguistically based methodology called variously “ethnographic semantics,” “the new ethnography,” and most commonly “ethnoscientific” [Tyler, 1969]. Basing their work loosely on the Sapir-Whorf formulations, the most enthusiastic of these researchers maintained that if an ethnographic researcher could understand the logic of categorization used by people under ethnographic study, it would be possible to understand the cognitive processes underlying their cultural behavior. The more extreme cognitive claims for ethnoscientific were quickly called into question [Burling, 1964, 20-28] but the technique of ferreting out the logic of categorization proved useful for the understanding of specific domains of cultural activity. Ethno-botany, ethno-zoology, and the comparative study of color categorization (Berlin and Kay 1969) proved to be enduring lines of research.

An important collateral development growing out of structural linguistic study was the elaboration of markedness theory by Joseph Greenberg [1966]. Drawing from the phonological studies of the formal linguists of the Prague School of the 1930s, Greenberg showed that some categories of linguistic phenomena are more “marked” vis-à-vis other categories. The “unmarked” member of a pair is more general, and includes reference to a whole category of phenomenon as well as to a specific sub-category of that phenomenon. The “marked” member refers
exclusively to a specific sub-category. Thus “cow” is unmarked vis-à-vis “bull,” which is marked because the former refers both to the general category of the animal and to the female, whereas the latter refers only to the male member of the species. Greenberg shows that these distinctions pervade all formal grammatical systems, as well as other semantic domains, such as kinship.

POST-COMMISKIAN ANTHROPOLOGICAL LINGUISTICS

In 1957 Noam Chomsky published his revolutionary work, *Syntactic Structures* [Chomsky, 1965] and from this point onward linguistic anthropology began to diverge in its purpose and activity from linguistics as an academic discipline. Chomsky’s theoretical orientation took linguists away from the descriptive study of phonology and morphology and focused activity on syntax as the central formal structure of language. Although it has been modified considerably since 1957, Chomsky’s rule-based Transformational-Generative Grammar has been the basic paradigm within which formal linguists have worked. Basing much of their work on the exploration of intuitive understanding of language structures, and often working only with English, formal linguists largely abandoned the practice of linguistic fieldwork. Ultimately, under Chomsky’s direction, formal linguistics saw itself as a branch of cognitive science. The syntactic structures detected by linguists would, Chomsky believed, be shown to be direct emanations of the neural structures of the brain.

Anthropological linguists began during the same period to direct their work away from the study of formal linguistic structures, and toward the study of language use in social and cultural context. Work in phonology and morphology was largely directed toward the investigation of historical interconnections between language groups.

One important development was a growing interest in the investigation of linguistic communication as a “uniquely human” phenomenon.

LINGUISTIC COMMUNICATION AS BEHAVIOR

Communication in its broadest sense is behavior resulting in transfer of information among organisms, with the purpose of modifying the behavior of all participants involved in the process. Communication is basic to all life, and essential to living things whose lives are carried out in a social environment.

Anthropologists have long used complexity of communication abilities and practices as one measure of the differences between human beings and other life forms. Whereas many animals embody some form of information interchange in their primary behavioral repertoires, it has long been thought that only humans are capable of the complex form of communication known as language. The exclusiveness of this human ability has been called into question by experiments undertaken in recent years in communication with other animal species, notably chimpanzees.
and other great apes. However, it is reasonable to maintain that no other species has developed communication to the level of complexity seen in human life.

THEORETICAL MODELS OF COMMUNICATION

Although the study of linguistics in some form dates back almost to the time of writing, theoretical models of communication as a general process, with language seen as only a particular instance, are fairly recent. Both the semiotician and linguist Ferdinand de Saussure and the pragmatic philosopher, Charles Sanders Peirce provide the basis for much later work on the general structure of communication through their development of theories of the functions of signs.

Edward Sapir provided one of the earliest general formulations of a behavioral approach to communication in 1931 for the Encyclopedia of the Social Sciences [Sapir, 1931, 78-81]. In this article Sapir establishes that “every cultural pattern and every single act of social behavior involve communication in either an explicit or an implicit sense.” He also maintains that communication is fundamentally symbolic in nature, and is therefore dependent on the nature of the relationships and understandings that exist between individuals.

The German linguist Karl Bühler developed a field theory of language in his Sprachtheorie in 1934 [Bühler, 1990] which proved to be a sturdy model for mathematicians, linguists and social scientists. Bühler saw language as consisting of four elements: speaker, hearer, sign and object; and three functions: the expressive (coordinating sign and speaker), the appeal (coordinating sign and hearer), the referential (correlating sign and objects).

Claude Shannon and Warren Weaver of the Bell Telephone Laboratories collaborated in 1948 to develop a mathematical model of communication, which, though influential, eliminated any account of social and cultural factors from the communicative process [Shannon and Weaver, 1949]. Shannon and Weaver’s formulation contained six elements: a source, an encoder, a message, a channel, a decoder and a receiver. These general elements could be realized in many different ways, but a common formulation would be to recognize the speaker as the source, the mind and vocal system as the encoder, a code system such as language or gesture as the message; sound waves in air, or electronic signals as the channel; the auditory system and brain as the decoder; and the hearer as the receiver.

Shannon and Weaver also included in their model the concept of “noise” in the system. The mathematical description of noise later became known as entropy and was the subject of study in its own right. Information in this formulation is seen as the opposite of entropy. Both concepts are described in terms of probability. The less probable an event is within a system, the greater its information content. The more probable the event is, the smaller the information content, and the closer the event approaches entropy. The existence of a bounded system with evaluative parameters within which the probability of an event can be calculated is essential to this definition, otherwise an unexpected event will be seen as random in nature, and thus have little information content.
Roman Jakobson, drawing on Bühler developed a model for communication similar to that of Shannon and Weaver in 1960 [Hockett, 1966, 1-29] using the following diagram.

In the above diagram each of what Jakobson calls the “constitutive factors... in any act of verbal communication” is matched with a different “function” of language (here indicated in italics). According to Jakobson, in each instance of verbal communication one or more of these functions will predominate. His particular interest in this article was to explain the poetic function of language, which he identifies as that function of language which operates to heighten the message.

ANIMAL COMMUNICATION VS. HUMAN COMMUNICATION

Bühler’s work and Jakobson’s suggestive extension of it were also the basis for the study of animal communication. The semiotician Thomas Sebeok [Sebeok, Ramsay, and Wenner-Gren Foundation for Anthropological Research, 1969; Sebeok, 1977] used their model, but extended it by pointing out that visual and tactile channels are as important as auditory ones in the total spectrum of animal communicative behavior, thus the term “source” and “destination” are more inclusive than “speaker” and “hearer.”

Anthropologists have long identified linguistic communication as one of the principal elements — if not the principal element—distinguishing humans from other animal forms. In the 1950s and 1960s a number of researchers began to explore the continuity of animal vs. human communication systems. Most work in this early period was speculative and programmatic, but nevertheless influential in setting research agendas.

Charles D. Hockett, writing at approximately the same time as Sebeok identified 13 Design-Features of animal communication, some of which he saw as unique to human beings. Hockett called these “pattern variables,” to delineate the principal
characteristics of human language [Hockett, 1966, 1-29].

Hockett’s pattern variables are summarized in Figure 2 below. Of the thirteen features the last four: displacement, productivity, traditional transmission, and duality of patterning; were seen by later researchers as embodying exclusively human capacities. They were therefore seen as constituting a test for the presence of linguistic abilities in other animal species.

Both Hockett and Sebeok’s work have been used in evaluating the linguistic capacity of chimpanzees since the 1960’s. The first of the so-called “linguistic chimps” was named Washoe, and was trained in American Sign Language by psychologists Allen and Beatrice Gardner at the University of Nevada in Reno [Gardner et al., 1989]. Hockett’s list was widely adopted not only by anthropologists, but also by formal linguists and psychologists. In the 1970’s the list was used as a kind of inventory to measure the linguistic abilities of chimpanzees, who were being taught to communicate with humans using American Sign Language and other non-verbal techniques. Hockett’s pattern variables were seen as a way to evaluate how “human” the communications of Washoe and other “linguistic primates” were. Of particular interest were the capacity for “displacement” (being able to speak about things not present or imaginary, also to lie), “productivity” (being able to generate new and original expressions), and “duality of patterning” (the ability to select symbolic elements from an array and combine them and recombine them in regular patterns. Formal linguists in particular seized on duality of patterning as a test of syntactic capacity. The Gardeners objected, pointing out that their experiment was designed merely to test the proposition of interspecies communication, not to measure Washoe’s capacity for human language, but to no avail. Their research took on a life of its own as a number of researchers began to test chimpanzees under different conditions. One of the most successful of these research efforts was conducted with bonobos, closely related to chimpanzees, by Sue Savage-Rumbaugh and her husband Dwayne Rumbaugh [Savage-Rumbaugh et al., 1998; Savage-Rumbaugh and Lewin, 1994; Savage-Rumbaugh, 1986].

Hockett’s research also led him to speculate on the behavioral origins of human speech. This work was later carried forward by a small number of biological anthropologists, including Philip Lieberman [Lieberman, 2006], and was supplemented by work among the animal psychologists looking at chimpanzee communication.


The period from 1970-1990 saw anthropological linguistics concerned with the development of more sophisticated models for the interaction of language and social life. Sociolinguistics, which had begun in the 1950’s was one important area of new activity embraced by anthropological linguistics. This later developed into a new activity called “the ethnography of communication” by Dell Hymes and John Gumperz, two of the pioneers in the field [Gumperz and Hymes, 1986; Hymes, 1974].
**FEATURE** | **CHARACTERISTICS**
---|---
1. Vocal Auditory Channel | Information is encoded vocally and decoded aurally
2. Broadcast Transmission and Directional Reception | Information is transmitted through sound waves broadcast generally, but is received by hearing apparatus that is able to detect the direction of the source of sound
3. Rapid Fading (Transitoriness) | Information decays rapidly allowing for transmission of new information in sequential order
4. Interchangeability | Information that is encoded vocally is perceived as equivalent to information received aurally. Consequently, that which is heard can be mimicked or repeated by the hearer.
5. Total Feedback | The information produced vocally by the originator of communication is also heard by that same originator, thus providing a feedback loop, and self monitoring.
6. Specialization | Different sound patterns are used for different communicative purposes. In humans, speech sounds are used primarily if not exclusively for communication
7. Semanticity | Sign phenomena are able to be understood as representations for referenced objects.
8. Arbitrariness | There need be no intrinsic resemblance or connection between signs and the things for which they serve as reference
9. Discreteness | The continuum of sound is processed cognitively into discrete meaningful patterns
10. Displacement | Communication about object outside of the physical presence of the communicators or imaginary or speculative in nature is possible.
11. Productivity | New and original communications can be created by communicators freely without their having experienced them previously.
12. Traditional Transmission | Communication structures and information conveyed through communication are transmitted and acquired as a result of social behavior rather than genetic capacity.
13. Duality of Patterning | Meaningful communication units are differentiated from each other in patterns of contrast. They simultaneously combine with each other in patterns of combination.

Figure 2. Thirteen design-features of animal communication (after [Hockett, 1966])
Sociolinguistics came to be called by Hymes “socially realistic linguistics,” since it dealt with language as it was found in the structures of social life. Much of sociolinguistics consists of seeing variation in the language forms of a particular community and showing how that variation correlates with or is produced by social and cultural divisions and dynamics in the community. These divisions can be based on gender, ethnicity, class differences or any other culturally salient division within the community. Variation can be a property of the language of a given social division (e.g. male vs. female speech, or the different vocabularies exhibited by different generations). It can also be produced by social processes that govern relations within and between divisions. Such factors as group solidarity in the face of external challenges, desire for prestige, and inter-divisional conflict can manifest themselves in linguistic behavior that contributes to the variability seen within the community.

The ethnography of communication was first seen as a form of sociolinguistics, but it quickly took on a life of its own. Termed “socially constituted linguistics” by Hymes, the ethnography of communication deals with the ethnographic study of speech and language in its social and cultural setting. In a manner reminiscent of Malinowski, language is viewed not just as a form, but also as a dynamic behavior. This “functional” linguistics shows what language does in social life. To this end, each society can be shown to have its own unique cultural pattern of language use that can be accounted for by looking at its interrelationship with other cultural institutions.

Hymes developed Jakobson’s original list of constitutive elements and functions as shown in Figure 2 above in several publications [Hymes, 1974]. The most elaborate of these used the mnemonic SPEAKING as shown in Figure 3.

1985–PRESENT–DISCOURSE AND EXPRESSIVE COMMUNICATION

It was not long before linguistic anthropologists began to realize that to study language in its full cultural context, it was necessary to study highly complex linguistic behaviors. These became known widely under the general rubric of “discourse.” John Gumperz, one of the pioneers in this area of study, points out that the careful scientific study of discourse would be impossible if technology in the form of audio and video recorders had not been available when they were [Gumperz, 1982]. Indeed, the study of discourse processes involves painstaking recording, transcription and analysis of verbal interaction that would have been impossible in Sapir’s day.

Discourse structures are seen to be highly patterned, with beginnings, endings, transitions and episodic structures [Schegloff, 1968, 1075-95; 1982, 71-93; 2007; Goffman, 1981; Silverstein and Urban, 1996]. They are, moreover collaborative in their production. Therefore it is impossible to study speakers apart from hearers in a linguistic event; all persons present are contributing participants, even if they remain silent. Additionally, it can be seen that all participants are not equal in every discourse event. Some participants are conventionally licensed to do more
than others in their communicative roles. Discourse allows for the exercise of strategic behavior, so an adroit individual can seize an opportune moment in communication and advance an agenda. Here too, strategic silence may be as effective as strategic verbal behavior [Basso, 1970, 213-230].

Within societies different social groups may have different discourse styles. These differences can impede communication between groups even when the individuals involved feel that they “speak the same language.” Deborah Tannen [1996; 1991; 1989] has been successful in bringing popular awareness to the discourse differences seen between males and females in American society. Jane Hill has likewise investigated the differences in discourse structures in different bilingual Spanish/English communities in the American Southwest. Structures in many other languages expressing hierarchy, intimacy, politeness and deference have been explored by a variety of linguistic anthropologists drawing on the pioneering work of Gumperz and Hymes [Beeman, 1986; Errington, 1988; Moerman, 1988; Ochs et al., 1996; Duranti, 1994; Beeman, 1987; Inoue, 2006].

Expressive communication in the form of poetry, metaphor, and verbal art also constitute important elaborated communication genres in human life. Paul Friedrich has been a pioneer in the investigation of poetic structures in communicative behavior [Friedrich, 1986]. Deriving his work in part from a direction suggested by Roman Jakobson in his seminal paper in 1960 cited above, Friedrich concludes that the creation of poetic structures is a central feature of all linguistic behavior. The study of metaphor and symbols has been important in the study of ritual and religious life, but in this period anthropologists began to see the centrality of the creation of metaphor as a discourse process. George Lakoff and Mark Johnson’s *Metaphors We Live By* [Lakoff and Johnson, 1980] set the stage.

---

**Figure 3. Elements of communication (after [Hymes, 1974])**

- S Situation (Setting, Scene)
- P Participants (Speaker or sender; Addressor, Hearer or Receiver, Addressee)
- E Ends (Purposes–outcomes, Purposes–goals)
- A Act Sequence (Message form, Message content)
- K Key (Tone, manner or spirit in which communication is carried out)
- I Instrumentalities (Forms of speech, Channels of Speech)
- N Norms (Norms of Interaction, Norms of Interpretation)
- G Genres (Culturally recognized types of communication)
for other research in this area. James Fernandez' investigation of tropic structures throughout cultural life bridges the gap between linguistic anthropology and cultural anthropology [Fernandez, 1986; 1991]. Expressive culture is the principal conveyer of emotion in culture, and this too has been an important subject of research in anthropological linguistics [Beeman, 2001, 31-57; Wulff, 2007; Lutz and White, 1986, 405-436; Lutz and Abu-Lughod, 1990].

Verbal art in the form of oration, narration, theatrical performance and spectacle is perhaps the most directed and complex form of discourse for human beings. Richard Bauman has written extensively on the properties of verbal art and performative aspects of culture [Bauman, 2003]. One of the most interesting aspects of this area of human communication is its “emergent” quality. Of course all communication is to some extent emergent, in that its shape and direction is continually modified by ongoing events and participants. However performance is of special interest because it usually involves a fixed body of material that, despite its fixed character, is still modified by presentational conditions. In short, although it is possible to identify the roles of “performer” and “audience,” all participants are in fact co-creators of the piece being performed. Their collaborative effort gives the final form to the work, the nature of which cannot be understood until it is completed. Consequently, every performance is a unique event. This being the case, the analysis of a given performance is of less interest than the analysis of the social and communicative processes that engender it. A number of recent works have pursued the study of the use of poetry, poetic discourse and political rhetoric as performative aspects of language in social life [Duranti, 1994; Beeman, 1993, 369-393; Caton, 1990; Miller, 2007; Abu-Lughod, 1999].

HUMOR

One aspect of verbal art that has attracted a great deal of attention in anthropology is humor. Humor is a performative pragmatic accomplishment involving a wide range of communication skills including, but not exclusively involving, language, gesture, the presentation of visual imagery, and situation management. Humor aims at creating a concrete feeling of enjoyment for an audience, most commonly manifested in a physical display consisting of displays of pleasure including smiles and laughter. Because the content of humor and the circumstances under which it is created are cross-culturally variable, humor is subject to ethnographic investigation — a project in the ethnography of speaking.

The basis for most humor is the manipulation of communication to set up a surprise or series of surprises for an audience. The most common kind of surprise has since the eighteenth century been described under the general rubric of “incongruity.” Basic incongruity theory as an explanation of humor can be described in linguistic terms as follows: A communicative actor presents a message or other content material and contextualizes it within a cognitive “frame.” The actor constructs the frame through narration, visual representation, or enactment. He or she then suddenly pulls this frame aside, revealing one or more additional...
cognitive frames which audience members are shown as possible contextualizations or reframings of the original content material. The tension between the original framing and the sudden reframing results in an emotional release recognizable as the enjoyment response we see as smiles, amusement, and laughter. This tension is the driving force that underlies humor, and the release of that tension — as Freud pointed out [1960] — is a fundamental human behavioral reflex.

Humor, of all forms of communicative acts, is one of the most heavily dependent on equal cooperative participation of actor and audience. The audience, in order to enjoy humor must “get” the joke. This means they must be capable of analyzing the cognitive frames presented by the actor and following the process of the creation of the humor.

Typically, humor involves four stages, the setup, the paradox, the dénouement, and the release. The setup involves the presentation of the original content material and the first interpretive frame. The paradox involves the creation of the additional frame or frames. The dénouement is the point at which the initial and subsequent frames are shown to coexist, creating tension. The release is the enjoyment registered by the audience in the process of realization and the release resulting therefrom.

The communicative actor has a great deal to consider in creating humor. He or she must assess the audience carefully, particularly regarding their pre-existing knowledge. A large portion of the comic effect of humor involves the audience taking a set interpretive frame for granted and then being surprised when the actor shows their assumptions to be unwarranted at the point of dénouement. Thus the actor creating humor must be aware of, and use the audience’s taken-for-granted knowledge effectively. Some of the simplest examples of such effective use involve playing on assumptions about the conventional meanings of words or conversational routines. Comedian Henny Youngman’s classic one-liner: “Take my wife . . . please!” is an excellent example. In just four words and a pause, Youngman double-frames the word “take” showing two of its discourse usages: as an introduction to an example, and as a direct command/request. The double framing is completed by the word “please.” The pause is crucial. It allows the audience to set up an expectation that Youngman will be providing them with an example, which is then frustrated with his dénouement. The content that is re-framed is of course the phrase “my wife.”

The linguistic study of Jokes is widespread in humor studies. Because jokes are “co-created,” they are difficult as a literary genre. They almost beg to be performed [Sachs, 1974, 337-353; Norrick and Chiaro, 2009; Norrick, 1993; Oring, 2010; 2003].

In this way the work of comedians and the work of professional magicians is similar. Both use misdirection and double- framing in order to produce a dénouement and an effect of surprise. The response to magic tricks is frequently the same as to humor—delight, smiles and laughter with the added factor of puzzlement at how the trick was accomplished.
Humans structure the presentation of humor through numerous forms of culture-specific communicative events. All cultures have some form of the joke, a humorous narrative with the dénouement embodied in a punchline. Some of the best joke-tellers make their jokes seem to be instances of normal conversational narrative. Only after the punchline does the audience realize that the narrator has co-opted them into hearing a joke. In other instances, the joke is identified as such prior to its narration through a conversational introduction, and the audience expects and waits for the punchline. The joke is a kind of master form of humorous communication. Most other forms of humor can be seen as a variation of this form, even non-verbal humor.

Freud theorized that jokes have only two purposes: aggression and exposure. The first purpose (which includes satire and defense) is fulfilled through the hostile joke, and the second through the dirty joke. Humor theorists have debated Freud’s claims extensively. The mechanisms used to create humor can be considered separately from the purposes of humor, but, as will be seen below, the purposes are important to the success of humorous communication.

Just as speech acts must be felicitous in the Austinian sense [Austin 1962], in order to function, jokes must fulfill a number of performative criteria in order to achieve a humorous effect and bring the audience to a release. These performative criteria center on the successful execution of the stages of humor creation.

The setup must be adequate. Either the actor must either be skilled in presenting the content of the humor or be astute in judging what the audience will assume from their own cultural knowledge, or from the setting in which the humor is created.

The successful creation of the paradox requires that the alternative interpretive frame or frames be presented adequately and be plausible and comprehensible to the audience.

The dénouement must successfully present the juxtaposition of interpretive frames. If the actor does not present the frames in a manner that allows them to be seen together, the humor fails.

If the above three communicational acts are carried out successfully, tension release in laughter should proceed. The release may be genuine or feigned. Jokes are such well-known communicational structures in most societies that audience members will smile, laugh, or express appreciation as a communicational reflex even when they have not found the joke to be humorous. The realization that people laugh when presentations with humorous intent are not seen as humorous leads to further question of why humor fails even if its formal properties are well structured.

One reason that humor may fail when all of its formal performative properties are adequately executed is—homage à Freud—that the purpose of the humor may be overreach its bounds. It may be so overly aggressive toward someone present in the audience or to individuals or groups they revere; or so excessively ribald that it is seen by the audience as offensive. Humor and offensiveness are not mutually exclusive, however. An audience may be affected by the paradox as revealed in
the dénouement of the humor despite their ethical or moral objections and laugh in spite of themselves (perhaps with some feelings of shame). Likewise, what one audience finds offensive, another audience may find humorous.

Another reason humor may fail is that the paradox is not sufficiently surprising or unexpected to generate the tension necessary for release in laughter. Children’s humor frequently has this property for adults. Similarly, the paradox may be so obscure or difficult to perceive that the audience may be confused. They know that humor was intended in the communication because they understand the structure of humorous discourse, but they cannot understand what it is in the discourse that is humorous. This is a frequent difficulty in humor presented cross-culturally, or between groups with specialized occupations or information who do not share the same basic knowledge.

In the end, those who wish to create humor can never be quite certain in advance that their efforts will be successful. For this reason professional comedians must try out their jokes on numerous audiences, and practice their delivery and timing. Comedic actors, public speakers and amateur raconteurs must do the same. The delay of the smallest fraction in time, or the slightest premature telegraphing in delivering the dénouement of a humorous presentation can cause it to fail. Lack of clarity in the setup and in constructing the paradox can likewise kill humor. Many of the same considerations of structure and pacing apply to humor in print as to humor communicated face-to-face.

GESTURE AND NON-VERBAL COMMUNICATION

Because anthropology is concerned with the human soma, gesture and non-verbal communication have been especially important areas in the intersection between linguistics and anthropology.

Most human communication is vocal in nature. However anthropologists have long understood that much communication takes place using non-verbal behavioral mechanisms. These range from gesture and “body language” to the use of interpersonal space, the employment of signs and symbols and the use of time structures.

Non-verbal behavior has been seen to have many sequential and functional relationships to verbal behavior. It can “repeat, augment, illustrate, accent or contradict the words; it can anticipate, coincide with, substitute for or follow the verbal behavior; and it can be unrelated to the verbal behavior [Ekman et al., 1972] (see also [Ekman and Friesen, 1975]). In all of these situations humans have learned to interpret non-verbal signals in conventional ways. However, just as words must be taken in context to be properly understood, so must non-verbal behavior be interpreted in the whole context of any given communication.

Perhaps the most important form of non-verbal communication is facial expression. Human beings are capable of interpreting an exceptionally large number of variations in facial configuration. This form of non-verbal behavior may also be one of the oldest forms of communication in evolutionary terms. Based on research
on present-day groups of primates, such common facial movements as smiles or eyebrow raises may have been postures of hostility for prehistoric hominids. Facial expression is one of the most important sources of information about affect for human beings today.

Movement of hands or other body parts in clearly interpretable patterns are likewise important forms of non-verbal communication. These are generally classified as gestures. Birdwhistell called the study of body movement kinesics. Many gestures “stand alone” for members of a particular society. Gestures of insult, of invitation, of summoning or dismissal, and of commentary appear to be universal for human society.

Edward T. Hall pioneered the study of body distance (proxemics) and time usage (chronics) as forms of non-verbal communication. According to Hall [1966; 1959; Hall and Society for the Anthropology of Visual Communication, 1974], there are important cultural differences in body distance for different social purposes. In American society, for example, normal social conversation takes place at about eighteen inches distance between participants. In Egyptian society normal social distance may be as close as six inches. Americans who are unaware of this difference may find themselves uncomfortable in an Egyptian social conversation. Likewise Hall points out that different conceptions of time are communicative. These include the scheduling of daily routines such as meal time and meeting times; and ideas of punctuality. In some societies lack of punctuality conveys an insult, whereas in other societies rigid use of time creates discomfort.

Ekman and Friesen have developed a typology of non-verbal behavior following the work of Efron [1941]. Their categories are 1) Emblems — non verbal acts that have a direct dictionary translation well known by members of a particular culture. 2) Illustrators — body movement that accompanies speech and can either reinforce the words being said, or show a contradictory, ironic, or other attitudinal posture toward the verbal message. 3) Affect displays — primarily facial expressions conveying emotional states or attitudes. 4) Regulators — acts that maintain and regulate the back-and-forth nature of speaking and listening, usually taking place during the course of face-to-face interaction. 5) Adaptors — often unconsciously performed body movements that help persons to feel more comfortable in social interaction, to deal with tension or to accommodate themselves to the presence of others. Hall’s proxemic and chronemic dimensions of non-verbal behavior fall under this category.

Gesture is certainly one of the oldest communicative behavioral repertoires in the history of humanity. Students of primate behavior note that chimpanzees and other great apes have a fairly elaborate vocabulary of gesture. Lieberman [1991] and others speculate that the brain’s capacity for verbal language evolved as an elaboration of the centers controlling manual dexterity. This makes the universal use of hand gesture as an accompaniment to speech seem to be a survival from a pre-linguistic human state.

Human gestures differ from those of other animals in that they are polysemic — that is, they can be interpreted to have many different meanings depending on
the communicative context in which they are produced. This was pointed out by pioneering researcher Ray Birdwhistell [1970] who called the study of human body movement “kinesics.” Birdwhistell resisted the idea that “body language” could be deciphered in some absolute fashion. He pointed out that every body movement, like every word people utter, must be interpreted broadly, and in conjunction with every other element in communication. The richness of human communicative resources insures that gesture will also have a rich set of meaning possibilities. Contemporary students of human gesture, such as Adam Kendon [2004; 1990; Kendon et al., 1976], David McNeill, [2000; McNeill et al., 2007] and Starkey Duncan [Duncan and Fiske, 1985; 1977] note that gesture can often be used as an additional simultaneous channel of communication to indicate the mood or spirit in which verbal communication is to be understood. The actions of the body, hand and face all serve to clarify the intent of speakers. Often humans display several kinds of gesture simultaneously with verbal language.

Over the centuries deaf persons have elaborated gestures into a full-fledged linguistic system which is fully utilizable for all forms of face-to-face communication — including technical and artistic expression. There are many varieties of deaf “sign language,” but most share certain structural similarities. All combine different hand-shapes with distinctive movements in order to convey broad concepts. The semiotic system of these languages thus represents to some degree a pictographic communication system, such as written Chinese. Gestural languages have also been used as a kind of pidgin communication for trade between people who do not share a mutually intelligible verbal language.

ANTHROPOLOGY AND LINGUISTICS IN YEARS TO COME

It seems certain that the mission of linguistic anthropology will remain the exploration of human communicative capacity in all of its forms and varieties. While analysis of the formal properties of language will play a role in this work, it is not likely to have the central place in the work of linguistic anthropology that it does in linguistics. New technology will bring not only increasingly sophisticated investigative techniques for the study of language in human life, but also will provide for new forms of human communication. Some of these are already being studied by linguistic anthropologists.

Computer mediated communication in particular has taken many forms. Electronic mail (e-mail), direct “chat” via computer, and the use of electronic “bulletin boards” are only a few. Computer and satellite transmission of words and images over the planet has made it possible for people living at great distances to communicate regularly. Many thousands of such electronically constituted “speech communities” based on shared interests have already come into being. The rules for communication via these new channels are now being formulated by the communities that use them, and should provide fertile ground for research in the future [Axel, 2006, 354-384; Silverstein, 1998, 401-426; Wilson and Leighton C. Peterson, 2002, 449-467].
BIBLIOGRAPHY


Efron, 1941] D. Efron. Gesture and environment; a tentative study of some of the spatio-temporal and "linguistic" aspects of the gestural behavior of eastern jews and southern italians in new york city, living under similar as well as different environmental conditions. New York: King's crown press, 1941.


INDEX

?∃x.Fo(x), 387  
?∃x.Tn(x), 387  
Fo, 387  
Ty, 387  
⟨↓∗⟩, 389  
⟨L−1⟩, 389  
⟨↑∗⟩, 387  
*, see Kleene star

abduction, 118  
acceptability, 151  
accommodation, see presupposition, accommodation of accomplishments, 188  
achievements, 189  
aquisition, 361, 366  
concept, 310  
first-language, 5, 101, 445-472  
phonological, 420-434  
the logical problem of, 455  
action, 486  
active formula, 66  
adaptors, 546  
additive, 85  
Adjukiewicz, K., 81  
adjunction, 135–137  
adverbial  
temporal, 234  
affect displays, 546  
agency, 497  
ascription of, 524  
agentive passivization, 75  
Aktionsart, 187  
algebra, 4  
alliances, 489  
allophone, 410  
α-conversion, 71  
alphabetism, 412, 415  
altruism, 488  
ambiguity, 63, 165–166, 363, 364, 371, 376, 381–383, 390, 393  
attachment, 165  
collective-distributive, 165  
semantic, 147  
syntactic, 165, 166  
American-Indian languages, 532  
American sign language, 538  
analogical generalisation, 422  
Andersson, K., 490  
aminal behaviour, 477, 479  
annotated, 468  
antecedent, 66, 68, 72  
antecedent-contained ellipsis: see ellipsis, antecedent-contained  
anthropological linguistics, 531, 532, 538  
anthropology, 4, 532  
antiquity, 482  
apes, 483, 488, 491, 498  
appropriateness, 240  
arcaeology, 532  
Arctic, see Inuit  
Ardener, E., 534  
argument, 485  
event, 254  
instrumental, 254  
optional, 254–255  
referential, 247  
argument from the poverty of the stimulus (APS), 446  
argument structure, 168, 169  
Aristotle, 6, 22, 64  
Armstrong, D. M., 23
Arnauld, A., 6
Arnold, K., 498
artificial intelligence (AI), 149, 152–153
imperfective, 264
perfective, 233, 263
progressive, 233
assertion, 215
assignment function, 65
attribute value matrices, 273
audience, 542
auditory feedback, 492
Austin, J. L., 13, 272, 487, 521, 544
automata, 461
automata theory, 161
probabilistic, 168
auxiliary verbs, 497

Bühler, K., 536, 537
babbling, 492
babies, 483, 488, 489
Bach language, 466
backward-chaining, 68
bag-of-words model, 158
Baker, G. P., 16
Balaban, M., 486
Bar-Hillel, Y., 81
base component, 74
Bates, E., 495
Bauman, R., 542
Bayes assumption
naïve, 159
Bayes’ rule, 145, 155–164
Bayesian, 457
Bealer, G., 10
Begriffsschrift, 64
behaviorism, 15, 398, 434
Bell telephone laboratories, 536
Benveniste, E., 6
β-conversion, 71
Bezuidenhout, A., 9
bilingual, 541

binary branching, 470
binding
semantic, 230
binding presuppositions, 248
binding problem, 340, 344, 345n, 350–352
biolinguistic, 34, 35, 96, 100
biological anthropology, 532
biological evolution, 479, 480
biologists, 494
birdsong, 498
Birdwhistell, R. L., 546, 547
Black, M., 519
Blackburn, P., 386
Bloom, P., 478
Bloomfield, L., 14
Boas, F., 532
body language, 545–547
bonobo, 538
Kanzi, 489
Booth, A. E., 486
Borker, R., 506
bound variable pronoun see pronoun
boundedness, 190
Brăuer, J., 488
Bréal, M., 20
bracket operator/modality, 85, 88
Brandom, R., 14, 17, 24
breathing, 491
Bresnan, J., 76
Browman, C. P., 491
Bureau of American Ethnology, 532
Buszowski, W., 73
Bybee, J. L., 477

c-command, 43
c-structure, see grammar, lexical-functional
Call, J., 488
Cameron, D., 506n
Cann, R., 385, 389
captive animals, 486
Carnap, R., 6, 8, 12, 13, 21, 23
Carpenter, B., 81
Carston, R., 13
Cartesian mind-body dualism, 122
Cartesian product, 68
case statement, 70
Cassirer, E., 26
categorial grammar see grammar, categorial
categorial hierarchy, 127
categorical judgement, 485
categorical perception, 489
categories, 405–407
categorisation, 534
CCG, see grammar, categorial, combinatorial
centre-embedded clauses, 496
Champollion, L., 190
changes, 493
chat, 547
Cheney, D. L., 484
child language development, 495
CHILDES, 471
children
cognitive development of, 492
humour of, 545
chimpanzees, 484, 488–489, 535, 546
chinchillas, 489
Chinese, 547
Chomsky hierarchy, 33, 40, 46, 52–55, 74, 448
Chomsky, N., 2, 4, 5, 19–21, 24–26, 63, 74, 76, 149, 325–331, 357–359, 477, 478, 481, 486, 494, 535
Church, A., 68
Church–Turing Thesis, 68
class of natural languages, 447
classical logic, see logic, classical
first-order, 68
propositional, 67
Clayton, N. S., 485
clustering, 462
coevolution, 480, 482, 492, 495, 496
cochlea, 489
code conception of language, 6, 17, 25
code cracking, 461
aspeccial, 233–234, 259–261
cognitive frame, 542, 543
cognitive science, 89, 152, 325, 460
coherecne, see discourse structure
coherecne theory of truth, 22, 23
colour categorisation, 534
combinator, 89
combinatorial property, 458
combinatoriality, 486
combinatory categorial grammar, see grammar, categorial, combinatorial
comedyans, 543
common ground, 524, see context, discourse
communication, 95, 96, 399, 484, 486, 487, 489, 496
computer mediated, 547
expressive, 541, see conceptual baggage
mathematical theory of, 536
communication system, 96
communicative, 115
communicative use, 114, 115
community of practice (CofP), 515
commutativity, 66
competence, xiii, 325–331, 336–339, 359, 360, 375, 378, 397, 403, 427, 434, 495, 496
competitive, 486
completeness, 66, 73
complex NP constraint, 372
complexity, syntactic 479
complexity of learning, 461
composition, 83
compositionality, 19, 21, 75, 107, 321
369, 378, 394–396, 494
comprehension, 90, 325, see parsing
computability, 101
computation, 68, 89
computational complexity, 454, 460
computational efficiency, see efficiency, computational
computational hardness, see hardness, computational
computational learning, see learning, computational
computational system, 96, 98, 127
computational theory of mind, 199-200, 375, 408, 409, see Fodor, J.A.
computationally efficient, see efficiency, computational
computer mediated communication, 547
computer science, 25
conception of time, 546, see semantics, temporal
concepts, 310, 483
conceptual baggage (CB), 516–517, 525
compared to conversational implicature, 519
conceptual-intentional system, 134, 135
conceptual/intentional interface, 359
compared to conversational implicature, 520
conclusion, 66, 68
conditional, see operator, conditional
congruence classes, 465
connotations, see conceptual baggage, implicature
conscious awareness, 429
constant, 69
constituency, 164
constituent structure, 470
constituent structure grammar, 47
constituent-structure, see grammar, lexical-functional (LFG)
construction, 101, 102
Construction Grammar, 494, 495
content meaning, 520
context, xiii, xv, 9, 10, 14, 545
ellipsis and, 390, 392
discourse, 232, 239, 264
partitioning of, 215, 216
predicational, 264
context sensitive syntactic structure, 466
context word, 158–159
context-free, 39, 40, 44
context-free grammar, see grammar, context-free
context-free languages, 39, 40, 44; see grammar, context-sensitive
context-sensitive, xvi, 44
context-sensitive grammar, see grammar, context-sensitive
context-sensitive languages, 452
continuity, 494
contraction, 66, 68
tcontrol, 366
convention, xviii, 414, 421, 433, 438, 439
conventional implicature, see implicature, conventional
convergence, 454
conversational implicature, see implicature, conversational
conversational maxims, 13; see maxims, Gricean
conversational record, 521
Cook, R., 485
Cook, V. J., 5n
cooperation, 488, 489
coordination, 46, 48, see type, coordination
semantic, 316, 317, 319, 321
non-standard constituent, 77, 78, 80, 81
standard constituent, 78
copy-theory of movement, see movement, copy-theory of
coreferential pronoun, see pronoun, coreferential
corpora, 148, 159, 468
  bilingual, 163
corpus linguistics, 145, 148
correspondence theory of truth, 22, 23
cortex, 491
counting, 186
courtship, 487
Coward, R., 519
Creole, 478, 480
Cresswell, M., 21, 25
Croft, W., 494
Culler, J., 4n
culmination, 189
cultural evolution, 479, 480
cultural anthropology, 532
Curry-Howard correspondence, 70
Curry-Howard proofs-as-programs, 77
Curry-Howard type-logical semantics, 87; see grammar, categorial, type-logical
cut formula, 66
cut rule, 66
cut-elimination, 68, 73, 89
CV structure, 492, 493
cybernetics, 152
cycle, 109
cyclicity, 107
dénouement, 543, 544
Dalrymple, M., 77, 380
Darwin, C., 498
Davidson, D., 15, 21, 23, 98, 179
Davidsonian event, 254–258
de Boer, B., 492
de Saussure, F., 2–5, 12, 19, 20
de Waal, F. B. M., 489
decidability, 68
decoding, 157
deduction, natural, see natural deduction
deep structure, 74
definite descriptions, 8, 9
deflationary theory of truth, 23
DeGusta, D., 491
demonstratives, 497
denotation, 176 483
dependency, xvi
discontiguous, 168
long-distance, see dependency, unbounded
unbounded, 78, 79, 83, 387, 388
word-word, 168
dependency grammar, 168
derivation, 56, 57
derivational economy, 105
Derrida, J., 18, 19
Descartes, R., 116, 123
design perfection, 111–113
Dessalles, J.-L., 489
deterministic finite state automata, 463
Devitt, M., 8
Dewey, J., 12
DFAs, 463
dialogue, 153, 365, 366, 380, 382, 391, 392
dialogue systems, 153
Dickinson, A., 485
diférence, 18
dimension, 190
disambiguation, 151
word-sense, 146–147, 157–159
discontinuity, 371, 387, 388
discontinuous constituent, 54; see dependency, discontinuous
discourse, xiv, 229, 497, 510, 540, 541; see DRT, SDRT
discourse analysis, 153, 523
discourse context, see context, discourse
discourse marker, 230
discourse relations, 230, 264–268
discourse representation structure (DRS), 376–380
discourse representation theory (DRT), 230, 277, 308, 376–381, 384, 385, 394, 396
segmented (SDRT), 230–234, 264–268
discourse structure, 264–268, 540, 541
  partial ordering of, 267–268
discrete infinity, 127
disjoint union, 68
displacement, 538
distinctiveness, 493
distinguished subconfiguration, 73
distributed cognition, 438
distribution, 457
distributional lattice grammars (DLG), 454, 466
distributional learning, 445, 464
distributivity, 190
Dobzhansky, T., 480
dog, 484, 490
domain, 65
domain general learning procedures, 446
Doré, F., 484
double-framing, 543
downward entailing contexts
  identifying, 154
Dowty, D., 83
Dretske, F., 24
DRT, see discourse representation theory (DRT)
DS, see dynamic syntax (DS)
dualism, 121
duality of patterning, 538
duality of semantics, 130
Dummett, M., 6, 23
Duncan, S., 547
dynamic predicate logic, see logic, dynamic predicate
dynamic semantics, see semantics, dynamic
dynamic syntax (DS), 385–397
  linked trees in, 389
  modalities in, 386
  tree logic of (LOFT), 386
dynamics, 359–361, 374, 384, 385, 393, 397, 398
E-language, 5, 26, 41, 52, 407, 421
E-physical, 408–410
E-type pronoun, see pronoun, E-type
ease of articulation, 493
Eckert, P., 512, 513, 515
Eco, U., 11
economy of effort, 480
economy principle, 102
Eddington, W., 12
effectiveness
  full semantic, 522
  utterance, 523
  efficiency,
    computational, 105, 129
  learning, 461
Efron, D., 546
Egli, U., 9
Ehrlich, S., 524, 525
Ekman, P., 546
electronic mail, 547
  antecedent-contained, 381
  sloppy readings of, 364, 381, 382, 391
  strict readings of, 391
  VP-ellipsis, 391
embedding, 496
emblems, 546
emic/etic, 406
emotion, 542
English progressive, 179
entailment, 65, 233–234, 177
entropy, 536
episodic memory, see memory, episodic
EPP-requirement, 107
epsilon calculus, 386n
equivalence 3; see type equivalence
equivalent derivations, 57–59
essence, 10
η-conversion, 71
ethno-botany, 534
ethnoscience, 534
ethno-zoology, 534
ethnographic semantics, 534
ethnography of communication, 538, 540
ethnography of speaking, 542
evaluation methods, 468n
event strings, 319
eventology, 187
events, 179, 180, 187, see semantics, event
eventuality, 262–263
evolution, 478, 481
evolution of language, 477, 479, 488, 493
evolutionary remnant, 492
exapted, 490
expectation maximization (EM), 470
experimental, 472
explanation, 93, 101
explanatory adequacy, 112
expletive, 371
explicature, 13
exponential, 85
exponential complexity, 460
exponential function, 460
expressive communication, 541, see communication, expressive
expressive culture, 542
extension, 8, 21, 22
externalism, 440
eyebrow raises, 546

$F_1$-score, 468
f-structure, 76, 77
facial expression, 545, 546
feature percolation, 78n
feature structure, 272, 277
features, 177
felicitous, 544
Fernandez, J., 542
Fetzer, J., 10
Fibonacci, 116
Field, H., 24
Fillmore, C., 273, 494

finite class of languages, 450
finite state languages, 40
finite state Markov processes, 37, 52
first-order judgement, 486
first-order logic, see logic, predicate
first-order term, 65
Firth, J. R., 534
Fitch, W. D., 491
Fitch, W. T., 478
FLB, 481
Fodor, J. A., 20, 24, 385
food-sharing, 489
form and function, 96
formal grammar, see grammar, formal
formal language theory, 167
formal language, 97, 359, 360, 363, 367–370, 374, 448
formal semantics, see semantics, model-theoretic
formal syntax, see syntax, formal
formal system, 74
formalisation, 63
formulas-as-types, 70
Foucault, M., 18, 508
“four field” discipline, 532
fragment, see ellipsis
FrameNet, 273, 274, 289, 311, 312, 314
frames, 273, 289, 304, 305, 309, 310, 321
Frank, F., 519
Frege, G., 7, 8, 21, 64, 98, 117, 212
frequency, 457
Freud, S., 543, 544
Friedrich, P., 541
Friesen, W. V., 546
full interpretation, 103, 105
fully combinatorial language, 494
function, 478
function letter, 65
function of language, 486
functional abstraction, 70
functional annotation, 76
<table>
<thead>
<tr>
<th>Term</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>generative capacity of</td>
<td>170</td>
</tr>
<tr>
<td>lexicalized</td>
<td>166</td>
</tr>
<tr>
<td>tree-adjoining</td>
<td>167–168</td>
</tr>
<tr>
<td>grammar induction</td>
<td>454, 468</td>
</tr>
<tr>
<td>unsupervised</td>
<td>469</td>
</tr>
<tr>
<td>grammatical function</td>
<td>76, 88</td>
</tr>
<tr>
<td>grammatical inference</td>
<td>459</td>
</tr>
<tr>
<td>grammatical markers</td>
<td>497</td>
</tr>
<tr>
<td>grammatical number</td>
<td>191</td>
</tr>
<tr>
<td>grammaticality</td>
<td>145, 149–150</td>
</tr>
<tr>
<td>grammaticality judgements</td>
<td>xiv, 34, 35n, 359</td>
</tr>
<tr>
<td>empirical basis for</td>
<td>149</td>
</tr>
<tr>
<td>grammaticalization</td>
<td>496, 497</td>
</tr>
<tr>
<td>granularity</td>
<td>190</td>
</tr>
<tr>
<td>Gray, J.</td>
<td>506</td>
</tr>
<tr>
<td>great apes</td>
<td>536, 546</td>
</tr>
<tr>
<td>Grice, H. P.</td>
<td>13, 153, 521</td>
</tr>
<tr>
<td>Grimm’s Law</td>
<td>531</td>
</tr>
<tr>
<td>Grimm, J.</td>
<td>531</td>
</tr>
<tr>
<td>grounded</td>
<td>417</td>
</tr>
<tr>
<td>groundedness of phonology</td>
<td>403</td>
</tr>
<tr>
<td>group readings</td>
<td>187</td>
</tr>
<tr>
<td>growth of trees</td>
<td>361, 367, 368, 384, 385, 386n, 390, 393, 394</td>
</tr>
<tr>
<td>Gumperz, J.</td>
<td>506n, 538, 540</td>
</tr>
<tr>
<td>Gupta, A.</td>
<td>185</td>
</tr>
<tr>
<td>Hacker, P. M. S.</td>
<td>16</td>
</tr>
<tr>
<td>Hacking, I.</td>
<td>1n, 118</td>
</tr>
<tr>
<td>Hale, K.</td>
<td>187</td>
</tr>
<tr>
<td>Hall, E. T.</td>
<td>546</td>
</tr>
<tr>
<td>Hamilton, W. D.</td>
<td>488</td>
</tr>
<tr>
<td>happenings</td>
<td>183</td>
</tr>
<tr>
<td>hardness, computational</td>
<td>462</td>
</tr>
<tr>
<td>Hare, B.</td>
<td>488</td>
</tr>
<tr>
<td>Harris, R.</td>
<td>4n</td>
</tr>
<tr>
<td>Hashiya, K.</td>
<td>490</td>
</tr>
<tr>
<td>Haste, H.</td>
<td>508</td>
</tr>
<tr>
<td><em>Haupatz</em></td>
<td>68</td>
</tr>
<tr>
<td>Hauser, M. D.</td>
<td>478, 490, 494</td>
</tr>
<tr>
<td><em>he-man</em> language, see generic masculine</td>
<td></td>
</tr>
<tr>
<td>head-dependency grammars</td>
<td>470</td>
</tr>
<tr>
<td>head-driven phrase structure grammar</td>
<td></td>
</tr>
<tr>
<td>(HPSG), see grammar, head-driven phrase</td>
<td></td>
</tr>
<tr>
<td>structure (HPSG)</td>
<td></td>
</tr>
<tr>
<td>Heffner H. E.</td>
<td>490</td>
</tr>
<tr>
<td>Heffner, R. S.</td>
<td>490</td>
</tr>
<tr>
<td>Heidegger, M.</td>
<td>17, 18</td>
</tr>
<tr>
<td>Heine, B.</td>
<td>497</td>
</tr>
<tr>
<td>here and now</td>
<td>484</td>
</tr>
<tr>
<td>heterosexuality</td>
<td>524</td>
</tr>
<tr>
<td>assumptions of</td>
<td>524</td>
</tr>
<tr>
<td>heuristic</td>
<td>472</td>
</tr>
<tr>
<td>Hewitt, G.</td>
<td>491</td>
</tr>
<tr>
<td>hidden Markov model, see model, hidden</td>
<td></td>
</tr>
<tr>
<td>Markov (HMM)</td>
<td></td>
</tr>
<tr>
<td>hidden structure</td>
<td>463</td>
</tr>
<tr>
<td>higher-order logic, see logic, higher-</td>
<td></td>
</tr>
<tr>
<td>order</td>
<td></td>
</tr>
<tr>
<td>Hill, J.</td>
<td>541</td>
</tr>
<tr>
<td>Hinzen, W.</td>
<td>5</td>
</tr>
<tr>
<td>Holdcroft, D.</td>
<td>4n</td>
</tr>
<tr>
<td>holism</td>
<td>200</td>
</tr>
<tr>
<td>homomorphism</td>
<td>75</td>
</tr>
<tr>
<td>honeybees</td>
<td>487</td>
</tr>
<tr>
<td>honorifics</td>
<td>509</td>
</tr>
<tr>
<td>Hoopes, J.</td>
<td>11</td>
</tr>
<tr>
<td>Hopcroft, J. E.</td>
<td>5n</td>
</tr>
<tr>
<td>Hopekins, W. D.</td>
<td>488</td>
</tr>
<tr>
<td>Hopi, 175, 533</td>
<td></td>
</tr>
<tr>
<td>Hopper, P. J.</td>
<td>497</td>
</tr>
<tr>
<td>Hornsby, J.</td>
<td>521</td>
</tr>
<tr>
<td>Horwich, P.</td>
<td>23</td>
</tr>
<tr>
<td>HPSG, see grammar, head-driven phrase</td>
<td></td>
</tr>
<tr>
<td>structure</td>
<td></td>
</tr>
<tr>
<td>Hultsch, H.</td>
<td>498</td>
</tr>
<tr>
<td>human infants</td>
<td>492</td>
</tr>
<tr>
<td>human nature</td>
<td>120</td>
</tr>
<tr>
<td>human syntactic capacity</td>
<td>494</td>
</tr>
</tbody>
</table>
Hume, D., 123
humour, 542
Humphreys, P. W., 10
Hurford, J., 485, 496
Hymes, D., 538, 540
hypothesis
  closed, 68
  open, 68
hypothesis space, 448, 449
hypothetico-deductive method, 424
I-language, xiv, 5, 26, 37, 41
icon, 11
idempotency, 66
identification in the limit (IIL), 449
identities, 3,
  indexing of, 504–509
identity axiom, 66, 68
identity group, 66
ideologies, 509, 511, 520
idioms, 495
IID, see independently and identically distributed
illocution, 487
illocutionary act, 13, 212, 521–522
illustrators, 546
implicature, 177, 197, 222, 234
  conventional, 177
  conversational, 13, 177, 519–520
implicit learning, 432
inclusiveness, 115
incongruity, 542
incrementality, 360, 378, 379, 384, 385, 389, 395, 396
indefinites, 236–237
independently and identically distributed (IID), 457
index, 11
indexical, 176, 361, 376
indexical field
  definition of, 513
indexical meanings, see meaning, indexical
indirect indexing, 512
indirect negative evidence, 455
individual constant, 65
individuals, 178, 185
  manifestation of, 185
inductive generalisation, 422
inference, 358n, 368, 369, 399
inferential role, 17
inferentialism, 358n
information retrieval, 468n
infons, 276
information, 536
information extraction, 153, 468n
information packaging, 264–265
information state update, 317
information theory, 145
initial state, 99
injection, 70
innate dispositions, 481
innate principles, 479, 494
innateness, 446
instantaneous telic events, 189
instantiation, 404
instrumental, 413, 438
instrumentalist, 405, 406, 413
intension, 8, 9, see logic, intensional
intensional interpretations (senses), 182
intensional logic, see logic, intensional
intensional semantics, see semantics, intensional
intensionality, 182
intention, 426
intentionality, 24
interaction, 510
interface condition, 137
interface representation, 103
interfaces, 133
internal Merge, see Merge, internal
internalisation, 430
internalism, 438, 440
interpersonal space, 545
interpretation function, 65
interpretation of first-order logic, 65
interrogatives, 454
intersubjective, 409, 421, 427, 432
<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>intractable</td>
<td>461</td>
<td></td>
</tr>
<tr>
<td>intuition</td>
<td>325, 327–337</td>
<td></td>
</tr>
<tr>
<td>intuitionistic logic, see</td>
<td>logic,</td>
<td>intuitionistic logic, see</td>
</tr>
<tr>
<td>Inuuit</td>
<td>532</td>
<td></td>
</tr>
<tr>
<td>invalid</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>‘is a’ relations, 58, 59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>island constraint, 85, 372, 390, 392, 393</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jackendoff</td>
<td>336-338, 489</td>
<td></td>
</tr>
<tr>
<td>Jakobson, R.</td>
<td>4, 537</td>
<td></td>
</tr>
<tr>
<td>Janssen, T. M. V.</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Jespersen, O.</td>
<td>498</td>
<td></td>
</tr>
<tr>
<td>Johnson, M.</td>
<td>81, 541</td>
<td></td>
</tr>
<tr>
<td>joint attention, 488</td>
<td></td>
<td></td>
</tr>
<tr>
<td>joke</td>
<td>544</td>
<td></td>
</tr>
<tr>
<td>Jones, Sir W.</td>
<td>531</td>
<td></td>
</tr>
<tr>
<td>Kamp, H.</td>
<td>9, 181, 230, 308, 374–379</td>
<td></td>
</tr>
<tr>
<td>Kanzi, 486</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaplan, D.</td>
<td>xv</td>
<td></td>
</tr>
<tr>
<td>Kaplan, R. M.</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>Kasper, R. T.</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td>Katz, J. J.</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Katz, P. A.</td>
<td>486</td>
<td></td>
</tr>
<tr>
<td>Kay, P.</td>
<td>494</td>
<td></td>
</tr>
<tr>
<td>Kay, R. F.</td>
<td>491</td>
<td></td>
</tr>
<tr>
<td>Kendon, A.</td>
<td>547</td>
<td></td>
</tr>
<tr>
<td>Ketchum, S.</td>
<td>521</td>
<td></td>
</tr>
<tr>
<td>kin selection, 488</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kinds, 178, 185</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kinesics</td>
<td>546</td>
<td></td>
</tr>
<tr>
<td>King, R.</td>
<td>525</td>
<td></td>
</tr>
<tr>
<td>Kiritani, S.</td>
<td>489</td>
<td></td>
</tr>
<tr>
<td>Kitzinger, C.</td>
<td>517, 524</td>
<td></td>
</tr>
<tr>
<td>Kleene star</td>
<td>77, 387</td>
<td></td>
</tr>
<tr>
<td>‘knowing how’, 396–398</td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘knowing that’, 396</td>
<td></td>
<td></td>
</tr>
<tr>
<td>knowledge, 97, 99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>knowledge of language, 359–361, 366, 375, 396–398</td>
<td></td>
<td></td>
</tr>
<tr>
<td>knowledge representation theory, 152</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kojima, S.</td>
<td>489, 490</td>
<td></td>
</tr>
<tr>
<td>Kripke, S.</td>
<td>9, 10, 16</td>
<td></td>
</tr>
<tr>
<td>Kulick, D.</td>
<td>510</td>
<td></td>
</tr>
<tr>
<td>Kusch, M.</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Kuteva, T.</td>
<td>497</td>
<td></td>
</tr>
<tr>
<td>label</td>
<td>133</td>
<td></td>
</tr>
<tr>
<td>labelled data, 458</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labov, W.</td>
<td>511, 512</td>
<td></td>
</tr>
<tr>
<td>Lakoff, G.</td>
<td>19, 541</td>
<td></td>
</tr>
<tr>
<td>Lakoff, R.</td>
<td>504</td>
<td></td>
</tr>
<tr>
<td>lambda, 369, 372, 381, 383</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lambda calculus, 237, 246–247, 254–255</td>
<td></td>
<td></td>
</tr>
<tr>
<td>typed, 68, 70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>untyped, 68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lambda conversion, 70, 71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lambda reduction, 71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lambda-categorial language, 179</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lambek calculus, 68–71, 73, 81, 85, 89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(standard) intepretation of, 72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lambek, J.</td>
<td>71, 73, 81n, 85</td>
<td></td>
</tr>
<tr>
<td>Lancelot, C.</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Langton, R.</td>
<td>521</td>
<td></td>
</tr>
<tr>
<td>language acquisition, see</td>
<td>acquisition,</td>
<td></td>
</tr>
<tr>
<td>first-language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>language and gender, 533</td>
<td></td>
<td></td>
</tr>
<tr>
<td>language as a natural object, 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>language engineering, 89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>language faculty, 5, 420</td>
<td></td>
<td></td>
</tr>
<tr>
<td>language faculty (broad sense (FLB)), 478</td>
<td></td>
<td></td>
</tr>
<tr>
<td>language faculty (narrow sense (FLN)), 478, 479</td>
<td></td>
<td></td>
</tr>
<tr>
<td>language games, 16, 271</td>
<td></td>
<td></td>
</tr>
<tr>
<td>language learning, see</td>
<td>acquisition first-language</td>
<td></td>
</tr>
<tr>
<td>language model, 156–157</td>
<td></td>
<td></td>
</tr>
<tr>
<td>language of thought, 375, 419, 421, 425; see representationalism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>language specific learning biases, 446</td>
<td></td>
<td></td>
</tr>
<tr>
<td>language vs. speech, 439</td>
<td></td>
<td></td>
</tr>
<tr>
<td>larynx, 490, 491</td>
<td></td>
<td></td>
</tr>
<tr>
<td>last resort, 102</td>
<td></td>
<td></td>
</tr>
<tr>
<td>law of large numbers, 457</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
learnability, 447, 480
learnable class, 448, 449
learnable languages, 448
learned arbitrary signals, 487
learning algorithm, 448
learning biases, 468
learning by forgetting, 405
learning, computational 445
Leavens, D. A., 488
left extraction, 79, 81, 83, 87, 88
left hemisphere, 490
left node raising, 83, 84
Levine, R. D., 494
Lewis, D., 21
lexical-functional grammar, see grammar, lexical-functional (LFG)
lexical mapping theory, 77
lexical semantics, 178, 273, 309
lexicalized grammar, see grammar, lexicalizedlexicon, 268, 297–301, 305–309
generative, 342
lexicon-syntax continuum, 495
LFG, see grammar, lexical-functional
Liberman, A. M., 490
Lieberman, P., 477, 490, 538
linear logic, see logic, linear
linear precedence, 81
linearity, 127
linguistic anthropology, 547
linguistic nativism, 447
linguistic relativism, 16, 25, 26, 532
linguistic sign, 418, 419
linguistic turn, 1, 6
linguistic universals, see universal, linguistic
link relation, 389
linked trees in DS, 389
LISP, 68
LNR, see left node raising
locality, 366, 390, 393
localized probabilistic context-free grammar (LPCFG), see grammar, localized probabilistic context-free
locutionary act, 13
LOFT, see tree logic
logic, 9, 13, 21, 23, 25, 68, 73
dynamic predicate, 236, 378–379
 glue, 265
higher-order, 68, 75, 179, 237
intensional, 75, 263
intuitionistic, 68, 69, 87
linear, 71, 77, 85
modal, 386
predicate, 65, 68, 292, 293, 485
propositional, 67, 179
tree, LOFT, 385, 386
typed feature, 81
logic of finite trees (LOFT), 385, 387
logic programming, 89
logical consequence, 65
logical form, 19, 236–237, 241, 249–261, 263–264, 394n
logical grammar, 64
logical rule, 66
logical semantics, see semantics, logical
logical syntax, see syntax, logical
logical words, 178
logocentrism, 18
LOT, see language of thought, 425
Lyotard, J. F., 16

machine learning, 152, 468
machine translation (MT), 143, 144, 149, 155
interlingual, 146
speech-to-speech, 146
statistical, 145, 162
surface realization algorithm for, 150–151
MacLarnon, A., 491
magical formulas, 534
magicians, 543
Malinowski, 540
Malinowski, B., 534
Maltz, D., 506
mammalian hearing, 489
markedness theory, 534
Markov process, 156
Marten, L., 385
Martin-Löf, 275–276
masculine generic, 525
mass, 186
massive storage, 498
materialism, 119
mathematical linguistics, 143, 144
mathematics, 4, 5, 7, 19, 21, 24–26
Mattingley, I. G., 490
maxims, Gricean, 13, 223
Maximum Likelihood Estimation (MLE), 456, 469
Maxwell, J. T., 387
McConnell-Ginet, S., 515, 521, 525
McDowell, J., 16
McNeill, D., 547
McVay, S., 498
Mead, G. H., 14
meaning, 6–8, 10, 14–20, 23–27, 271–273, 375, 376, 465, 482; see semantics
content, 515–520
discursive, 520, 526
indexical 510, 511
lexical, see semantics, lexical representations of, see representationalism, representations of meaning
social, 500, 516; see gayspeak
sentence, 210
speaker, 200, 519
Grice's definition of, 521
meaning postulates, 177
measurement, 190
membership oracle, 466
memorandum, 145
memory, episodic 484
memory

chimpanzee auditory 490
increased storage of 480
limitation of, 477, 495
retrospective, 485
mentalism, see representationalism
Menzel, C., 484
mereology, 179, 190
Merge, 128, 133, 134
external, 130
internal, 130
message, 537
metacognition, 484
metaphor, 541
metaphysics, 6, 175, 179
metarule, 78
methodological dualism, 120
methodological vs. substantive minimalism, 110
Meurers, W. D., 494
Meyer-Viol, W., 386
mind design, 117
mind-external, 408, 409, 412
mind-internal, 403, 408, 409, 411, 412
mind-reading, 488
mind/body phenomenon, 63
Minimal Link Condition, 104
Minimalism, 99, 374n
minimalist program, 99, 358, 394n, 478, 494
minimum description length, 457
misdirection, 543
MIX, 466
modal, see modality
modal subordination, 261
modality, 253–259, 261–264
epistemic, 262–263
necessity, 263
possibility, 263
model
abnegation of, 149–150
bag-of-words, 158
channel, 157
distortion, 163
hidden Markov (HMM), 148, 161–162
knowledge-lean, 149; see model, statistical
knowledge-rich, 147–149
probabilistic, 170; see model, statistical
semantic, 180–183
statistical, 145–147
model structure, 176
model-theoretic semantics, see semantics, model-theoretic
modification, 230
adjectival, 249
adverbial, 262
predicate, 242
spatial, 259, 262–263
temporal, 255, 258–259, 262–263
verbal, 254–259
modules of the grammar, 102
modularity,
emergence of, 425
monkey, 498
Mönnich, U., 10
monotonicity, 66
Montague Grammar, 375
Moortgat, M., 85, 369
Moro, A., 132
morpheme, 164, see also morphological representations
morphological analysis, 164
morphological representations, 164, 358, 374n, 389n
morpho-syntax, 482
Morris, C., 11, 23
movement, 169
copy theory of, 129
multiplicative, 85
multistratality, 77
multivocality, 520–521, 524–526
mutual understanding, 487; see common ground, context, discourse
n-gram language model, 156
n-grams, 146, 156–157
names, see proper names
Narrangansett, 532
narration, 542
native speaker intuitions, 359
nativism, 120
linguistic, 447
natural adequacy, 98
natural deduction, 68, 69
intuitionistic, 70
natural kinds, 9, 10
natural language metaphysics, 175
natural language processing, 468n
natural language understanding, 153
natural philosophy, 116
natural sciences, 2, 4, 5, 15, 24, 25
natural selection, 100, 478
naturalism, 4, 5, 14, 15, 24, 407
naturalist, 411, 420, 421
naturalistic inquiry, 99
Neale, S., 9
negative evidence, 451
negative polarity, 366
neo-Darwinian synthesis, 100
neural nets, 89
new ethnography, 534
new theory of reference, 10
newborn human infants, 490
Newton, Sir I., 123
Nicaraguan sign language, 497
node raising
multiple, 83
noise, 536
“noisy” channel model, 157
non-commutative intuitionistic linear logic, 72, 85, 89
non-context fee languages, 466
non-deterministic languages, 463
non-terminals, 465
<table>
<thead>
<tr>
<th>Index</th>
<th>567</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-verbal communication, 545</td>
<td></td>
</tr>
<tr>
<td>nonmonotonic, 344</td>
<td></td>
</tr>
<tr>
<td>nonmonotonic logic, 327, 343</td>
<td></td>
</tr>
<tr>
<td>normativity, 24</td>
<td></td>
</tr>
<tr>
<td>nouns, 183</td>
<td></td>
</tr>
<tr>
<td>numbers, 186</td>
<td></td>
</tr>
<tr>
<td>object-permanence, 484</td>
<td></td>
</tr>
<tr>
<td>Ochs Keenan, E., 506n</td>
<td></td>
</tr>
<tr>
<td>Ogden, C. K., 8, 483</td>
<td></td>
</tr>
<tr>
<td>one-to-one correspondence, 463</td>
<td></td>
</tr>
<tr>
<td>ontogeny, 490</td>
<td></td>
</tr>
<tr>
<td>ontological, 191</td>
<td></td>
</tr>
<tr>
<td>operation, 65</td>
<td></td>
</tr>
<tr>
<td>operator</td>
<td></td>
</tr>
<tr>
<td>conditional, 236, 238–239</td>
<td></td>
</tr>
<tr>
<td>iteration, 258</td>
<td></td>
</tr>
<tr>
<td>opposition, 3</td>
<td></td>
</tr>
<tr>
<td>oration, 542</td>
<td></td>
</tr>
<tr>
<td>outside world, 482</td>
<td></td>
</tr>
<tr>
<td>over-generalisation, 467</td>
<td></td>
</tr>
<tr>
<td>overlap, 181</td>
<td></td>
</tr>
<tr>
<td>P-marker, 58</td>
<td></td>
</tr>
<tr>
<td>PAC-learning, 457</td>
<td></td>
</tr>
<tr>
<td>pair formation, 70</td>
<td></td>
</tr>
<tr>
<td>Pāṇini, 151</td>
<td></td>
</tr>
<tr>
<td>paradigmatic, 160</td>
<td></td>
</tr>
<tr>
<td>paradox, 543, 545</td>
<td></td>
</tr>
<tr>
<td>parametric, 461</td>
<td></td>
</tr>
<tr>
<td>paraphrase, 63, 154</td>
<td></td>
</tr>
<tr>
<td>parrot, 485</td>
<td></td>
</tr>
<tr>
<td>parsing, 150, 151, 164, 359, 385, 387–389, 392, 393, 396, 468, 496</td>
<td></td>
</tr>
<tr>
<td>statistical, 165, 167–168</td>
<td></td>
</tr>
<tr>
<td>parsing algorithm, 151</td>
<td></td>
</tr>
<tr>
<td>parsing-as-deduction, 89</td>
<td></td>
</tr>
<tr>
<td>part-of-speech tagging, 159</td>
<td></td>
</tr>
<tr>
<td>Partee puzzle, 273, 304</td>
<td></td>
</tr>
<tr>
<td>Partee, B., 9</td>
<td></td>
</tr>
<tr>
<td>partial structure, 365, 387, 388, 392, 396</td>
<td></td>
</tr>
<tr>
<td>parts of speech, 125, 183</td>
<td></td>
</tr>
<tr>
<td>pattern variables, 538</td>
<td></td>
</tr>
<tr>
<td>patterns</td>
<td></td>
</tr>
<tr>
<td>meaningful, 520–521, 523–524</td>
<td></td>
</tr>
<tr>
<td>Payne, R. S., 498</td>
<td></td>
</tr>
<tr>
<td>Peano, G., 64</td>
<td></td>
</tr>
<tr>
<td>Peirce, C. S., 11, 118, 536</td>
<td></td>
</tr>
<tr>
<td>Penn Treebank, 468</td>
<td></td>
</tr>
<tr>
<td>Pepperberg, I. M., 485</td>
<td></td>
</tr>
<tr>
<td>perception, 275, 486, 489</td>
<td></td>
</tr>
<tr>
<td>Peregrin, J., 6, 22, 25</td>
<td></td>
</tr>
<tr>
<td>perfect, 119</td>
<td></td>
</tr>
<tr>
<td>perfection, 112, 116</td>
<td></td>
</tr>
<tr>
<td>performative, 210, 212</td>
<td></td>
</tr>
<tr>
<td>performer, 542</td>
<td></td>
</tr>
<tr>
<td>perlocutionary act, 13</td>
<td></td>
</tr>
<tr>
<td>permutation, 66</td>
<td></td>
</tr>
<tr>
<td>PF deletion, 382</td>
<td></td>
</tr>
<tr>
<td>pharynx, 490</td>
<td></td>
</tr>
<tr>
<td>phase structure grammar, 45</td>
<td></td>
</tr>
<tr>
<td>phases, 109</td>
<td></td>
</tr>
<tr>
<td>philosophy of language, 117</td>
<td></td>
</tr>
<tr>
<td>philosophy of phonology, 403</td>
<td></td>
</tr>
<tr>
<td>phoneme, 404–407, 412–414, 416, 492, 493, 533</td>
<td></td>
</tr>
<tr>
<td>grounding of, see groundedness of phonology, 408, 417</td>
<td></td>
</tr>
<tr>
<td>phonetic segments, 492</td>
<td></td>
</tr>
<tr>
<td>phonetics, 481, 489</td>
<td></td>
</tr>
<tr>
<td>phonetics/phonology distinction, 403</td>
<td></td>
</tr>
<tr>
<td>phonological knowledge, 403</td>
<td></td>
</tr>
<tr>
<td>innate, 403</td>
<td></td>
</tr>
<tr>
<td>phonological representations, 357, 358, 361, 370, 372n</td>
<td></td>
</tr>
<tr>
<td>phonological units, 492</td>
<td></td>
</tr>
<tr>
<td>phonology, 358, 481, 482, 491</td>
<td></td>
</tr>
<tr>
<td>P/phrase-/markers, 47</td>
<td></td>
</tr>
<tr>
<td>phrase structure, 304</td>
<td></td>
</tr>
<tr>
<td>phrase structure grammar, 166</td>
<td></td>
</tr>
<tr>
<td>phrase-marker, 48n, 57</td>
<td></td>
</tr>
<tr>
<td>phylogeny, 490</td>
<td></td>
</tr>
<tr>
<td>physicalism, 119, 127, 415</td>
<td></td>
</tr>
</tbody>
</table>
physiology, 2, 5
Pinker, S., 5n, 478, 489
Pirahê, 186
pivot language, 147
plans, 485
Plato, 6
plurals, 186
Podesva, R., 510, 515
poetic function of language, 537
poetry, 541
pointing, 488
Pollard, C. J., 81, 494
polynomial, 454
polynomial function, 460
polysemeic, 546
positive evidence, 450
possible individuals, 192
possible worlds, 180
posterior probability, 155
postmodernism, 16
poststructuralism, 12, 18, 19
poverty of the stimulus, 5, 479
PP attachment, 470
pragmatic intent, 496
pragmatics, xvii, 23, 213, 215–218, 385, 481, 482, 486, 487
pragmatism, 12, 14–16, 23
Prague School, 534
Prawitz, D., 71
precedence, 181
precision, 468
predicate, 485
predicate letter, 65
predicate logic, see logic, predicate
premises, 66, 68
prepositions, 497
presentation of $L$, 449
prestige, 489
presupposition, 177, 212–215, 230, 238–243, 245–251
accommodation of, 239, 248–249, 251
binding of, 239, 248
clausal, 249

justification, 245
ordering, 250
parameter, 245–246
speech-act analysis of, 214
type, 238, 249, 251, 258, 264
primary linguistic data (PLD), 445
primates, 546
primitive languages, 478
primitives, 472
principle of compositionality see compositionality
Principles and Parameters (P& P), 95, 99, 461
prior probability, 155
private (proto-)concepts, 486
private concepts, 483, 484
probabilistic context free grammar (PCFG), see grammar, probabilistic
context-free (PCFG)
probabilistic learning, 454
probability
Bayesian, see Bayes’ rule
conditional, 155
joint, 155
lexical transfer, 163
maximum, 160
theory of, 145, 455
word emission, 160–162
probability distribution, 457
probably and approximately correct (PAC), 457
processes, 188
processing, 361
production, 90, 359, 388, 392
productivity, 495, 538
projection, 70
projective dependency analysis, 169
pronoun, 361–366, 371, 376–379, 387, 391, 393; see anaphora
coreferential 361, 364
E-type 362, 376, 377
bound variable, 361, 362, 367
proof normalisation, 71
proof-theoretic, 145
proofs, 71
proofs-as-programs, 70
proper names, 198–199, 298
properties, 178, 182
propositional content, 496
propositional language, see logic, propositional
propositions, 117, 179, 276
propositions as types, 276
prospective memory, 485
proto-transformational derivation, 75
Proto-World, 480
protolanguage, 494
protracted telic events, 188
proverbs, 495
PS grammars, 41
psychological reality, 406, 472
psychology, 2, 5, 15, 20, 26
PTQ, 179
public labels, 483, 484
public symbolic labels, 486
punchline, 544
Putnam, H., 9, 10
P&P, 101
qualitative shift, 495
quantification, 308–309, 361, 362, 377, 379
quantitative limitation, 496
question answering, 153
question words, 497
questions, 215–218
  concealed, 232
  interrogative, 454
Quine, W. V. O., 14, 15, 21, 24, 98
Révész, G. E., 5n
RAM, 89
rational design, 115
rationalism, 116
rationalist, 118
rats, 485
real metaphysics, 192
realisation, 404, 407, 408, 410, 440
reasoning
invalid, 66
valid, see validity
relative clause markers, 497
Reby, T., 491
recall, 468
Recanati, F., 13
reconstruct, 497
record type, 272, 277, 278, 284, 285, 287, 291
records, 289, 295, 321
reciprocal altruism, 488, 489
recursion, 40, 44, 53n, 478
recursive languages, 450
recursively combining constructions, 498
recursively combining words, 494
reduced phrase marker, 59
reference, 6–11, 13, 15, 20, 21, 24, 26
causal theory of, 203–204
cluster theory of, 199–200
reflexivity, 65, 66
re-identification, principle of, 185
regular grammar, 74
regulators, 546
Reichenbach, H., 11
Reimer, M., 9
relation, 65
relative boundary, 392
relative clause, 360, 360n, 372, 381, 389, 390, 392, 393, 454
relative pronoun, 389, 390
release, 543
representation, 6, 7, 10–12, 404, 409, 412, 415, 419, 427, 432, 434, 436, 438, 440
levels of, 357
representational economy, 105
representationalism, 357–359, 369, 370, 372, 373, 375, 378, 379, 380, 382, 384, 385, 392, 393, 398,
representations of meaning, see representation, meaning
requirement(s) in DS, 387
Rescher, N., 23
residual languages, 463
resumption, 360n
rhythm, 489
Richards, I. A., 8, 483
right node raising, 77, 78, 80, 81, 83
RNR, see right node raising
Rogers, H., 506, 507
role and reference grammar, 373
Rorty, R., 1n, 15, 16, 23
Rounds, W. C., 81
Rousseau, J.-J., 498
Ruhlen, M., 480
rule-to-rule hypothesis, 370
rule-following, 16, 17
Rumbaugh, D., 538
Russell, B., 7–9, 12, 98
Ryle, G., 13
s-structure, 77
Sag, I. A., 81, 494
saltation, 495
Sandbothe, M., 12
Sanskrit, 531
Sapir, E., 26, 533
Sapir-Whorf hypothesis, 419, 533
Saussure, F. de, 2–4, 63, 477, 531, 536
Savage-Rumbaugh, S., 486, 490, 538
Schwartz, B. L., 484
scrub jays, 485
Searle, J., 24
Sebeok, T., 12, 537
second-order fact, 486
second-order judgements, 485
secondary rationalisations, 532
selectional restrictions, 238–240
self-organisation, 492, 493, 498
Sellars, W., 16, 17
semantic component, 103
semantic interpretation, see meaning;
semantics
semantic parameters, 177
semantic representation, see representations of meaning
semantic role, 169
semantic role labelling, 170
semantic structure, 77; see representation, meaning
semantic triangle, 8
semantic values, 176
semantically compositional combinatoriality, see semantics, compositional
semantically compositional syntax, see syntax; semantically compositional
semantics, 23, 481, 482
compositional, 272, 273, 309, 494, 496
context-sensitive, 237
continuation-style, 237, 246
denotational, 243
determiner, 247
dynamic, 236–243, 522
event, 179, 180, 187, 254
frame, 273, 304–311
intensional 8
lexical, 178, 229–267, 273, 289–305, 309,
logical, 64, 75, 77, 89
model-theoretic, 21, 25, 63, 64, 201, 375
possible-world, 198–200
pragmatics vs., 218
programming language, 236
situation, 236
Tarskian, 236
temporal, 265
truth-conditional, 197–210, 217, 237, 244, 282, 375
two-dimensional 205, 208
type-driven, 229, 245–253
semigroup, 72
semiotics, 6, 10–12
sense, 7–9, 13, 182
sententialist, xv
sentiment analysis, 154
separability parameter, 462
sequent, 66, 72
sequent calculus, 66–68, 71, 74, 86
set-theory, 179
setup, 543
‘sexist’ language, 517–521
sexual dimorphism, 491
sexual selection, 491
sexuality
  definition of, 503
Seyfarth, R. M., 484
Shannon, C., 157, 536
Shiffer, S., 20
shortest derivation requirement, 105
side-formula, 66
Σ, F grammar, 39, 40, 44, 48, 56
sign, 63, 81, 545; see linguistic sign
sign language, 547
signified, 63, 81, 89
signifier, 63, 81, 89
silence, 541
simple type accommodation, 248
simulated computationally, 493
single-cycle generation, 106
singly transformations, 48n
situation semantics, 297
situations, 181, 276
sloppy reading, see ellipsis, sloppy readings of
smiles, 546
Smith, J. D., 484
Smyth, R., 506, 507
social convention, 403, 426, 431
social distance, 546
social meaning, see meaning, social
socially constituted linguistics, 540
socially realistic linguistics, 540
socially-established convention, 428, 429
sociolinguistics, 538, 540
sociology, 2
songbirds, 498
sorts, 186
soundness, 66, 73
sparseness of word distribution, 157
spectacle, 542
speech, 489, 492
speech act, 12, 13, 210–219, 522, 544
speech circuit, 63
speech communities, 547
speech event, 272, 295, 297, 318
speech perception, 489, 492, 498
speech processing, 490
speech production, 489–491, 498
speech recogniser, 157
speech sound type, 404, 414
Sperber, D., 13, 487
‘splitt’ utterance, 365, 367, 374, 382, 392
stages, 178, 185
states, 187
statistical modelling, 457
statistical techniques in computational linguistics, 145
stative, 260–261
Steedman, M., 81, 83
Stokhof, M., 27
strategic behaviour, 541
stratification, 276, 281
Strawson, P. F., 9
strict reading, see ellipsis, strict readings of
string theory of events, 273
strings, 294
strong generative capacity, 34, 35, 37, 45, 59, 465
structural analogy, 417, 418
structural operator/modality, 85, 87
structural rule, 66, 71, 74
structuralism, 3–5, 18, 21, 358, 394n
structure, 2–6, 18, 25
style
  persona, 513
subcategorization, 78, 81
subformula property, 68, 74
subjacency, 55n
subject, 497
substantive minimalism, 111
substitutable languages, 453
substitution, 70
  free, 70
succedent, 66, 68, 72
successor function, 125
summarisation, 154
  speech and text, 154
supervaluation 208–209
supervised learning, 468
supra-finite classes, 451
surface realisation, 150
surface structure, 74
surprise, 542, 543
Svoboda, V., 27
syllable structure, 492
symbol, 8, 11, 16, 479, 541, 545
symmetry, 131
synchronic, 477, 480
syntactic categories, 494, 497
syntactic complexity, 498
syntactic concept lattice, 454, 466
syntactic congruence, 464
syntactic movement, 371, 372, 374n, 388
syntactic representations, see syntactic structure
syntactic structure, 361, 363, 369, 370, 372, 373, 379, 384, 394, 396
syntagmatic, 160
  formal, 63, 64
  logical, 64
  semantically compositional, 498
tagging, 148
  part-of-speech, 148, 159–162
tagset, 159
tamarin monkeys, 489
Tannen, D., 506, 541
target language, 147
Tarski, A., 21–23, 65
tense, 179, 299, 340, 363, 386n, see semantics, temporal
tension, 543
ter Meulen, A., 25
term, 69–71
term insertion, 75
textual entailment, 153
theatrical performance, 542
theorem, 66, 73
theory of mind, 487, 489, see representationalism
they
  notionally singular, 505n
things, 183
thought, 134, 486
time structures, 545, see semantics, temporal
Tincoff, R., 489
tit-for-tat, 488
TLCG, see grammar, categorial, type-logical
toddlers, 488
Todt, D., 498
token, 414
token frequency, 423, 435
Tomasello, M., 488
topic, 77
topic-comment structure, 497
topicalization, 77–79
tractability, 460
tractable cognition thesis, 460
traditional transmission, 538
training data, 471
transduction, 404, 410, 411
transformational grammar, 74, 358, 373, 394n, 535
  proto-, 74
transformations, 44, 51
transitivity, 65, 66
translation, 534
transmogrification, 411
Traugott, E. C., 497

Tree growth, 361, 367, 385, 386–394
Tree structures of sentences, 494
Tree-adjoining grammar, 167
Treichler, P., 519

Trigram language model, 156

Trivers, R. L., 488

Truth, 15, 21, 22, 200, 368–370, 372, 374, 375, 379, 394, 396n, 399
Truth value, 7, 21, 176
Truth-conditions, see semantics, truth-conditional

TTR, 273, 277, 288, 317, 320, 321

Turing machine, 68, 89, 101

Turner, K., 24

Twin Earth, 203

Type, 69, 71, 72, 243–245
Constraint, 251
Coordination, 252–253
Event, 251
Informational object, 244, 246
Modifier, 246
Physical object, 244, 246–247
Polymorphic, 245, 251, 253, 258, 261
Underspecified, 252

Type domain, 69

Type equivalence 291–292

Type frequency, 435

Type-logical categorial grammar, see grammar, categorial, type-logical

Type raising, 83
Type shift, see coercion
Type theory, 273, 278
Type theory with records, see TTR
Type/token, 404, 405, 407–409, 435

Typed feature logic, see logic, typed feature

Typed lambda calculus, see lambda calculus

Uchida, H., 90

UG, see universal grammar

Ullman, J. D., 5n

Unannotated, 468

Unbounded, 454

Unbounded dependency, see dependency, unbounded

Underspecification, 267–268, 367, 368, 375, 385, 386n, 387, 388, 390, 393, 397

Unfixed node, 387

Unification, 344–350

Unification grammar, 81

Unification problem, 124

Uniformitarian, 478

Universal grammar (UG), 98–99, 420, 479

Universal, linguistic 437

Unrestricted/Turing powerful grammar, 74

Unsupervised grammar induction, see grammar induction

Unsupervised learning, 469

Update, 368, 374, 375, 384, 385, 387, 388, 390, 391, 393, 397

Updating, 521, 522

Upper bound, 452

Uptake, 521, 522

Usage-based approaches to phonological knowledge, 403

Usage-based phonology, 434

Use theories of meaning, see speech acts

Vagueness, 208

Validity, 66, 73

Valuation, 70

Value, 3

Van Benthem, J. F. A. K., 25, 85

Variable, 70
Bound, 70
Free, 70
Linguistic, 511

Variant
Linguistic, 511

VC-dimension, 458
vector space, 158
verbal agreement, 497
verbal art, 541, 542
verbal aspect, see aspect
verbal modifiers, see modification, verbal
verbs, 183
main, 492,
semantics of, 289–305; see semantics, temporal
vervet monkey, 484, 486
vocabulary, 492
vocal motor scheme, 427
vocal tract, 489–491
vocal tract normalization, 489
vocalization, 491
voice
phonetic properties of, 507
voice onset time, 489
voluntary control, 484, 487, 491
von Heusinger, K., 9
von Humboldt, W., 26
vowel inventories, 493
vowel qualities, 490, 491
vowel space, 493
vowels, 493

Wall Street Journal, 468
Washburn, D. A., 484
Washoe, 538
Watson, J., 484
Waxman, S., 486
weak bias learning, 471
weak generative capacity, 34, 38, 40, 46, 52
weak generative power, 465
weakening, 66, 68
Weaver, W., 145, 536
wellformedness
semantic, 243–244
Werning, M., 19
Westerståhl, D., 20
wh-questions, 470
whales, 498

Whorf, B. L., 26, 175, 533
Whorfian hypothesis, 533
Wierzbicka, A., 25
Willias, R., 532
Wilson, D., 13, 487
Wittgenstein, L., 12, 15–17, 26, 271–273
women’s language, 504, 510
word meaning, 273, 311, see semantics, lexical
word-sense disambiguation, 158
WordNet, 158
Xu, F., 486
Youngman, H., 543
Zaenen, A., 77
Zipf’s first law, 157
Zoloth, S. R., 490
Zuberbühler, K., 498
Zuidema, W., 492