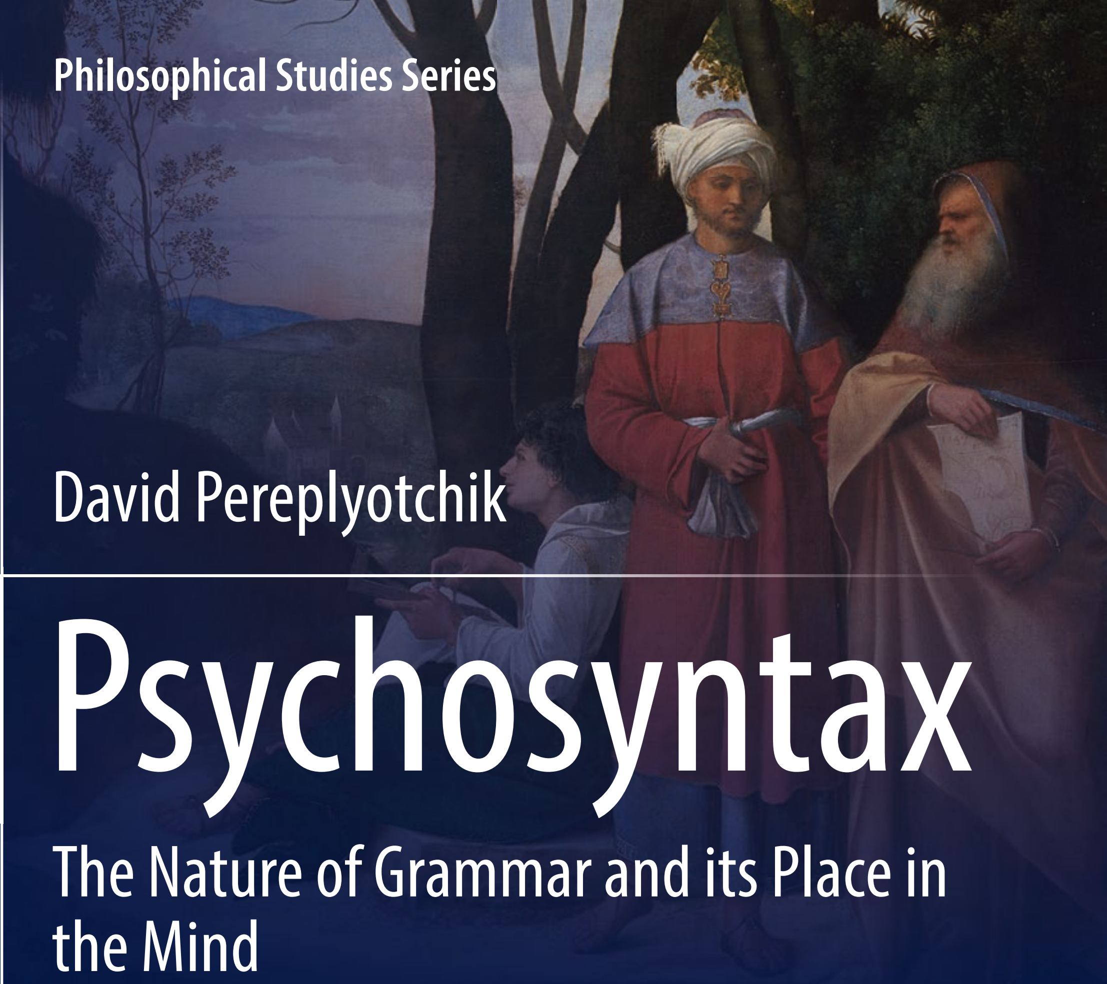


Philosophical Studies Series

David Pereplyotchik

Psychosyntax

The Nature of Grammar and its Place in the Mind

Springer

Philosophical Studies Series

Volume 129

More information about this series at http://www.springer.com/series/6459

David Pereplyotchik

Psychosyntax

The Nature of Grammar and its Place in the Mind

Springer

David Pereplyotchik
Department of Philosophy
Kent State University
Kent, OH, USA

Philosophical Studies Series
ISBN 978-3-319-60064-2 ISBN 978-3-319-60066-6 (eBook)
DOI 10.1007/978-3-319-60066-6

Library of Congress Control Number: 2017948924

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To my family and friends—
past, future, and possible.

Acknowledgements

The present work started out life as my Ph.D. dissertation, *The Psychological Import of Syntactic Theory* (2012), which I wrote with the help of funding provided by the Mario Cappelloni Dissertation Fellowship. To Michael Devitt I owe the debt of seeing that project through with me, despite many substantive disagreements, from which I invariably learned a great deal. Michael was responsible for getting me excited about the psychological reality of grammar in the first place, and also for getting me to rethink the cumbersome title for the book version. From Janet Dean Fodor, I learned more about psycholinguistics than I had ever dreamt possible, and felt wonderfully encouraged to pursue this path. Major sections of this work were written with the benefit of conversations with faculty in the philosophy and linguistics departments at both CUNY and NYU, including Dianne Bradley, Marcel Den Dikken, Bob Fiengo, Michael Levin, Stephen Neale, Gary Ostertag, Paul Horwich, David Rosenthal, Stephen Schiffer, and Christina Tortora. Jennifer Corns kindly read and commented on early drafts of the entire dissertation, and many others commented on portions of the book manuscript, including Endre Begby, John Collins, Daniel Harris, Fritz Macdonald, Russell Marcus, Hal Morris, Frank Pupa, Georges Rey, and three anonymous reviewers for Springer. I am grateful to them, and to the audiences at innumerable conferences and symposia, for helping me to whip this thing into its final shape. Finally, my biggest personal debts are to my family and friends, who have shepherded me through this journey, bringing two key elements—meaning and structure—into my life. And a great deal of joy besides.

The journal *Philosophical Psychology* has kindly permitted the reprinting of major portions of Chap. 6. That material, with scattered stylistic variations, appears also as the following:

"Representations in Language Processing: Why Comprehension is not Brute-Causal," *Philosophical Psychology*, 29 (2), 2016, pp. 277–291.

Contents

Abbreviations

ATN	Augmented transition network
CFG	Context-free grammar
CGR	Covering grammar realization
CI	Conceptual-intentional system
CTM	Computational theory of mind
DCG	Definite clause grammar
DTC	Derivational theory of complexity
GB	Government and binding
GPSG	Generalized phrase structure grammar
HPSG	Head-driven phrase structure grammar
HSPM	Human sentence processing mechanism
LC	Late closure
LF	Logical form
LFG	Lexical functional grammar
LOTH	Language of thought hypothesis
MA	Minimal attachment
MCP	Minimal chain principle
MPM	Mental phrase marker
P&P	Principles and parameters
PAD	Parsing as deduction
PCFG	Probabilistic context-free grammar
PF	Phonological form
PLR	Precomputed lexical realization
PPP	Preliminary phrase packager
RTM	Representational theory of mind
SBC	Strong brute-causal model
SM	The sausage machine model
SSS	Sentence structure supervisor
TAG	Tree adjoining grammar

Introduction

Historical Background: Recent Shifts in the Philosophy of Cognitive Science

Cognitive science was the outcome of several major theoretical advances in the 1950s and early 1960s. The development of recursive function theory in the formal sciences (Rogers 1967), as well as advancements in electrical engineering, led to the construction of physical computing machines with impressive information-processing abilities (Newell and Simon 1963). At the same time, the operant conditioning paradigm in psychology was shown to be methodologically groundless (Fodor 1968) and both predictively and explanatorily barren (Chomsky 1959). These developments in turn gave rise to a variety of functionalist proposals in philosophy (Armstrong 1968; Dennett 1969, 1978; Fodor 1968; Lewis 1972; Putnam 1960, 1967). These were all significant intellectual achievements, by any measure. But it is arguable that chief among them—indeed the thread that bound them all together—was Noam Chomsky's development of generative grammar in linguistics (Chomsky 1957, 1965). Here, finally, was a plausible research program that promised to yield fruit in a domain that is absolutely central to our conception of human psychology.

Early excitement in the 1960s and 1970s gave rise to rapid progress in the theory of formal syntax (Chomsky 1965, 1972, 1975b, 1977), which in turn fostered the development of detailed parsing models in computer science (Wanner and Maratsos 1978; Marcus 1980). Mounting psycholinguistic data made it possible to ask increasingly probing questions about the relationship between generative grammars and computational models of real-time parsing (Fodor et al. 1974; Bresnan 1978; Berwick and Weinberg 1984). Philosophical debate thus arose about the place of transformational grammar in the mind.

By the 1980s, such debates in turn generated questions about cognitive architecture, particularly issues concerning the modularity of cognitive systems (Fodor 1983). Insights from linguistics were applied also to theories of "higher" cognition, including reasoning, planning, and decision-making. The language of thought hypothesis (LOTH) and the associated computational theory of mind (CTM) were formulated clearly enough to warrant investigation into the "semantics" of psychological states (Fodor 1975, 1987). In the early 1990s, these advances led many researchers in psychology and philosophy to formulate increasingly detailed proposals about the nature of concepts (Peacocke 1992; Brandom 1994; Fodor 1998) and other types of mental representation.

The emerging framework of connectionist computation, or parallel distributed processing (PDP), served to sharpen the issues and made clear, once again, the need for an explicit account of the relation between cognitive processes and neural implementation (Clark 1991; Churchland 1995). Impressive connectionist models of various cognitive processes came to be developed. Given the long-standing centrality of language to our conception of human cognition, it was only natural that questions would arise about the relation between these models and the latest incarnations of generative grammar—government and binding theory (GB), lexical functional grammar (LFG), head-driven phrase structure grammar (HPSG), and the like.

Philosophical debates around this time centered on the extent to which systematicity and productivity are real properties of cognition—again, most centrally, of language processing—and whether neural networks would be capable of modeling the compositional operations that seemed necessary to explain whatever degree of systematicity and productivity there might be (Davies 1987, 1989; Macdonald and Macdonald 1995).

From the point of view of a philosopher of mind, everything changed in the mid-1990s. The topics of central concern shifted markedly, as debates about consciousness took center stage (Chalmers 1990; Dennett 1991; Block 1996). Concomitantly, interest in propositional, intentional, and linguistic cognition declined. A concerted effort was launched to solve outstanding problems in the theories of sensation, perception, attention, emotion, and self-knowledge—anything having to do with the qualitative or so-called "phenomenal" states of mind. This agenda continues to dominate the field today. At present, when philosophers of mind advance substantive claims about intentionality, it is often to motivate some proposal about "cognitive phenomenology" or "phenomenal concepts." At the same time, research into the neural correlates of consciousness, perception, and attention is burgeoning, providing philosophers with vast pools of data from cognitive neuroscience that bear on such positions as representationalism and disjunctivism.

Efforts to construct a plausible psychosemantics have all but ceased, despite the widespread recognition that nothing even resembling convergence has taken place and that no fully satisfactory theory of intentionality exists. Indeed, some theorists are explicit in their doubt that such a theory will ever come to light. Philosophers of language became preoccupied with debates that are, to a large extent, divorced from the concerns of computational modeling and neural implementation. Simultaneously, the computational branch of cognitive science—which once employed a great many philosophers and "GOFAI" enthusiasts—was largely taken over by mathematically minded specialists in dynamical systems theory and other frameworks of statistical modeling. Such theorists are drawn to information-processing models that are often characterized as nonlinguistic and nonsymbolic; their interest lies more with the "embedded/embodied" conception of cognition.[1]

This shift has had the unfortunate effect of pushing into the background foundational issues at the interface between philosophy, on the one hand, and computational psychology, formal linguistics, and experimental psycholinguistics, on the other. Relatively few philosophers today have an abiding interest in issues concerning the psychological reality of grammar—what was once a hot topic in the philosophy of language and mind. Fewer still have in-depth knowledge of the formalisms employed in contemporary syntactic theory and of their bearing on computational-level descriptions of various parsing models. Yet such models are continually being developed and refined. And the philosophical issues they present are as pressing as ever. Indeed, coupled with the accumulated knowledge now available in psycholinguistics and neurolinguistics, these models can shed a great deal of light on the psychological reality of various theoretical constructs from formal linguistics.

The Contemporary Psychological Reality Debate

Chomsky (1965) first made explicit what is now the dominant view concerning the objects and aims of linguistic inquiry. Rather than studying the sounds and inscriptions that we produce and comprehend, or the social conventions that govern linguistic usage, Chomsky argued that the primary target of linguistic theorizing must be the *tacit knowledge* that underlies every competent speaker-hearer's linguistic competence. On a common and natural interpretation, this knowledge is encoded in *mental representations* of grammatical principles.

Chomsky's "cognitivist" conception of linguistic inquiry generated a great deal of excitement and debate in many areas of cognitive science, particularly in philosophy. Philosophers immediately questioned the coherence of the notion of tacit knowledge. What sort of knowledge can be neither learned nor taught, neither spoken nor recollected (Quine 1970; Stich 1971)? And why should we count as knowledge something that does not play a role in our everyday practical and theoretical inferences? (Most New Yorkers know that there's great pizza in Brooklyn. Do they really *know* that only the features on functional heads are subject to parametric variation?) Finally, what prevents us from seeing the linguists' construction of grammars as simply an effort to delineate the class of grammatical sounds and inscriptions (Soames 1984; Devitt 2006)? Why must linguistics traffic in any psychological notions at all?

A number of philosophers have defended Chomsky's approach by offering clarifications of the concepts of "mental representation" and "subpersonal mechanism" (Davies 1989; Peacocke 1989; Rey 2003). But it is safe to say that there is, at present, no consensus in the cognitive science community—and certainly none among

[1] See Anderson (2003), Clark (2006), Gallagher (2005), Gibbs (2003), Hutto and Myin (2012), Robbins and Aydede (2008), Shapiro (2004), and Wilson (2002).

philosophers—concerning the viability and conceptual coherence of Chomsky's program. Detailed analyses of many languages have, of course, been produced, and much more is now known concerning the character of language acquisition and processing. But the philosophical underpinnings of the project, particularly the notion of psychological reality, continue to defy generally acceptable explication.

In addressing these issues, I take as my point of departure Michael Devitt's recent book-length treatment of the debate. Devitt (2006) challenges nearly every aspect of Chomsky's theoretical framework. He argues that the objects of linguistic theory are public languages (construed "nominalistically" as sounds and inscriptions) and that the grammars emerging from formal syntax tell us only about the rules that govern those languages, not the cognitive mechanisms involved in acquiring or using them.

My own project can be seen as a critical response to both Chomsky's and Devitt's positions. The aim is to construct a philosophy of linguistics that combines what I take to be the best-supported aspects of each. On broadly philosophical issues—the epistemological status of syntactic theory, its ontological commitments, and its methodology—I side, in large part, with Devitt. Where we part company is over the psychological reality of syntactic rules and principles. Devitt (2006) argues that there is no evidence for the claim that successful grammars are represented, or even embodied, in the human mind/brain. Moreover, he entertains the view that comprehending and producing language is "a fairly brute-causal associationist process, rather than a process involving representations of the syntactic properties of linguistic expressions" (220). Drawing on a wide range of neurocognitive and behavioral studies, I marshal several lines of evidence against both of these claims.

Importantly, my arguments for the psychological reality of syntactic rules and principles make no appeal to the controversial innateness hypothesis (Chomsky 1965). My view is not, therefore, hostage to the outcome of ongoing debates between nativist and empiricist approaches to language acquisition (Cowie 1999; Laurence and Margolis 2001). My reliance on contemporary work in psycholinguistics extends, in large part, to studies of real-time language comprehension. This allows me to draw on established, up-and-running computational models in giving substance to technical notions like "tacit knowledge," "procedural rule," "cognitive module," and "subpersonal state." This is especially evident in Chaps. 8 and 9, where I undertake a historical survey of the coevolving fields of formal syntax and parsing theory.

Summary of the Findings

Before delving into the details of the arguments, let me set out the main conclusions of my research, in as nontechnical a fashion as the subject matter permits.

1. Contrary to the dominant Chomskyan position, the formal syntactician is *not* best seen as engaging in a distinctively psychological inquiry. Even the suc-

cessful grammars emerging from syntactic theorizing cannot be assumed without argument to be psychologically real. A great deal of support from psycholinguistic research must be marshaled before we can make substantive claims about the psychological role of a successful grammar. This is true even if we equate the success of a grammar with what linguists call "explanatory adequacy."

2. The explanatory adequacy of a grammar is best construed in the following neutral fashion: a grammar of some particular language is explanatorily adequate if it fits well with the maximally general, simple, and unified theoretical coverage of *all* human languages. Construed in this way, achieving explanatory adequacy is desirable *whether or not the cognitivist conception of linguistics is true*. Although explanatory adequacy is often seen as a proprietary notion within the cognitivist framework, I argue that it is in practice assessed *independently* of psychological data. The pursuit of explanatorily adequate grammars can and should be motivated by general methodological principles that have nothing in particular to do with the cognitivist conception of linguistics.
3. Reflection on the actual practice of both syntacticians and language acquisition theorists reveals their commitment to the reality of *public languages*. Although this commitment is often disavowed, there is simply no other way to make sense of the persistent and ineliminable references to public languages in the canonical texts of formal syntax and acquisition theory.
4. In practice, the primary target of syntactic theorizing is not an idealized idiolect—what Chomsky calls an "I-language"—but rather an idealized *speech community* and its public language. Public languages have some of the aspects of what Chomsky calls "E- languages," in that they are collections of sounds, marks, muscle movements, and the like. But the classes of these items that constitute one or another public language are delineated on entirely theory-internal grounds, *without* reference to political organizations or prescriptive goals. The individuation conditions on E-languages are, at present, not precise and require idealization away from messy variation within a community. But this is no less true of I-languages, whose individuation conditions are, at this stage of inquiry, only dimly understood. We should not expect to provide precise individuation conditions for either E- or I-languages in advance of sustained inquiry.
5. The public E-languages that syntacticians study are not "abstract objects" in the philosopher's sense. That is, they are not causally inert things that exist outside of spacetime, and they are not uncountably infinite. The aims of linguistics are in no way furthered by thinking of language in this "platonist" fashion. Linguists' claims concerning the infinitude of language do *not* signal an ontological commitment to abstract entities. Rather, they reflect the modal force of linguistic generalizations, as well as a principled idealization away from mortality, memory constraints, and the like. Nor is it useful to think of the truths of linguistics as being discovered through "nonempirical rational intuition" (Katz 1985, 2000).

6. Though the target of syntactic theorizing need not be a psychological state or mechanism, there is in fact a very strong case to be made for the claim that some class of grammars is psychologically real. The mind/brain of each competent speaker-hearer employs a specific set of syntactic rules or principles, which play an important role in the cognitive processes underlying language comprehension.
7. Comprehending linguistic input requires constructing and manipulating mental representations of its syntactic structure. Any psychological or computational model that eschews such representations will be unable to account for a vast range of behavioral and neurocognitive data. The models thus far proposed in this genre exhibit a striking pattern of failure.
8. Representations of incoming linguistic structure are *not* best seen as beliefs, thoughts, or even perceptual judgments. They differ in kind from all of the psychological states that figure in commonsense explanations of behavior. Unlike those states, the mental representations posited by the psycholinguist are invariably nonconscious, inferentially insulated from most beliefs, and inexpressible in speech. Moreover, ascribing such states requires making far less demanding normative assumptions. Whereas erroneous judgments and inferences leave a person open to rational criticism, misrepresentations in language processing are best seen as malfunctions in a dedicated *subpersonal* mechanism within a person's mind/brain.
9. Constructing and manipulating representations of syntactic structure requires either *representing* or *embodying* a grammar. Any psychological or computational model that neither represents nor embodies a grammar will be incapable of replicating human performance and effectively coping with the massive ambiguity of linguistic inputs. This is true regardless of what additional machinery the model requires, including statistical information about the frequency of various kinds of input.
10. The notion of embodiment is intermediary between the notions of full-blown representation and mere "conformity to a rule." Embodiment is distinct from both, in ways that are open to empirical test.
11. The hypothesis that grammatical rules or principles are embodied is more parsimonious than the hypothesis that they are represented. Moreover, we have, at present, no principled grounds for asserting that grammars are represented, rather than embodied, in the human mind/brain.
12. We should tentatively conclude that a common claim in generative linguistics—viz., that grammars are *represented* in the minds of competent language users—is either a conflation of the notions of embodiment and representation or simply an attractive but as-yet-ungrounded hypothesis. Although we confidently assert that the syntactic structure of linguistic input is explicitly represented in the course of language comprehension, the claim that *the rules or principles of a grammar* are likewise mentally represented, rather than embodied, must await significantly more fine-grained measuring techniques and sophisticated experimental paradigms in cognitive neuroscience.

Summary of the Arguments

In what follows, I sketch my arguments for several of the conclusions listed above. I begin with philosophical issues—metaphysics, epistemology, methodology—and then switch gears to review the results of psycholinguistic and computational research.

Metaphysics, Epistemology, and Methodology

There are, broadly speaking, three competing frameworks for answering the foundational questions of linguistic theory—cognitivism (e.g., Chomsky 1965, 1995, 2000), platonism (e.g., Katz 2000), and nominalism (e.g., Devitt 2006, 2008).

Platonism is the view that the subject matter of linguistics is an uncountable set of *abstracta*—entities that are located outside of spacetime and enter into no causal interactions. On this view, the purpose of a grammar is to lay bare the essential properties of such entities and the metaphysically necessary relations between them, in roughly the way that mathematicians do with numbers and functions. On this picture, the question of which grammar a speaker cognizes is to be settled by psychologists, using methods that are quite different from the *nonempirical* methods of linguistic inquiry.

The nominalist, too, denies that grammars are psychological hypotheses. But she takes the subject matter of linguistics to consist in concrete physical tokens—inscriptions, acoustic blasts, bodily movements, and the like. Taken together, these entities comprise public systems of communication, governed by social conventions. The purpose of a grammar, on this view, is to explain why some of these entities are, for example, grammatical, co-referential, or contradictory and why some entail, bind, or c-command others.

Cognitivism, by contrast, is the view that linguistics is a branch of psychology—i.e., that grammars are hypotheses about the language faculty, an aspect of the human mind/brain. A true grammar would be psychologically real, in the sense that it would correctly describe the tacit knowledge that every competent speaker has—a system of psychological states that is causally implicated in the use and acquisition of language.

In Chap. 1, I point out that the epistemological side of the platonist position faces a challenge from the Quinean attack on the tenability of the distinction between empirical and nonempirical modes of inquiry. Katz (2000) argues that Quine's epistemology is inconsistent, because it entails that the principles of reasoning are simultaneously revisable and unrevisable. I show that his argument is fallacious. It overlooks a distinction between our principles of reasoning and our theory *of* those principles. Drawing this distinction eliminates the threat of inconsistency.

As regards the ontology of linguistics, Katz (1985) argues that the optimal grammar for natural language generates nondenumerably many sentences and, hence, that linguistics cannot be about any aspect of the natural world. But I maintain that linguists' claims concerning the infinitude of language need not signal an ontological commitment to abstract entities. Rather, they reflect the lawlike, counterfactual-supporting character of linguistic generalizations, as well as a principled idealization away from mortality, memory constraints, and motivational factors, inter alia. As both Chomsky (2001) and Devitt (2006) agree, there is no reason to take as *literally true* the claim that there exists an uncountable infinity of sentences.

All in all, then, platonism provides a radically mistaken view of the methodology and ontology of linguistics. In Chap. 2, I examine the remaining interpretations of linguistic theory: the dominant cognitivist position and the sophisticated nominalist rival. As both Chomsky and Devitt point out, if one begins with the cognitivist conception, then the question of whether a syntactic rule or principle is psychologically real reduces to the question of whether we have grounds to believe the grammar that posits it. If Chomsky's cognitive conception is right, then, given that we typically do have such grounds, the psychological reality issue is already settled. By contrast, if one begins with Devitt's nominalist conception of grammars, then the ascription of psychological reality to one or another syntactic principle requires, in addition, powerful *psychological* assumptions.

Though I am ultimately neutral on which of these views provides a more satisfying conception of linguistic inquiry, I side with Devitt in thinking that the psychological reality debate cannot be settled without direct appeal to the results of psycholinguistic experiments. Thus, in an effort to block the trivialization of the psychological reality issue, I cast doubt in Chaps. 3 and 4 on the two standard arguments for Chomsky's cognitivist conception. If the cognitivist conception is not the only game in town, then the psychological reality claim becomes interesting; far from being trivially true, it stands in need of a sustained defense. I go on to develop such a defense in Chaps. 5–9.

One common motivation for cognitivism is that unlike its rivals, it has the resources to motivate the search for universal linguistic principles—a universal grammar (UG). On the cognitivist conception of syntax, a specification of UG tells us about the innate resources that a child brings to bear in the acquisition process. Given that language acquisition is an independently interesting phenomenon, the more help we get from syntax in theorizing about it, the more credibility accrues to the syntactic proposals. What analogous motivation can the nominalist provide?

To reply, I note in Chap. 3 that evidence for any claim about the structure of UG invariably rests on the putative existence of a linguistic universal. But we have no way of determining what linguistic universals there are except by constructing grammars for a variety of languages and checking whether the constructs employed by the grammar of L are applicable to another language, L*. The methodology of devising and comparing the grammars of various languages does not presuppose or require cognitivism. In particular, the conclusion of the following inference is a non sequitur.

> Since general linguistic theory describes the common resources of the grammars and there must be something common to all humans as acquirers of language, it looks as if general linguistic theory was all along an account of a universal human cognitive feature, that is, UG. (Collins 2008: 86)

It is of course true that "general linguistic theory describes the common resources of grammars," and it is likewise true that "there must be something common to all humans as acquirers of language." But it does *not* follow, and probably is not true, that every linguistic universal that we ever discover—if, indeed, we discover any (Evans and Levinson 2009)—must automatically be seen as encoded in the human genome, represented in the minds of competent speakers, or involved in language acquisition. For any putative linguistic universal, there are numerous possible explanations. Methodologically, we must first be clear about what universals there are—or even whether there are any—and then examine them, one by one, proposing and (dis)confirming competing genetic, environmental, social, and psychological explanations. An a priori commitment to the cognitive conception is out of place in an empirical discipline.

Still, the project of "general linguistic theory" is a worthy one. Syntacticians are right to borrow the resources of a grammar of one language in theorizing about another. And, in accordance with the general principles of theory choice—the desirability of explanatory unification and maximum generality—they are right to prefer grammars that fit well with what is known about other languages. For this reason, the explanatory adequacy of a grammar is best construed as a successful fit with the maximally general, simple, and unified theoretical coverage of *all* human languages. Construed in this way, explanatory adequacy is desirable *whether or not* the cognitivist conception of linguistics is true.

Another common argument in favor of the cognitivist conception is that the notion of a "public language" is irrelevant to scientific inquiry. On behalf of the opposition, I argue in Chap. 4 that the notion of a public language is indispensable in the study of language acquisition. The data and explananda of acquisition theory are routinely couched in terms that make ineliminable reference to public languages. The empirical findings that animate the poverty of the stimulus argument, and inquiry in acquisition theory more generally, typically concern quantitatively described patterns of error. It is very difficult to see how such findings might be formally recast as claims about the relation between the child's usage and some *specific* I-language. Against *whose* I-language would a child's usage be quantitatively compared?

One might reply that the acquisition theorist is comparing the child's grammar to an *idealized* I-language. But it is not at all clear what import the appeal to idealization has in this context. If we press on the notion of an idealized I-language, we find that it amounts to no more than a consistent setting of parameters (assuming a principles-and-parameters grammar). But there are many such settings, most of which fail to match the language of the child's linguistic community. Thus, reference to the grammar of a public language seems unavoidable in singling out the language that the theorist identifies as the child's "target grammar." If there is an idealization in the vicinity, it is one that abstracts away from the variation *within the community*—i.e., the differences between individual speakers—and yields an *idealized speech community* as the theoretical object of interest.

Psycholinguistics and Computational Linguistics

As a first step toward establishing the psychological reality of syntactic principles, I argue that in the course of comprehension, hearers construct *mental phrase markers*—i.e., mental representations of the syntactic structure of incoming linguistic stimuli. In Chap. 5, I give several arguments for this claim:

1. *Neurolinguistics*: EEG studies using the violation paradigm have found that early left anterior negativity is elicited by, and only by, syntactically ill-formed stimuli. Semantic and pragmatic violations elicit a distinct EEG signature (Bornkessel-Schlesewsky and Schlesewsky 2009). And, more recently, MEG data have revealed that the brains of competent speakers respond to sequences of words in ways that track not only their acoustic, syllabic, and prosodic features but also phrase-level groupings (Ding et al. 2016). These and other studies show that the human sentence-processing mechanism constructs distinctly syntactic representations.
2. *Structural priming*: In producing language, people tend to employ the syntactic structures that they recently produced or comprehended. Priming studies that make use of this fact allow researchers to identify some of the representations that people construct when processing language (Bock and Kroch 1989; Bock and Loebell 1990; Pickering and Ferreira 2008). The behavioral data can be used not only to demonstrate the psychological reality of mental phrase markers but also to determine their content to a degree of precision that ERP studies cannot yet achieve.
3. *"Garden-path" processing*: Linguistic input is rife with ambiguity. The language-processing system is remarkably effective in selecting the correct resolution of such ambiguities. When it fails to do so, the anomaly shows up in behavior—e.g., extended fixations on a crucial part of a sentence, in eye-tracking studies (Rayner et al. 1983), or a modulation of reaction times in cross-modal priming studies (Nicol and Swinney 1989). The principles of ambiguity resolution, minimal attachment, late closure, and the minimal chain principle, form the foundation of many psychologically plausible parsing models (Frazier 1979; DeVincenzi 1991). Taken together these principles predict and explain a vast range of behavioral and neurocognitive data. The principles themselves are not represented, nor even embodied, in the mind/brain, but they make ineliminable reference to the construction and manipulation of mental phrase markers. Competing accounts of the same data appeal to statistical information in the input, but the relevant frequencies are, once again, defined over mental representations of phrase structure.

Chapter 6 is devoted to a fourth argument for the psychological reality of mental phrase markers (MPMs). The main thesis is a negative one: no known model of language processing can explain the available data concerning human parsing preferences without positing MPMs. Computational models that eschew mental phrase markers have been developed in the classical AI tradition (e.g., Schank and Birnbaum

1984). More recently, Devitt (2006a) tentatively endorses an account on which comprehension does not involve constructing mental phrase markers but, instead, maps acoustic input directly into thoughts, whose syntactic structure is assumed to be more or less the same as the structure of public language sentences. I show that both models are unworkable in light of the available evidence, and that Devitt's argument for "brute-causal" models rests on an untenable distinction between mental representations and "mere responses" on the part of an organism.

Having established the reality of mental phrase markers, I go on to argue that their construction and manipulation can *only* be accomplished by a mechanism that either represents or embodies a grammar. Before proceeding to a defense of this claim, I clarify the central notions of representation and embodiment in Chap. 7. A system that *embodies* a grammar does not store a set of rules and principles in an explicit data structure (Stabler 1983) and does not "access" or "read" them during real-time operations. Rather, they are "hardwired" into the causal structure of the system, in such a way as to guide the construction of the mental phrase markers. In order to be an instance of embodiment, this hardwiring must, moreover, meet a condition that is stronger than the mere ability to process inputs of a certain type—i.e., stronger than mere "conformity" to a rule. For every rule or principle of the grammar, the hardwired system must have a causal mechanism that mediates the computation of *all* the syntactic representations that are in the domain and range of that rule or principle (Davies 1995). A language-processing system can *conform* to a particular grammar without *embodying* it in this sense. Embodiment is weaker than representation, but stronger than conformity, in ways that are open to empirical test.

These conceptual clarifications pave the way for a detailed account, in Chaps. 8 and 9, of the coevolution of syntactic theory and computational models of language processing. Chapter 8 lays out the details of context-free grammars and the Earley and CYK parsing algorithms (Jurafsky and Martin 2008), as well as probabilistic models of language processing. Chapter 9 examines the psychological plausibility of parsers that employ transformational grammars, as well as augmented transition networks (Woods 1973; Wanner and Maratsos 1978), principle-based parsers (Johnson 1989; Berwick et al. 1991), and minimalist parsers (Weinberg 1999; Harkema 2001; Stabler 1997, 2001). I sketch the grammar underlying each of these models and discuss their commitment to one of a range of positions on the psychological reality issue.

Along the way, I flesh out the idea that parsing is a species of natural deduction (Johnson 1989, 1991; Shieber et al. 1993). Within the parsing-as-deduction framework, a grammar can be treated as a set of declaratively represented axioms, which must be accessed and inserted as a "step" in the deduction. But it can also be seen as a set of embodied inference rules, in accordance with which the deduction proceeds. At present, I argue that there are no decisive reasons for thinking that grammars are declaratively represented as data structures in the mind/brain, rather than embodied as hardwired procedural dispositions.

Chapter 1
The Ontology of Language and the Methodology of Linguistics

Abstract There are three competing frameworks for answering the foundational questions of linguistic theory. Platonism holds that linguistics is about abstract entities, whose essential properties grammarians discover, by using nonempirical reasoning, as in mathematics. Nominalism takes linguistics to be about concrete physical tokens that comprise conventional systems of communication; grammars explain how inscriptions and the like can be, e.g., grammatical, co-referential, or contradictory. Cognitivism takes linguistics to be a branch of psychology, seeing grammars as hypotheses about the tacit knowledge that every competent speaker possesses. I argue that the epistemological side of the platonist position is undermined by W. V. Quine's attack on the notion of nonempirical modes of inquiry. Jerry Katz contends that Quine's epistemology is inconsistent, because it entails that principles of reasoning are simultaneously revisable and unrevisable. I show that Katz' "revisability paradox" overlooks the distinction between our principles of reasoning and our theory of those principles. Drawing this distinction eliminates the threat of inconsistency. Further, I argue that linguists' claims concerning the infinitude of language need not signal an ontological commitment to abstract entities. Rather, they reflect the lawlike, counterfactual-supporting character of linguistic generalizations, as well as a principled idealization away from mortality, memory constraints, and motivational factors.

Keywords Generative linguistics • Innateness • Platonism • Cognitivism • Nominalism • J. J. Katz • W. V. Quine • Noam Chomsky • Quinean epistemology • A priori knowledge • The revisability paradox • The principle of noncontradiction • Abstract entities • Infinitude of language • Lawlike generalizations • Idealization • Explanatory adequacy • Observational adequacy • Descriptive adequacy • The cognitive revolution • Language acquisition • Paul Postal • D.Terence Langendoen • Size laws • Nondenumerable infinity • J. S. Mill • EEG

D. Pereplyotchik, *Psychosyntax*, Philosophical Studies Series 129,
DOI 10.1007/978-3-319-60066-6_1

1.1 Introduction

Our topic in this chapter is the metaphysics of language. The methodological naturalism that guides this inquiry counsels us to give priority to the following question: What conception of language is playing a role in the working linguist's theories? Unfortunately, the question, so formulated, fails of presupposition; nothing answers uniquely to the description "*the* working linguist." Linguistics is comprised by dozens of schools of thought, each propounding competing theories pitched at any of a range of "levels of analysis." A theorist who takes herself to be "doing linguistics," and is so taken by others in her field, may have any of the following goals in mind:

- To lay bare the phonological, morphological, syntactic, semantic, and prosodic properties of the expressions comprising every human language
- To show how such expressions are used to communicate in various contexts
- To characterize the effects of socio-economic status on the acquisition and use of language
- To map out the perceptual and motor functions involved in the comprehension and production of language (spoken, written, signed, etc.)
- To determine the history of both the structure and use of various aspects of language
- To describe the social and cognitive processes whereby human children acquire language
- To detail the neurophysiological structures and mechanisms that underpin or implement language acquisition, comprehension, and production
- To ascertain the genetic apparatus responsible for the construction of the initial neural structures that make acquisition and processing possible
- To offer an account of the origins and evolution of our capacity to acquire and use language
- To describe the relations between the acquisition and use of language, on the one hand, and the acquisition and use of other cognitive, perceptual, and motor capacities, on the other
- To characterize language-specific disorders, and to isolate their genetic, neurological, perceptual, and articulatory underpinnings
- To devise methods for diagnosing and assisting persons who suffer from such disorders
- To program artificial machines in such a manner as to make them suitable for useful or engaging interaction with competent human speakers
- To formulate systematic methods for translating between distinct languages
- To describe and compare the literary contributions of various cultures

It's safe to say that no monolithic conception of language will do justice to all of these diverse goals. To take a prominent example, consider "Language is a social art"—the opening line of W. V. Quine's seminal *Word and Object* (1960). Though valuable in its own right, the conception of language that is embodied in this claim can hardly be expected to highlight the concerns of a syntactician who is bent on refining the latest conception of *Merge* and *Move*, to say nothing of a neurolinguist striving to ascertain the significance of early left anterior negativity.[1] Still, while we cannot reasonably expect a tidy account of what language *is*, careful attention to both the methodology and the results of actual linguistic theorizing can have profound consequences for a wide range of philosophical projects. Or so I hope to show.

In what follows, I restrict my attention to the generative tradition, pioneered in the mid-1950s by Noam Chomsky and others. This is not to suggest that other traditions are not worth exploring, much less that they carry no merit; indeed, it would surprise me if that were the case. Rather, my choice of focus reflects the fact that the generative tradition has much to recommend it as an object of historical and philosophical scrutiny. It is at once a field young enough to admit of open foundational questions and, at the same time, sufficiently developed as to enable one to draw on substantive results. (In philosophy, fixed points are all too rare and always welcome.) Furthermore, generative linguistics is in a prime position to make deep contributions to the broader field of cognitive science—another longstanding interest of naturalistically inclined philosophers.

Linguists in the generative tradition seek a precise characterization of the syntactic and semantic structure of linguistic expressions, stated in a recursive formalism. What is characteristic of the generative enterprise is the confidence that (i) such a formalism can be arrived at by methods that are, in essence, no different from those at play in other empirical sciences, and (ii) that a statement of this formalism is absolutely central to the resolution of outstanding questions in other branches of linguistics.[2] But, while the introductory texts in generative grammar present what initially appears to be a well-articulated and adequately motivated conception of language, a closer look at the history and sociology of the field reveals deep fissures. A useful starting point in examining these issues is an anthology edited by the late Jerry Katz, aptly titled *The Philosophy of Linguistics* (1985).

Katz argues that there have been two major revolutions in recent thinking about language. First, under the influence of Russell, Wittgenstein, Carnap, and Quine, the analysis of linguistic expressions displaced metaphysics and epistemology as the central focus of philosophical theorizing in the first half of the twentieth century. Accordingly, philosophers became interested in linguistics, in an effort to put their theories on what seemed to be a more secure footing—a project that remains alive in much contemporary work. Second, under Chomsky's influence, generative grammar came to prominence as the most fertile approach to linguistic analysis.

---

[1] The second line of *Word and Object* will fare no better with contemporary acquisition theorists in the generative tradition: "In acquiring [language] we have to depend *entirely* on intersubjectively available cues as to what to say and when" (1960: ix, emphasis added).

[2] Needless to say, a commitment to both the existence and innateness of a Universal Grammar is characteristic of many theories in the generative tradition. But it is not, I think, a defining feature of that tradition. There has been no shortage of detractors from the innateness thesis, some of whom are nevertheless plainly generative grammarians.

Chomsky's seminal contributions were threefold. First, he introduced a new kind of syntactic formalism—transformational grammar—that was clearly superior to the extant taxonomic grammars and to Carnap's logical syntax. Second, he put linguistics on a rationalist and mentalist footing, claiming that linguistics studies *partially innate knowledge structures*. This required the introduction of the notion of tacit knowledge, which will serve as the primary focus of the present inquiry. The innateness thesis—which, by contrast, receives relatively little attention in the chapters to come—was advanced in order to account for the speed and facility with which children acquire language, in spite of an alleged poverty in the data to which they are exposed in the early years of development. Accordingly, the innate knowledge was[3] to have three components: (i) a specification or template of all possible human grammars, (ii) a simplicity metric, along which those grammars are ranked with respect to observational data, and (iii) a device for taking in the observational data and ranking the grammars along that metric, eventually ranking highest what is sometimes called the "target grammar."[4]

Chomsky's third lasting contribution was to decisively discredit the methodology favored by nominalists like Bloomfield, Harris, and Quine—the so-called "discovery procedures" that employ substitution criteria in constructing a taxonomy of linguistic forms. On Chomsky's then-radical view, a theoretical notion was to be considered legitimate for use in linguistic inquiry if it proved necessary for the construction of an optimal grammar for a language—where optimality goes well beyond mere observational adequacy. Chomsky argued forcefully that grammars are not to be seen as mere descriptive botanies of linguistic expressions, but are, instead, substantive empirical hypotheses, to be judged by their explanatory success. A grammar was to be evaluated on the basis of the underlying structures it assigns to various linguistic constructions—the requirement of *descriptive* ade-

[3] Chomsky's view of the matter has evolved considerably since the publication of *Aspects of the Theory of Syntax* (1965). The Principles and Parameters model (Chomsky 1981, 1986) provides an importantly different conception of the child's innate endowment.

[4] It is worth pausing over this common piece of terminology and taking stock of the connotations it carries with respect to the debates surrounding what Chomsky has called E-language and I-language. The term 'target grammar' carries the implication that there is a grammar external to the child at the time of acquisition, which the child is struggling to grasp. Chomsky and his followers would surely resist this implication, claiming that the use of the term is appropriate only in the informal presentation of a theory, not in its serious development—a distinction that Chomsky wields with worrying frequency. (See, e.g., his exchange with Rey, in Barber [2000]). Still, whether a linguist uses the term or eschews it, we are owed a formal, "serious" statement of the child's "goal" throughout the acquisition process. This would go a long way toward making sense of what acquisition theorists mean when they claim that a child has made a "mistake" in the course of acquisition, or, equally, that no mistake was made. I am not convinced, at present, that an explicit, formal, and "serious" account of this would make no mention of E-language. See Chap. 3 for further discussion of this issue, as well as of Chomsky's claim that the notion of E-language plays no role in linguistic theory.

quacy—as well as the role that those structures play in a child's acquisition of a specific language—the requirement of *explanatory* adequacy.[5]

Katz goes on to note that what he calls "the second linguistic turn" gave rise to the proliferation of a large variety of formal descriptions of linguistic structure:

- Transformational grammar (*Syntactic Structures*)
- The Standard Theory (*An Integrated Theory of Linguistic Descriptions*; *Aspects)*
- Extended Standard Theory (*Studies on Semantics in Generative Grammar*)
- Revised Extended Standard Theory (*Reflections on Language*)
- Government and Binding Theory (*Lectures on Government and Binding*)
- Lexical-functional grammar (Joan Bresnan, *Lexical-functional grammar*)
- Montague Grammar (Richard Montague, *Formal Philosophy*)
- Arc-pair grammar (Johnson and Postal, *Arc-pair grammar*)
- Abstract Phrase Structure Grammar (Gazdar)[6]

This staggering assortment presents a rather daunting challenge to an aspiring philosopher of linguistics. Katz writes:

> The philosopher who now wishes to make use of linguistics faces a bewildering complexity. With so many theories available in linguistics, philosophers unfamiliar with the issues either have to let their desire to apply linguistics to philosophy go unrealized or risk having happen to their applications of linguistics what happened to Quine's application of substitution criteria. One navigational aid would be a robust philosophy of linguistics. It would provide an ongoing examination of theoretical developments in linguistics, classifying the theories that emerge, highlighting the philosophically important differences between them, and putting them in a form that is more accessible to philosophers generally. Specialists in the philosophy of linguistics would present the philosophically significant issues between alternative theories of linguistic structure in the way that philosophers of logic have presented the philosophically significant issues between alternative logics. (Katz 1985: p. 11)

Noble as this vision was, it's hard to credit the thought that it was implemented to the fullest by contemporary philosophers. True, philosophical discussions of *semantic* and *pragmatic* theories have been lively and fruitful. But the number of philosophers who closely follow ongoing developments in phonology, syntax,

[5] For an impressively lucid discussion of the history sketched above, see Fodor et al. (1974), Chaps. 2 and 3. See also Blumenthal (1970) and Townsend and Bever (2001), Chap. 2. I discuss the conceptual link between explanatory adequacy and language acquisition in Chap. 4.

[6] Though relatively inclusive, this list is still radically incomplete, as Katz is well aware. (Indeed, he notes that he has mentioned "only the most prominent cases.") Langendoen and Postal (1984: p. 243), provide a more comprehensive, though still incomplete, list. I reproduce their list here, omitting the entries that appear also on Katz's list, as well as the dates of the relevant publications: Finite Grammar (Hockett), Finite State Grammar (Reich), Realistic Grammar (Brame), Stratificational Grammar (Lamb; Lockwood), Tagmemics (Longacre), Natural Generative Grammar (Bartsch and Vennemann), Semantically Based Grammar (Chafe), Functional Grammar (Dik), Daughter Dependency Grammar (Hudson; Schachter), Phrasal Core Grammar (Keenan), Corepresentational Grammar (Kac), Relationally Based Grammar (Johnson), Dependency Grammar (Hays), Categorial Grammar (Lambek), Cognitive Grammar (Lakoff and Thompson), Meaning-Text Models (Melcuk), The Abstract System (Harris), Configurational Grammar (Koster), Neostructural Grammar (Langendoen), String Adjunct Grammar (Joshi, Kosaraju, and Yamada), Equatorial Grammar (Sanders), and Systemic Grammar (Hudson).

psycholinguistics, neurolinguistics, acquisition theory, or computational linguistics is depressingly small.[7]

Katz thought that there was an important reason for philosophers to engage in a concerted study of contemporary linguistic theory: it is here, he claimed, that the infusion of carefully defended *metaphysical* positions can have a striking effect on the direction of empirical research. Indeed, he argued that the *foundational* questions in linguistics are ontological: What sorts of things is linguistics *about*? To what ontological categories do they belong? Traditionally, there have been three positions on the matter. Nominalists (e.g., Bloomfield, early Quine, and Goodman) claim that linguistics is about token physical objects or events. Cognitivists or "conceptualists" (e.g., Chomsky, Fodor) argued that it is, rather, about mental states and cognitive processes.[8] The philosophers whom Katz calls "realists" (e.g., Frege, Katz, Postal and Langendoen) take the objects of inquiry to be *abstracta*—entities lacking causal powers and locations in spacetime, and accessible to the mind only via a special faculty of intuition. To avoid the vexed terms 'realism' and 'Platonism', I will call this position *abstractism*.

Each of the three traditional ontological positions admits of substantive dispute among its adherents. The core nominalist claim, for instance, can be cashed out in the fashion of resemblance nominalism, predicate nominalism, and trope theory. Abstractists come in many stripes as well; Katz (2000: ch. 1) lists "classical Platonism," "contemporary Aristotelianism," and "naturalized realism" as the going candidates. The same is true of cognitivist approaches, as we will see in the chapters ahead. Nevertheless, Katz's coarse-grained tripartite distinction allows him to paint a historical picture that is characterized by what he sees as the two major scientific revolutions in linguistics. On this story, the first revolution consisted in Bloomfield's revolt against nineteenth-century "mentalistic" linguistics, which established nominalism as the reigning ontological orthodoxy within the field. The second was Chomsky's cognitive revolution, against Bloomfield and the behaviorists, which swung the pendulum back in the other direction, replacing nominalism with cognitivism, which remains the received view today. Having sketched this history, Katz writes:

> It is not surprising that the ontological issue has had such historical importance when one realizes that all the major questions in the foundations of linguistics depend on how the issue is resolved. What kind of science one takes linguistics to be—whether it is put with empirical sciences like psychology or with non-empirical sciences like mathematics and

---

[7] As a heuristic, compare the number of philosophers who can rehearse the difference between S4 and S5 modal logics at the drop of a hat, and the number of philosophers who can say anything of substance about the difference between the Government and Binding theory of the 1980s and 1990s, and the Minimalist Program that has slowly replaced it. Though the latter contrast is surely more pronounced, it is rarely discussed in the philosophical literature, while the former is common ground among philosophers of language. The present work is intended as a corrective to this unfortunate trend.

[8] Katz wavers between the terms 'conceptualism' and 'mentalism'. I will use the label 'cognitivism', which is sufficiently inclusive and has the added virtue of bypassing distracting issues surrounding technical uses of the term 'concept' in psychology and philosophy of mind.

> logic—depends on how one resolves the ontological issue. Similarly, what the nature of grammatical argumentation is, how we are to understand the claim that a statement about sentences or a language is true, what a grammatical fact is, how such facts are known, and what the essence of a language is, are all questions to which we will give different answers depending on whether we are nominalists, conceptualists, or realists. (Katz 1985: 14–15)

This brief passage sets the agenda for the rest of this chapter. In Sect. 1.2, I take issue with Katz's distinction between empirical and nonempirical (or "intuitional") sciences. Even with respect to the hard cases of mathematics and logic, I follow Quine (1986) in holding that Katz's position rests on an untenable dualism. I show that Katz's master argument against Quine's epistemological holism does not work. I then argue in Sect. 1.3 that his considerations regarding the infinitude of natural language do not militate in favor of abstractism.

1.2 Katz's Argument Against Epistemological Holism

Recall that Katz writes, "What kind of science one takes linguistics to be—whether it is put with empirical sciences like psychology or with non-empirical sciences like mathematics and logic—depends on how one resolves the ontological issue." Katz gives two examples of sciences that are allegedly non-empirical: mathematics and logic. The claim that these disciplines are non-empirical encompasses a number of related epistemological and metaphysical theses. One is that their subject matter consists of so-called "abstract objects"—e.g., numbers, functions, and propositions. The mathematician and the logician are taken to be studying the "essences" of such objects, as well as the "metaphysically necessary" relations that hold between them. Another such claim is that the truths in these fields are discovered or justified in a way that does not rely, in any significant sense, upon perception, observation, or experiment. Rather, they are established by the methods of deductive proof or conceptual analysis. The body of knowledge that they deliver is, thus, *a priori*. And, on a traditional conception of *a priori* knowledge, it follows that the results in these fields—unlike the results in, say, physics—are immune to revision on the basis of empirical and pragmatic pressures. Summing up, we can say that abstractism is the view that some fields provide us with *a priori* knowledge of the metaphysically necessary truths concerning a realm of nonphysical entities that transcend space-time and enter into no causal relations. Although abstractism is most often concerned with logic and mathematics, Katz (2000) holds that it is also the correct view of the ontology and methodology of linguistics.

Though common in the philosophical literature, abstractism has its share of opponents. A particularly influential view that denies the main tenets of abstractism has been championed by W. V. Quine and his followers. Here is a characteristic statement of Quine's position.

> Mathematics and logic are supported by observation only in the indirect way that those aspects of natural science are supported by observation; namely, as participating in an organized whole which, way up at its empirical edges, squares with observation. I am concerned

> to urge the empirical character of logic and mathematics no more than the unempirical character of theoretical physics; it is rather their kinship that I am urging, and a doctrine of gradualism. … A case in point [is] the proposal to change logic to help quantum mechanics. The merits of the proposal may be dubious, but what is relevant just now is that such proposals have been made. Logic is in principle no less open to revision than quantum mechanics or the theory of relativity. The goal is, in each, a world system—in Newton's phrase—that is as smooth and simple as may be and that nicely accommodates observations around the edges. If revisions are seldom proposed that cut so deep as to touch logic, there is a clear enough reason for that: the maxim of minimum mutilation. The maxim suffices to explain the air of necessity that attaches to logical and mathematical truth. … This much can be said for the linguistic theory of logical truth: we learn logic in learning language. But this circumstance does not distinguish logic from vast tracts of common-sense knowledge that would generally be called empirical. There is no clear way of separating our knowledge into one part that consists merely in knowing the language and another part that goes beyond. (Quine 1986: p. 100)

In this passage, Quine challenges the epistemological claims of abstractism, leaving aside its metaphysical claims.[9]

Katz (2000) presents Quine's view with palpable veneration, but goes on to develop what he takes to be a knockdown argument against it.

> Full appreciation of the power of Quine's empiricist account of logic and mathematics must compel admiration from even the staunchest rationalist. It is quite surprising that an account based on so extreme a form of empiricism can steer contemporary empiricism past the Scylla of Millian inductivism and the Charybdis of logical empiricism and come so close to capturing the special certainty of mathematics and logic. Nonetheless, Quine's holistic conception of knowledge does not in the final analysis enable contemporary empiricists to provide a satisfactory account of the special certainty of logical and mathematical truth, because the conception is inconsistent. (Katz 2000: p. 72)

He casts the alleged inconsistency—his sole objection to the Quinean approach in this book—in the form of a "revisability paradox," which runs as follows:

> On Quine's epistemology, noncontradiction, universal revisability, and simplicity are different from other principles in our system of beliefs in this respect: they are constitutive of the epistemology of the system. The epistemology is a belief-revision epistemology and those principles comprise the basic mechanism of belief revision. They thus serve as essential premises in every argument for reevaluating a belief. Every such argument has to assume the principle of noncontradiction as a rationale for departing from an assignment of truth to logically conflicting statements. The principle is required to initiate the process of revising presently accepted statements, otherwise we have to tolerate a radically laissez-faire epistemology on which anything goes. Further, every argument has to assume that the class of revisable statements is the class of all statements of the system, otherwise the epistemology will no longer be the uncompromising empiricism it was intended to be. Finally, every argument has to assume simplicity or something like it to narrow down the class of

[9] Quine's attitude toward abstract entities evolved over the course of his career. Having defended nominalism in his youth (Goodman and Quine 1947), he later went on to reluctantly admit the existence of *some* abstract entities—sets, individuated extensionally—on account of their alleged indispensability in the natural sciences. Orenstein (2002) calls Quine a "reluctant Platonist, admitting only as many abstract objects, such as sets, as are indispensable for the business of science" (p. 86). Whatever the case about that, Quine has always been unambiguous in his rejection of propositions and other abstract entities for the purposes of linguistics.

> potentially revisable statements. … Here is the paradox of revisability. Since the constitutive principles are premises of every argument for belief revision, it is impossible for an argument for belief revision to revise any of them because revising any one of them saws off the limb on which the argument rests. Any argument for changing the truth value of one of the constitutive principles must have a conclusion that contradicts a premise of the argument, and hence must be an unsound argument for revising the constitutive principle. (Katz 2000: p. 73)

In sum, Quine's epistemology is supposed to underwrite the claim that *every* part of our belief system is revisable, but the very principles that are *constitutive* of that epistemology are themselves unrevisable, on pain of internal inconsistency, so Quine's view is self-contradictory. Despite his frank admiration of Quine's epistemology, Katz sees this argument as a fatal blow, sufficient to disqualify Quine's view from further consideration:

> The revisability paradox shows that no form of uncompromising empiricism, at least none that we are presently aware of, can meet the epistemic challenge to antirealism. The paradox also undercuts the Quinean explanation of how truths of mathematics and truths of logic can be taken to be about natural objects in the Quinean (1961c, 44) sense of being part of "a device for working a manageable structure into the flux of experience." (Katz 2000: p. 74)

In contrast to Katz, I believe that Quine's view of the matter is essentially on target, that there is no revisability paradox, and that the illusion of paradox comes from failing to appreciate the status of the principles of noncontradiction, universal revisability, and simplicity. This failure is, in turn, the result of a broader misconception about the nature of epistemology.

Katz's mistake comes in precisely when he claims that the principles of noncontradiction, universal revisability, and simplicity "serve as essential premises in every argument for reevaluating a belief" and that "every such argument has to assume the principle of noncontradiction as a rationale for departing from an assignment of truth to logically conflicting statements." Katz believes that the principle of noncontradiction "is required to initiate the process of revising presently accepted statements." These are all erroneous statements. The abovementioned principles play no such role in Quine's epistemology. They are *not* premises that we ever have occasion to sincerely marshal in open debate, and they do *not* figure explicitly in our reasoning. Rather, they are useful *descriptions of* our reasoning practices. They are *true of* our belief-maintenance and revision processes in roughly the way that the principles of celestial mechanics are *true of* our solar system.[10]

Let us take these principles in turn. The principle of noncontradiction is not "required to initiate the process of revising presently accepted statements." Rather, there is a true psychological generalization, according to which, *as a matter of fact*, when two beliefs come into direct conflict with one another, one of the beliefs is revised. No supervisory faculty is needed for this to occur. That is, reasoners need

[10] I do not mean to suggest that this is *all* there is to say about the principles in question—a strong version of psychologism. The principles may well do more than serve as descriptive tools for the psychologist. They may, for instance, double as inferential norms. My claim is only that they do not play the role in our psychology or epistemology that Katz's argument requires.

not ask themselves, "Can I revise this belief?"—all the while holding the belief in their cognitive grip, as it were—and then answer in the affirmative after having consulted their explicit commitment to the principle of universal revisability. Rather, it's that rational beings are *in fact* prepared—indeed, in a certain sense, *programmed*—to revise any belief, given a suitable range of conflicting evidence. Indeed, their ability to do this is partially constitutive of their rationality.[11] Finally, when revision does take place, reasoners do not begin by restating anew their strong desire to achieve a maximally simple overall worldview—not even tacitly. Rather, the belief revision process *in fact* takes the path of least resistance (according to a metric that no one has yet fully elucidated). All in all, it appears that the old distinction between *fitting* a rule and *being guided* by it—a distinction that we will examine closely in later chapters—has never been more apt.[12]

Why does Katz fall prey to this error? There are two interpretive possibilities. First, he may be in the grip of a model of reasoning according to which there is a "self" or an "overseer" who supervises and directs the belief-revision process, taking epistemic principles as explicit premises in his or her practical decision-making. This quasi-voluntarist model of reason is, I believe, very difficult to sustain. Its main strength lies in the fact that various quirks of introspection make it a tempting option for generation after generation of philosophers.[13] The second possibility is that Katz has failed to appreciate the force of Quine's naturalist approach to epistemology. According to Quine, the epistemologist does not furnish us with "first principles," on which to base our inferences and arguments, nor with normative injunctions to achieve certain epistemic ends. Rather, a naturalized epistemology shores up *descriptive*, *explanatory*, and *predictive* theories of perception and inference.[14] It is

[11] Such talk of rationality also raises the possibility that the principle of noncontradiction is an inferential *norm*—something we should aspire to conform to, teach children to abide by, and hold others accountable for respecting. Like psychological generalizations, such norms need not play any role in actual reasoning.

[12] The fitting/guiding distinction is, I believe, not fine-grained enough to capture differences that will become important later. In Chap. 7, I will draw a tripartite distinction between (1) merely *fitting* a rule, (2) *embodying* a rule, and (3) *representing* a rule—i.e., using it "as a premise" or "as data" in the course of cognitive or computational operations. My claim in the main text is that, *pace* Katz, the principles of universal revision, noncontradiction, and minimal mutilation do *not* belong in class (3). Whether they belong in class (1) or class (2) is, I take it, an open empirical issue. To settle it, we would have to determine whether, in the brain, there is a *common causal mechanism* (Davies 1989, 1995) that is responsible for *all* of the cognitive operations that conform to each individual principle of reasoning. I suspect that there are no such mechanisms, so my money is on class (1).

[13] See Dennett (1991) for an excision of the "central executive" and "central meaner"—vestiges of the Cartesian philosophy that retains its grip on the field, despite facing seemingly insurmountable problems. For a more recent attack on the view that state-consciousness serves an important epistemic function, see Rosenthal (2008, 2012).

[14] Michal Devitt and Hilary Kornblith have argued that empirical discoveries could show us not only that our theory of our own inferences was wrong, but also that the inferences we in fact make are not ones that we *should* be making. Indeed, Devitt points out that results like those of Kahneman (2011) play precisely that role. This sort of view would allow us to see naturalized epistemology as having room for a normative element that is not merely a "chapter of psychology." This issue is orthogonal to the debate between Quine and Katz.

plain that Katz does not subscribe to this approach. But Quine does. And the charges of inconsistency and paradox cannot be made to stick as long as that is the case.

Whatever the source of Katz's error, I think it's safe to say that we've effectively defused his alleged paradox.[15] To elaborate, there is no contradiction whatsoever in a theorist's coming to reject explicit statements of the principles of noncontradiction, simplicity, and universal revisability, in the course of reasoning *in accordance with those very principles*. All this requires is that the theorist be factually mistaken about his or her own psychological processes—a position that many psychologists have had the misfortune of occupying. Nor is there any reason to subscribe to a weaker version of the unrevisability claim, according to which the principles that correctly describe our reasoning today will *always* correctly describe it. For all that we presently know, reasoning practices may very well evolve over time, under any number of conceivable pressures—environmental, social, or biological. Indeed, if the change were to take place quickly enough, a theorist who rejected some explicit principle of reasoning may start out being mistaken about her own psychology but, by a curious kind of fortune, *end up* being correct.

Issues concerning the status of logical and epistemic principles are large and ramify quickly. My intent is not to settle major disputes in epistemology *en passant*. Rather, my goal in this section has been to resist just one influential line of reasoning in favor of an abstractist view of linguistics. I turn now to another motivation for this view, this time pertaining to topics in metaphysics.

1.3 The Infinitude of Language and the Ontology of Linguistics

As noted above, Katz conceives of the history of linguistics as being characterized by two completed revolutions—Bloomfield's temporary victory over the mentalists and, eventually, Chomsky's triumph over the nominalists. He suggests, however, that the evolution of linguistic theory is not yet at an end, for there is one final revolution left to go—a revolution that will overturn the dominant Chomskyan paradigm and herald in the reign of a superior alternative. Wielding the label 'realism' for what I have been calling 'abstractism', he takes some then-current work in linguistics to be moving in precisely this direction.

> The interest of the ontological issue in linguistics is increased by the fact that there is now a number of linguists and philosophers who are trying to formulate a realist view of language and argue that it should replace the traditional conceptualist views that are presently accepted widely. [In a footnote, Katz lists R. Montague, E. Itkonen, J. Ringen, J. Katz, D. T. Langendoen and P. Postal.] By making the realist position in linguistics more than a logical possibility, these linguists and philosophers have raised the question of whether linguistics might take the next step from conceptualism to realism.

---

[15] If the argument of this section hits the mark, Katz's case (2000: pp. 17–19) against the position taken by Penelope Maddy in her book *Realism in Mathematics* (OUP 1990) is correspondingly weakened.

> As we have seen in the historical development from Bloomfield to Chomsky, there has been a correlation between the ontological position that linguists accept and the methodology and kind of grammar they propose. *A different ontology might thus open up a new approach to the study of language. The formulation of a realist position for linguistics give linguists some reason to think that new insights about grammatical structure may be forthcoming*. (Katz 1985: p. 15, emphasis added)

Katz cites Langendoen and Postal's book, *The Vastness of Natural Languages* (1984), in which the authors argue for the absence of a size constraint on natural languages and, on this basis, for the impossibility of describing language with certain types of grammars.

Before turning to Langendoen and Postal's arguments, let me first take issue with Katz's vision of the historical and theoretical dialectic in linguistics. As is evident from the italicized passage in the quotation above, Katz believes that paying particular attention to ontological issues may well give rise to his envisioned revolution in both philosophy and linguistics, provided of course that the abstractist option prevails. But there are at least two problems with the suggestion that a thesis about the metaphysics of language can yield "new insights about grammatical structure."

First, Katz's suggestion implies that ontological issues are clear enough for us to be confident in challenging hard-won syntactic analyses on the basis of antecedently established ontological claims. That's a bizarre commitment to hold in the face of the lack of convergence on the part of metaphysicians regarding ontological issues. As compared with the healthy progress and steady evolution that we find in syntactic theorizing, the debates in ontology have made little progress. I am not aware of a single *result* in metaphysics that is comparable in explanatory power to, say, the shift from pre-GB-era syntax to the Principles and Parameters theory (Chomsky 1986) and the associated improvements in computational parsing models (Berwick and Fong 1995).

Second, Katz's claim seems to be out of sync with the way in which the linguists whom he cites actually proceed. Postal and Langendoen, for instance, first launch *empirical* arguments for their grammars, and only *then* draw ontological conclusions. Indeed, Katz's claim is out of sync with how he characterizes *his own* thinking in a later publication:

> [I]f the concerns of a partisan viewpoint in the philosophy of mathematics are allowed to decide questions of grammatical structure, distinctions reflecting no grammatical differences will be made over a wide segment of the language. Since such distinctions are only philosophically motivated, a multiform semantics would compromise the autonomy of linguistics. Linguistic argumentation would degenerate into philosophical debate—Why should the linguist let the concerns of antirealism decide? Why not the concerns of realism? And so on. To preserve the autonomy of linguistics and the integrity of its argumentation, only linguistic considerations can be allowed to determine the description of sentences. (Katz 2000: p. 31)

In this passage, Katz is insisting that empirical issues having to do with the description of sentences are to be settled *prior* to the defense of ontological theses.

Putting aside the apparent exegetical inconsistencies in Katz's *oeuvre*, as well as the implausibility of his prioritization of ontological issues in linguistics, let us focus briefly on whether Postal and Langendoen's argument for an abstractist grammar in fact succeeds. I believe it does not.

A natural place to begin the discussion is with two background assumptions that are commonplace in generative grammar. The first is that any sentence of any natural language is finite in length, in the sense that it is constituted by finitely many phonemes or morphemes (themselves drawn from a finite stock). The second is that there is no sentence, S, such that operating on S with a recursive rule of the grammar would not generate a longer sentence, S*. Put more simply, the claim is that there is no longest sentence. These two assumptions—i.e., that all sentences are finite in length, and that they are nevertheless unbounded—can be used to prove that the set of sentences in any natural language is *countably*, or, in a different terminology, *denumerably* infinite; the cardinality of this set is $\aleph_0$.

These assumptions are what Postal and Langendoen refer to as "size laws." Sufficiently formalized grammars of a natural language will typically include one or another such law. What Postal and Langendoen pointed out was that the size laws stated in the previous paragraph, standard in generative grammar, are not the only laws we might adopt. We can imagine, for instance, replacing the law that the length of natural language sentences is unbounded with a law that takes some fixed, finite number, N, to be an upper bound on the number of phonemes or morphemes in a sentence. This, of course, is *ad hoc* and unmotivated. The reasoning behind the unboundedness assumption is similar to what we encounter in the case of natural numbers. There is no largest positive integer because, for any integer we pick, we can always add 1 to it, thereby yielding a larger integer. Similarly, there is no longest sentence, because you can always add 'Wilfrid knows that' to any (declarative) sentence, thereby generating a longer sentence.

Postal and Langendoen's argument, however, is not aimed at replacing the standard size laws with more *restrictive* ones, like the law just described. Rather, they seek to *abolish* the size laws altogether. On their view, parsimony dictates that linguists should adopt a grammar that places *no limit at all* on the length of sentences, thereby allowing the grammar to output sentences of literally *infinite* length. They go on to prove that such grammars, when encompassing plausible ancillary principles, yield the consequence that there are *transfinitely* many sentences—i.e., that the set of sentences is uncountable or nondenumerable, its cardinality exceeding $\aleph_0$.

Before evaluating Postal and Langendoen's claims, it is worth pointing out that, contrary to the conclusions that the authors draw in the book—the very conclusions that Katz seeks to promote in the passages quoted above—the claim that there are transfinitely many sentences in fact has no significant consequences whatsoever for the ontological dispute between abstractists and their opponents. In subsequent remarks, Langendoen has written:

> Paul and I spilled a lot of ink trying to clarify Chomsky's "conceptualist" view of natural languages, and having clarified it, to show that it is incorrect. But it now strikes me as quite easy to work out a coherent conceptualist theory of natural languages that does not incorporate a size law. In fact, any principles and parameters theory that does not do so, but that does include what strikes me as the patently correct principle of coordinate compounding, will yield the result that natural languages (albeit, E-languages, in Chomsky's terminology) are of transfinite size. (Langendoen, T., "Is language infinite? Replies to Manaster-Ramer," in *Linguist List* 2.602, Sept. 30th, 1991.)

Notably, although Langendoen has changed his mind, Postal continues to attack the dominant Chomskyan paradigm. In Postal (2009) he writes:

> The incoherence of Chomsky's position is worth putting slightly differently. To say a collection is discretely infinite is to say its members can be put in one to one correspondence with the members of one of its own sub-collections and also with the positive integers. What Chomsky's ontology then asserts is that the human brain component he calls FL [i.e., faculty of language] or a particular state of it defining a particular NL [i.e., natural language] embodies as aspects a collection of things which can be put in such a correspondence with the integers 1, 2, 3,…. But there can't even begin to be enough of *anything* in a human brain or its functioning to ground such a correspondence. It is no accident then that across the decades of adumbrating his biolinguistic view, Chomsky has never even sought to specify the nature of supposedly biological entities which manage to have a transfinite cardinality. (p. 110)

Postal goes on to quote a passage from Langendoen and Postal (1984).

> Even for an attested NL like English, the claim that a grammar, even a psychogrammar, generates mental representations immediately creates otherwise unnecessary fundamental problems. Evidently, either standard or radical conceptualist must minimally assume that any actual human mind or brain is finite, and thus that its very nature limits the objects which are in fact representable therein, for trivial non-linguistic reasons. Consequently, if psychogrammars generate mental representations of sentences and mental representations are, as the term suggests, things actually present in real minds, in something like the sense in which, say, data or computations are present in real computers, the radical conceptualist position claims that NL grammars have a finite output, one containing no representation of cardinality greater than some finite *k*. This is inconsistent not only with the traditional generative position but with claims in the very works where radical conceptualism is advocated that the domain of grammar is infinite. (pp. 131–132; quoted in Postal 2009: p. 110)

What both of these passages seem to miss is that the phenomena that the linguist seeks to describe, predict, and explain include not only *past* cognitive and verbal episodes—acts of production, comprehension, and perhaps metalinguistic judgment—but also *future* ones. And while the number of these—both past and future—is finite, there is no way of telling in advance *which* particular such episodes we will come to encounter. What we do know from our observations thus far, however, is that such episodes have been *continually novel*. And we have no reason to expect that they will cease to exhibit *that* pattern.

Now, in the long run, advances in the biochemical and bioinformatic sciences may increase the human life span, alter our motivational states, and enlarge our memory resources. As a result, the sentences that we produce and understand may become not only continually novel (as they have been to date), but also significantly longer. And just as there is no telling in advance what the limits are on the linguistic creativity that gives rise, continually, to novel sentences, there is likewise no telling in advance just how long the longest sentence a human will ever utter might be.[16]

[16] My claim, then, is that the following remarks from Rohde (2002) are based on a failure of imagination: "Language is *pseudo-context-free* or *pseudo-context-sensitive*, but infinite recursion is an idealization ungrounded in any observable reality. Some like to say that there are an infinite number of possible sentences in a language, meaning that there are an infinite number of utterances that conform to the rules of the grammars that linguists design. But if we limit the definition of a sen-

This, I submit, is the underlying justification for adopting the standard size law of sentence unboundedness.

Note, however, that there is *no* such justification for abolishing the size laws altogether. For, while we do not know how long the longest sentence ever might be, we *do* know this: No such sentence will ever be infinite. Whatever advances are made in life-extending, memory-enhancing, and motivation-boosting technologies, it remains the case that a human being will never utter an infinitely long sentence. This we can be sure of. Therefore, there is no empirical pressure for us to posit such sentences, not even as mere potentialities.[17]

Postal and Langendoen share with Katz the view that the infinitude of language is a *datum*—one that can only be captured by the abstractist conception. (After all, in contrast with *real* objects, there's always room for more *abstracta*!) The reply that I favor is nicely articulated by Devitt (2006a).

> Katz has another objection to nominalism: grammars are about an infinite number of sentences but there cannot be an infinite number of tokens. If there were a problem for my sort of nominalism it would lie in its apparent commitment to nonactual possible sentences, a problem that would arise even if we were dealing with a finite language (e.g. English with a limit of one million words to a sentence). The only significance of any apparent commitment to an infinite number of sentences is that it would guarantee that some were nonactual. But talk of there being nonactual possible outputs of a competence can be a mere manner of speaking… So too can talk of there being an infinite number of such outputs. The truth behind the talk of the nonactual can be simply that the grammar is lawlike. [Footnote: So if something were a sentence, a *wh*-question, a passive, or whatever, it would have the properties specified for such items by the theory.] And the truth behind the talk of the infinite can be simply that there is no limit to the number of different sentence tokens that might be governed by the rules the grammar describes. (p. 26)

The idea here is that the infinitude of language is *not* a datum. Indeed, it's not even *true* that there are, in existence, infinitely many linguistic tokens or types (nor that

---

tence, as I choose to do, to those utterances that could possibly convey a useful and consistent meaning to at least a few fellow speakers of a language, even under optimal conditions, the claim of an infinite variety of sentences is simply not true. There is a finite bound to the length and complexity of sentences that humans can comprehend, and thus a finite bound to the possible sentences we can create, short of inventing new words. If I were to link all of the sentences in this thesis together with *and* or some other suitable conjunction, the result would not be another sentence. It would be an abomination. I'm not arguing that there aren't a vast number of possible sentences in English, just that there aren't an infinite number" (pp. 3–4).

[17] The position developed here is shared by Collins (2008a), who writes: "Note, the datum here is not that we understand an *infinity* of sentences. Given the vagaries of the notion of 'understanding' and metaphysical suspicions over the infinite, such a claim would needlessly attract irrelevant objections. The datum, rather, is that our competent use of language is 'indefinitely new' or 'continuously novel'; far from finding ourselves stumped at novel sentences, the vast majority of sentences we produce and consume are entirely novel to us… If our competence were somehow finitely bounded, then our apparent ability to project 'acceptable' or 'grammatical' across an indefinite range would be a kind of fortunate accident, as if there were a finite upper bound, but one sufficiently great to be unnoticed. Such a speculation is irrelevant and, anyhow, false" (p. 32).

there have ever been, nor that there will ever be).[18] Rather, the infinitude is an artifact of theory construction—specifically, a reflection of the fact that we seek maximally wide generalizations in our effort to accommodate *any* potential datum that we might encounter. Chomsky (2001) puts the point as follows:

> We also speak freely of derivations, or expression Exp generated by L, and the set of such Exps—the set that is called “the structure of L” in Chomsky (1986), where the I/E terminology is introduced. Evidently, these entities are not “internal.” That has led to the belief that some externalist concepts of “E-linguistics” are being introduced. But that is a misconception. These are not entities with some ontological status; they are introduced to simplify talk about properties of FL and L, and they can be eliminated in favor or internalist notions. One of the properties of Peano’s axioms PA is that PA generates the proof P of “2 + 2 = 4” but not the proof P’ of “2 + 2 = 7” (in suitable notation). We can speak freely of the property “generable by PA,” holding of P but not P’, and derivatively of lines of generable proofs (theorems) and the set of theorems, without postulating any entities beyond PA and its properties. Similarly, we may speak of the property “generable by L,” which holds of certain derivations D and not others, and holding derivatively of an expression Exp formed by D and the set {Exp} of those expressions. No new entities are postulated in these usages beyond FL, its states L, and their properties. Similarly, a study of the solar system could introduce the notion HT = {possible trajectories of Halley’s comet within the solar system}, and studies of motor organization or visual perception could introduce notions {plans for moving the arm} or {visual images for cats (vs. bees)}. But these studies do not postulate weird entities apart from planets, comets, neurons, cats, and the like. No “Platonism” is introduced, and no “E-linguistic” notions: only biological entities and their properties. (“Derivation by Phase,” pp. 41–42.)

As Chomsky makes clear, the point holds not only in the realm of linguistics—conceived as a mode of empirical inquiry—but also for the case of mathematics, geometry, and all of the so-called “formal sciences.” Indeed, the point holds with respect to *all* serious systematic inquiry, on account of its being a virtue of *any* descriptive-explanatory theory that its generalizations be maximally broad.

John Stuart Mill famously held that our confidence in the claims of arithmetic is based on induction from past instances (Mill 1874). Although Mill did not put the view as carefully as he might have—appealing to induction where abduction would have served better—I think he was quite correct to insist that the epistemology of arithmetic is no different, at bottom, from that of any other area of empirical inquiry. One begins with claims that facilitate one’s encounters with the world—claims about the quantities of, e.g., apples, miles, dollars, and hours—and later formalizes one’s theory, when necessary, by selecting as “basic” those statements that entail the very claims that one has found serviceable. Crucially, the formalization inevitably involves an element of idealization. One’s navigation of the natural and social environment—e.g., the building of bridges and stock portfolios—rarely requires thinking about *very* large, let alone *infinite*, quantities. But, in the course of formalizing one’s commonsense theory of quantity, there appears to be no principled reason to set an upper bound on the number of successors a given number might have. This is,

[18] Collins (2009) puts the point this way: “Controversy on this issue appears to arise from the common talk in linguistics and philosophy of natural languages being infinite. The infinity of English, as it might be, however, is not a phenomenon. The phenomenon is that particular organisms are continuously novel in their speech and understanding, which we theorize in terms of unbounded generation” (fn. 4).

however, nothing more than a convenience, of the same kind as we find in other cases of scientific idealization. It does *not* signal that the subject matter is now a set of abstract objects. This is what Chomsky (2001) means when he writes:

> One of the properties of Peano's axioms PA is that PA generates the proof P of "2 + 2 = 4" but not the proof P' of "2 + 2 = 7" (in suitable notation). We can speak freely of the property "generable by PA," holding of P but not P', and derivatively of lines of generable proofs (theorems) and the set of theorems, without postulating any entities beyond PA and its properties. (p. 42)

Making a similar point about idealization and "indefinitely large" quantities in linguistics, Devitt (2006a) says:

> [Talk of indefinitely large quantities] may appear to commit theories of outputs to the existence of unactualized possibilia, but it can be, and in my view should be, a mere manner of speaking. It is a convenient way of capturing that these theories, like all interesting ones, are lawlike. Strictly speaking, the theories quantify only over actual entities but the theories are, in some sense, necessary. So the talk captures the modal fact that if something *were* a horseshoe, a chess move, a *wff*, a bee's dance, or whatever, then it *would have* the properties specified by the appropriate theory of outputs. (How are we to explain modal facts? I don't know but, *pace* David Lewis, surely *not* in terms of unactualized possibilia.) (Devitt 2006a: p. 21, fn. 6)

The larger moral to draw from all of this, I think, is that debates about the cardinality of the set of natural language sentences are not debates about ontology, for they have no serious consequences in that direction. Rather, they concern the nature of the idealizations that linguists should make in pursuing their explanatory goals. Given that we have no idea how long a sentence can get—nor any grounds for advancing a principled estimate—and given that we have good reason to believe that the length of sentences is a function of life span, motivation, memory and the like, the natural idealization to make in constructing a grammar is that sentence length is unbounded but nevertheless finite. Postal and Langendoen's argument to the contrary misses the mark.

The best diagnosis of their failure is, I believe, that they misconstrue the aims and aspirations of linguistic theory. The construction of a grammar is not a formal game. In one way or another, it will have to enter into a theory of natural phenomena such as speech and comprehension. Hence, whatever one's prior metaphysical commitments might be, one's grammar will have to one day be integrated with a psychological theory of acquisition and use. This is not to say that it's a foregone conclusion that grammars are, in some sense, "in the mind." That thesis must be both clarified and defended (Chaps. 5–9). Rather, it's to say that whatever entities a grammar describes are going to have to enter *somehow* into an explanation of how flesh-and-blood creatures manage to acquire, and to put to use, their ability to produce and comprehend language.

The exact mode of integration of linguistics and psychology has, of course, not yet been specified. However, we can rule out in advance some extremely implausible suggestions that have been endorsed in the literature. For instance, Postal (2009) says the following: "Everett (2005) claimed *rightly* that if natural languages were biological, syntactic trees should be visible in CAT scans" (fn. 12, p. 111, my emphasis). It's clear, I think, that Postal and Everett's claim is either factually mistaken or simply confused. In later chapters, we will survey the extensive evidence for thinking that the human sentence processing mechanism builds mental representations of the syntactic structure of incoming linguistic stimuli. Some of this evidence draws on recent EEG data—recordings of real-time neuronal activity during comprehension tasks. Given that mental representations are ultimately patterns of neuronal activity, my argument will commit me to the claim that syntactic representations do, in a sense, "show up" in devices that are functionally similar to CAT scanners. However, it should go without saying that such representations do not *look* (visually) at all like hierarchical tree-structures, and cannot simply be "read" by mere casual observation, even by a trained linguist. The interpretation of neural activity—i.e., the assignment of representational content to various brain processes—requires painstaking functional analysis. We can conclude that Postal's claim is either false—because syntactic properties and relations *are*, in fact, neurally encoded—or else it rests on a naïve assumption to the effect that the brain uses *the very same notational format* that one finds in contemporary linguistics texts.

1.4 Conclusion

Considered as a package of views concerning the ontology and methodology of linguistics, abstractism offers us a mistaken picture of the field, of its aspirations, and of language itself. In Sect. 1.2, I sketched an alternative position, composed of the naturalism and epistemological holism of W. V. Quine. Given that Quine avoids commitment to an extravagant ontology of abstract entities and doesn't posit a suspect faculty of "rational intuition," I take his to be the default position—the one we should adopt, pending a powerful argument to the contrary. I went on to examine Katz's argument, according to which the Quinean view suffers from a fatal internal inconsistency—what Katz (2000) calls "the revision paradox". I pointed out that Katz fails to distinguish between our descriptive theory of our own psychological practices, and our actual psychological practices. When the distinction is properly drawn, no paradox arises. The Quinean position remains unscathed.

In Sect. 1.3, I argued that abstractism derives no support from the infinitude of natural language. Indeed, following Chomsky (2001), Devitt (2006a), and Collins (2009), I urged that this infinitude be seen not as a datum, but rather an artifact of theory construction. As such, talk of there being infinitely many sentences in natural language can and should be viewed as a commitment to the modal force of syntactic theory—i.e., its counterfactual-supporting character, or "lawlikeness," in the sense of Goodman (1954). The upshot is that neither the nominalist position nor Chomsky's "biolinguistic ontology" face a threat. Certainly, Postal (2009) is mistaken in his insistence that the latter is "incoherent."

Having argued against abstractism, I propose to examine the two alternative views of linguistic ontology—the sophisticated nominalist position developed by Devitt (2006a) and its dominant rival, the cognitivism of Chomsky (2000). These are the topics of coming chapters.

Chapter 2
Cognitivism and Nominalism in the Philosophy of Linguistics

Abstract Noam Chomsky equates an individual's idiolect with a hypothesized psychological structure, an "I-language", which the competent speaker/hearer "tacitly knows" or "cognizes" via "mental representations" of syntactic principles. Just what do these claims amount to? And what grounds are there for believing them? I attempt to pin down Chomsky's evolving commitments regarding the relation between an I-language and the performance systems that are involved in comprehension and speech. This in turn raises issues about the empirical methodology and naturalistic credentials of Chomsky's "cognitivist" conception of language. I then discuss Michael Devitt's "Linguistic Conception," a contemporary descendent of nominalism, according to which language consists of concrete entities—inscriptions, acoustic blasts, and the like. Devitt argues that linguists describe certain properties of these entities—e.g., grammaticality and entailment—that allow them to play a role in communication. Both approaches face non-trivial difficulties. Chomsky's view awaits an articulation of the relation between grammars and neurophysiological mechanisms. Devitt's requires a more detailed account of public language conventions and a satisfactory strategy for paraphrasing away apparent references to unvoiced and abstract elements of syntactic structure—empty categories. Finally, I argue that Devitt is right to distinguish a competence from its outputs, and hence linguistics from psycholinguistics.

Keywords cognitivism, nominalism, E-language vs. I-language, idiolect, structural description, tacit knowledge, competence vs. performance, parsing, internalism (see also: individualism), Michael Devitt, conventions, empty categories (see also: unvoiced elements, phonologically null constituents), outputs of competence, the Linguistic Conception, supervenience, discovery procedures, Representational Thesis (RT)

D. Pereplyotchik, *Psychosyntax*, Philosophical Studies Series 129,
DOI 10.1007/978-3-319-60066-6_2

2.1 Introduction

In this chapter, I sketch the main features of Noam Chomsky's influential perspective on language and linguistic theory and contrast it with the rival account developed by Michael Devitt. Chomsky equates an individual's idiolect with what he calls an "I-language"—a psychological structure that he takes to be the primary subject matter of linguistic theorizing. He characterizes the competent speaker/hearer as "cognizing" her I-language and as having "tacit knowledge" in the form of a "mental representation" of various syntactic rules or principles. Just what do these claims amount to, and what grounds are there for believing them? In Sect. 2.2, I attempt to pin down Chomsky's commitments regarding the relation between an I-language and the performance systems that are involved in the comprehension and production of language. This gives rise to a discussion of the competence-performance distinction, which in turn raises issues about the empirical methodology and naturalistic credentials of Chomsky's "cognitivist" conception of language.

Although Chomsky's philosophical views form the foundation of generative linguistics, many theorists have raised substantive questions about the relation that he attempts to forge between grammar and the mind. Such questions have become urgent in part because an alternative position in linguistic metatheory has recently been formulated and defended, viz., Michael Devitt's "Linguistic Conception"—a contemporary descendent of nominalism that stands in sharp contrast to Chomsky's cognitivism.

In Sect. 2.3, I discuss Devitt's view, according to which language consists of token inscriptions, acoustic blasts, and other spatiotemporally localizable entities.[1] Devitt argues that the task of linguistic theory consists in accurately describing certain properties of these entities—e.g., grammaticality, constituency, and entailment—which allow such entities to play the role that they do in linguistic communities.[2] In defending the coherence of this view, it is crucial to say something about the hard cases—viz. the empty categories *wh*-trace, NP-trace, PRO, *pro*, ∅, *Op*, and others. I spell out Devitt's notion of a higher-order relational property and discuss the import of his appeal to linguistic conventions, which I take to be at the core of his account of the metaphysics of empty categories. Finally, I follow Devitt in distinguishing sharply between a competence and its outputs, which in turn underwrites a robust and theoretically significant distinction between linguistics and psycholinguistics.

---

[1] In using the catch-all term 'entities', I slur over the differences between objects, events, and processes. Such distinctions will be explicitly noted only when relevant.

[2] Talk of properties does not fit well with Devitt's nominalist leanings. He takes it to be a mere *façon de parler*—harmless shorthand that can be regimented away when the chips are down. In Sect. 2.3.1, I point to a case where it is not clear what a suitable regimentation might be.

2.2 Chomsky, I-language, and the Cognitivist Conception of Linguistics

Over the course of more than five decades, in a series of groundbreaking works, Noam Chomsky has developed and continually refined a profound and highly influential conception of language. Central to his view is the idea that human beings are innately endowed with a dedicated neural structure for acquiring and using language. This "mental organ," which Chomsky calls the Language Faculty (FL), houses representations of various rules and principles for assigning internal structure to linguistic expressions.[3] On this view, the object of linguistic inquiry is, in the first instance, a *psychological* entity—i.e., a person's "linguistic competence." Chomsky identifies this with what he calls an "I-language"—a body of domain-specific *knowledge* or *information* that every competent speaker has at his or her disposal.[4]

Although Chomsky's specific hypotheses regarding what is represented within FL have evolved significantly over the course of his career—constituting a breathtakingly rich body of technical work—his basic ideological assumptions have remained largely intact. The following passage, taken from Chomsky (1995), serves as a paradigmatic statement of his view, which overlaps substantially with the position espoused in his early publications (e.g., Chomsky 1965, 1975a, b, c, 1980a, b, c, 1986).

> We are concerned, then, with the language faculty, which we understand to be some array of cognitive traits and capacities, a particular component of the human mind/brain. … We call the theory of the state attained [by the language faculty] its *grammar* and the theory of the initial state [of the language faculty] *Universal Grammar* (UG). … We distinguish between Jones's *competence* (knowledge and understanding) and his *performance* (what he does with that knowledge and understanding). … We may think of the language, then, as a finitely specified generative procedure (function) that enumerates an infinite set of structural descriptions (SDs). Each SD, in turn, specifies the full array of phonetic, semantic, and syntactic properties of a particular linguistic expression. (*The Minimalist Program*, pp. 14–15)

To fix ideas about what structural descriptions are, Fig. 2.1 below provides an example.

A notable aspect of the passage quoted above is Chomsky's stress on the claim that competence in a language involves "knowledge and understanding"—states

---

[3] Following Chomsky, I use the abbreviation 'FL' to avoid introducing yet further ambiguity into the abbreviation 'LF', which is commonly used to denote a level of formal representation in the Principles-and-Parameters (P&P) approach to syntax. The term 'LF' is already bedeviled by difficulties, which arise from the differences between its role in P&P theories and the role of 'Logical Form' in traditional logic and philosophy of language.

[4] As is customary, I use 'speech' as a catch-all term for vocalizations, inscriptions, hand signs, etc., and I use 'speaker' to refer to both the producers and the consumers of such signals. Where it matters to the discussion, I distinguish between spoken, written, and signed linguistic signals, as well as between speakers, writers, signers, and their audiences.

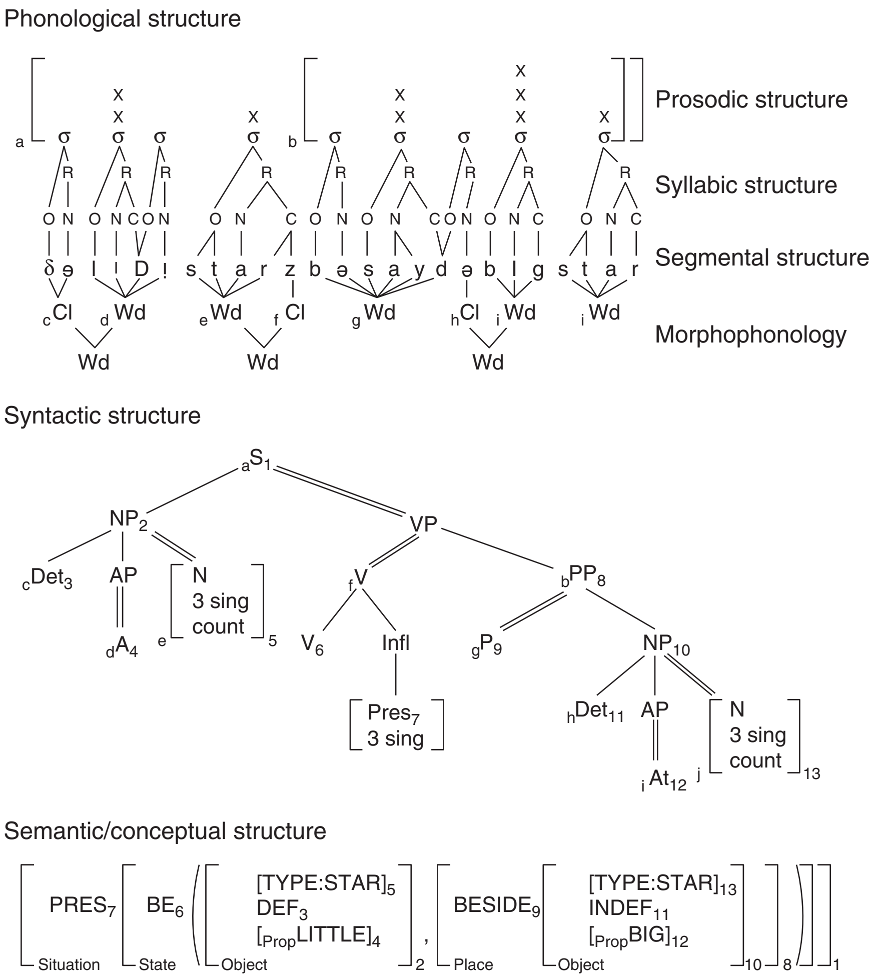


Fig. 2.1 A partial structural description (SD) of the sentence 'The little star's beside a big star'. (*Source*: Jackendoff, 2002. Chomsky (1995) characterizes SDs as "symbolic objects … [each of which] specifies the full array of phonetic, semantic, and syntactic properties of a particular linguistic expression" (pp. 14–15). The details of the analysis represented here are contentious and do not necessarily reflect Chomsky's precise theoretical commitments)

that are widely believed to be *representational* in character.[5] As noted above, the knowledge that the representations in the language faculty jointly constitute is what Chomsky refers to as the speaker's "I-language." In Chomsky's usage, an I-language is an explicitly psychologized grammar of a particular speaker's idiolect.

> To distinguish this concept of language from others, let us refer to it as an *I-language*, where *I* is to suggest 'internal,' 'individual,' and 'intensional'. The concept of language is internal, in that it deals with an inner state of Jones's mind/brain, independent of other elements of the world. It is individual in the sense that it deals with Jones, and with language communities only derivatively, as groups of people with similar I-languages. It is intensional in the technical sense that the I-language is a function specified in intension, not extension: its extension is the set of SDs (what we might call the *structure* of the I-language). (*The Minimalist Program,* p. 15).

Importantly, I-languages must be distinguished from E-languages, which are conceived of as systems of signs common to a community of speakers. The 'E' in 'E-language' is meant to suggest 'external to the mind/brain', which is what such signs would have to be if they were the physical tokens of the nominalist or the abstract types of the abstractist. By contrast, Chomsky's view is that linguistic entities are *in the mind*, rather than out in the world. He suggests, furthermore, that constructing a notion of E-language inevitably requires drawing arbitrary boundaries between linguistic communities, or making reference to socio-political constructs that are strictly irrelevant to a scientific study of language. The 'E' in E-language is also intended to suggest 'extensional'. Chomsky (1986) rejects the view, which he associates with Quine and Lewis, that grammars are functions-in-extension from sounds to meanings. All in all, his view is that E-language plays no serious role in linguistic theorizing.

> [E-language] has no known status in the study of language. One might define E-language in one or another way, but it does not seem to matter how this is done; there is no known gap in linguistic theory, no explanatory function, that would be filled were such a concept presented" (*The Minimalist Program,* p. 16).

On the basis of this cognitive conception of language, Chomsky claims that there is no difference between a grammar's being "psychologically real" and its being simply *true*. For, if a grammar is about a psychological entity—viz., I-language—then the distinction between the truth of a grammar and its psychological reality

---

[5] The term 'knowledge' is a notoriously slippery one. Chomsky does not insist on it, and it is debatable whether he intends the term to denote full-blooded *propositional attitudes*. Many philosophers, including Fodor (1983, pp. 4–5) have taken him to hold this commitment and Devitt (2006a) says that it is "natural" to do so (though his official view is noncommittal about this). Collins (2008a, ch. 5) argues against this interpretation. In an attempt to shed the unwelcome connotations of the term 'know', Chomsky (1980a, pp. 69–70) introduced the technical neologism 'cognize'. Determining the import of this notion is a crucial part of getting clear on Chomsky's view of the relation between a competent speaker and a grammar (i.e., "the generative procedure"). It is also important for determining whether his view entails that the states of the language faculty constitute what philosophers have called "knowledge-that," as against "knowledge-how." Chomsky (2003) provides an attempt at clarification. In Chap. 7, I highlight the significant consequences, for this issue and many others, of seeing the representational states of the human sentence processing mechanism (HSPM) as subpersonal states, rather than personal-level propositional attitudes.

collapses. This line of reasoning has been a recurring theme in Chomsky's work, as the passages below illustrate.

> We attribute psychological reality to the postulated representations. … In short, we propose … that our theory is true … and give substantial independent evidence for the theoretical constructions, showing that the postulated principles explain many other facts of a similar nature, withstand empirical tests in English and other languages…" (Chomsky 1978: pp. 206–208).

> [W]hat is 'psychological reality' as distinct from 'truth, in a certain domain'? … I am not convinced that there is any such distinction. […] the truth of the theories we are constructing; or if one prefers … their 'psychological reality', though this term is best abandoned, as seriously misleading" (Chomsky 1980a: pp. 46–48)

> We observe what people say and do, how they react and respond, often in situations contrived so that this behavior will provide some evidence (we hope) concerning the operative mechanisms. We then try, as best we can, to devise a theory of some depth and significance with regard to these mechanisms, testing our theory by its success in providing explanations for selected phenomena. Challenged to show that the constructions postulated in that theory have "psychological reality," we can do no more than repeat the evidence and the proposed explanations that involve these constructions. Or…we can search for more conclusive evidence…" (Chomsky 1980a: p. 191)

As we will see, Devitt (2006a) forcefully opposes this identification of a grammar's truth with its psychological reality.

According to Chomsky, the production and comprehension of speech requires not only the possession of an I-language, but also its participation in real-time, online computational processes. In the course of such processes, the representations that constitute a speaker's I-language generate *instructions* to what Chomsky calls the "performance systems" of the mind/brain. More specifically, the representations generated by a person's I-language instruct her articulatory system to engage in certain muscle contractions and her conceptual-intentional system to interpret incoming linguistic stimuli in highly specific ways.[6]

> Jones makes use of [her] competence to express [her] thoughts, to refer, to produce signals, to interpret what [she] hears, and so on. The language faculty is embedded in performance systems, which access the generative procedure. … We might think of the SDs as providing *instructions* to the performance systems that enable Jones to carry out these actions [i.e., to express her thoughts, to refer, to produce signals, to interpret what one hears, and so on]. When we say that Jones has the language L, we now mean that Jones's language faculty is in the state L, which we identify with a generative procedure, embedded in performance systems. (*The Minimalist Program*, p. 15, emphasis added.)

---

[6] Chomsky often suggests that an individual's language faculty assigns interpretations to languages that, intuitively, the individual does not know, so that "one might well learn about the languages of Jones and Wang [monolingual speakers of English and Chinese, respectively] by studying their reactions to utterances of Swahili" (*The Minimalist Program*, p. 16). Indeed, the suggestion is sometimes taken farther, as, for instance in the following passage: "The I-language … assigns status to a vast range of physical events, including the utterance 'John seems to be sleeping', the utterance 'John seems sleeping', a sentence of Hindi, and probably the squeaking of a door, if we could do careful enough experiments to show how speakers of English and Japanese might differ in the way they 'hear' the noise" (Chomsky 1988: p. 585).

We note, once again, the plausibility of interpreting Chomsky as claiming that states of the language faculty are "representational"—though, admittedly, in a sense that has yet to be defined. Indeed, in accordance with his slogan "No Computation without Representation," Jerry Fodor (1983, pp. 4–8) has argued that this interpretation of Chomsky is the *only* one on which talk of performance systems "*access*[*ing*] the generative procedure" and structural descriptions "providing *instructions* to the performance systems" (emphases mine) makes any sense. However, while Fodor's argument is compelling, it should be stressed that Chomsky's remarks, quoted above, are compatible with a broad range of perspectives on representation.

Fodor's influential view of representation casts it as a distinctive kind of *relation* between the mind/brain and various objects, properties, and events. However, alternative accounts of representation have been proposed and defended. In particular, Chomsky (2000) argues forcefully that linguistics is an "internalist" or "individualist" inquiry. By this, he means that linguistics—which he conceives of as a branch of psychology—deals solely with the properties of the mind/brain *irrespective* of the relations that it bears to anything in the environment. Thus, while it may be convenient to speak of various internal states as "representations," such talk should not, according to Chomsky, be interpreted as a commitment to the existence of a representation *relation*. In Chap. 7, I will discuss Chomsky's position on "representation" and argue that his internalism faces important difficulties. For the present, let us take "representation" as a black-box notion, and survey other aspects of Chomsky's position.[7]

Against the background sketched thus far, Chomsky formulates several guiding questions, the answers to which would jointly constitute a comprehensive theory of language.

(a) What does Jones know when he has a particular language?
(b) How did Jones acquire this knowledge?
(c) How does Jones put this knowledge to use?
(d) How did these properties of the mind/brain evolve in the species?
(e) How are these properties realized in mechanisms of the brain? (*The Minimalist Program,* p. 17).

Let us focus in on questions (a) and (c), which reflect the primary concerns of the present work.

Chomsky's answer to question (a) is this: In virtue of the fact that Jones' I-language generates structural descriptions for an infinite number of strings, Jones knows an infinite number of facts about the linguistic properties of expressions—properties such as rhyme (e.g., that 'pin' rhymes with 'bin'), grammaticality (e.g., that 'the child seems sleeping' is not completely well-formed), entailment (e.g., that the sentence 'Mary persuaded Bill to go to college' entails that Bill came to intend to go to college, though not that Mary did), ambiguity (e.g., that 'Mary is too angry

[7] See Devitt (2006a: ch. 4) for various interpretations of Chomsky's remarks on representation and the psychological reality of grammar.

to run the meeting' can mean either that *Mary* can't run the meeting, or that Mary is so angry that *we* can't run the meeting), and so forth.

Question (c) concerns the ways in which competent speakers put their "knowledge of language" to use. Chomsky writes that this question "calls for the development of performance theories, among them theories of production and interpretation" (*MP*, p. 18). Though he espouses a striking skepticism concerning the viability of this project,[8] Chomsky does say that "highly idealized aspects of the problem are amenable to study … on the standard empirical hypothesis … that one component of the mind/brain is a *parser*, which assigns a percept to a signal (abstracting from other circumstances relevant to interpretation)" (*ibid*). He then reveals what he takes to be the relationship between the parser and the mental representations that constitute an I-language: "The parser presumably *incorporates* the [I-]language *and much else*, and the hypothesis is that interpretation involves such a system, embedded in others" (p. 18, emphases mine).[9]

Note that Chomsky does *not* take the linguist's formal grammar—i.e., the theory of the generative procedure that constitutes an I-language—to be a theory of the computational processes that are effected in comprehension and production. In other words, Chomsky does not think of an I-language as a parsing or production mechanism. Rather, the parser "*incorporates* the [I-]language" in some manner. To clarify the issue, consider first a passage from Chomsky (1980a), in which he discusses a well-known syntactic phenomenon, namely, the impossibility of extracting material from *wh*-clauses. Chomsky argues for an explanation of this phenomenon that makes reference to phrase markers that belong to various "levels of representation" and are related by the formal operation of "movement". While the details are fascinating, what matters more for present purposes is Chomsky's commentary on the character of his proposed explanation—particularly its bearing on psychology.

> Tentatively accepting this explanation, we impute existence to certain mental representations and to the mental computations that apply in a specific way to these mental representations [and] [w]e attribute psychological reality to the postulated representations and mental computations. Have we gone beyond the bounds of what is legitimate and proper, in so doing? I think not. … [T]he argument sketched seems to me analogous in relevant respects to that of a physicist postulating certain processes in the interior of the sun. Of course, there are differences; the physicist is actually postulating physical entities and processes, while we are keeping to abstract conditions that unknown mechanisms must meet. We might go on to suggest actual mechanisms, but we know that it would be pointless to do so at the present stage of our ignorance concerning the functioning of the brain" (*Rules and Representations*, 1980: pp. 196–197).

---

[8] E.g., "Put generally, the problems are beyond reach: it would be unreasonable to pose the problem of how Jones decides to say what he does, or how he interprets what he hears in particular circumstances" (1995, p. 18).

[9] The phrase 'and much else' suggests that, on Chomsky's view, the parser may have access to information that goes well beyond what is contained in SDs, or in the syntactic rules and principles that "generate" SDs. In particular, the hint seems to be that the parser has access to the encyclopedic information required for "pragmatic" inferences regarding the speaker's message or communicative intent. Spelled out fully, the suggestion seems to be that the parser is not modular, in the sense of Fodor (1983).

The crucial point here is that the grammar does no more than impose "abstract conditions" on unknown mechanisms. We might ask, of course, how strong these conditions are. Specifically, must the mechanism construct syntactic representations in the course of on-line processing? Must a parsing algorithm literally *move* parts of some of these representations in generating the others? Chomsky's remarks in his earlier work make it clear that this *not* what he intends.

> [I]t is important to distinguish between the function and the properties of the perceptual model PM and the competence model G that it incorporates… Although we may describe the grammar G as a system of processes and rules that apply in a certain order to relate sounds and meaning, we are not entitled to take this as a description of the successive acts of a performance models. In fact it would be absurd to do so…. If these simple distinctions are overlooked, great confusion must result" (*Language and Mind*, 1968/1972: p. 117).

In a similar vein, Jackendoff (2000) writes:

> As has been remarked for decades, this framework of grammatical theory cannot serve directly as a model of processing. We were all taught as graduate students that one should not think of a syntactic derivation as modeling the course of processing: one doesn't think of the initial symbol *S* first, then gradually expand it till one chooses the words, then push the pieces around until one finally decides what the sentence means and how to pronounce it. So the notion of derivation has been distanced from processing by calling it "metaphorical"; this term has been applied especially to processes of syntactic "movement." (pp. 21–22)

Jackendoff's remarks in this passage echo an important footnote in *The Minimalist Program*, where Chomsky points out that "the ordering of [syntactic] operations is abstract … with no temporal interpretation implied. In this respect, the terms *output* and *input* [to a movement operation] have a metaphorical flavor…" (p. 380, fn. 3).

The distinction between theories of linguistic structure and theories of psychological processing mechanisms is an immediate consequence of another distinction that Chomsky drew in *Aspects of the Theory of Syntax* (1965), between *competence* and *performance*. As noted earlier, a speaker's linguistic competence is identified, in Chomsky's work, with her knowledge of the syntactic rules or principles that constitute her I-language.[10] While these are often characterized as components of a "generative procedure," the passages quoted above should make it clear that the "generation" or "derivation" of structural descriptions *need not* be a process that takes place in the mind/brain in real time.[11] Rather, Chomsky believes that the

[10] One might think of linguistic competence as an ability of some sort—an instance of what many philosophers, following Ryle, have called "knowledge-*how*." In his early writings, Chomsky described it in these terms. For instance, in his famous review of Skinner's book, *Verbal Behavior*, Chomsky wrote: "The construction of a grammar which enumerates sentences in such a way that a meaningful structural description can be determined for each sentence does not in itself provide an account of … actual behavior. It merely characterizes abstractly the *ability* of one who has mastered the language to distinguish sentences from nonsentences, to understand new sentences (in part), to note certain ambiguities, etc." (emphasis added). More recently, however, Chomsky has resisted this way of construing the nature of competence, insisting on an alternative construal that emphasizes what philosophers call "knowledge-*that*". To my mind, his reasons for doing so are not compelling. For a lucid discussion of this issue, see Devitt (2006a), ch. 6.

[11] It must be noted that, in his early work, Chomsky flirted with the idea of treating a grammar as a model of processing. Berwick & Weinberg (1984) write: "Miller and Chomsky (1963) identified rules of the grammar with computational operations of the parser in a one-to-one fashion. This

I-language is somehow *incorporated* or *recruited* by the psychological mechanisms that effect comprehension and production. Such performance mechanisms, in contrast to the I-language that they incorporate, *do* perform computations in real time. Their inputs in comprehension consist of both linguistic stimuli *and* the I-language, which can be seen (for these purposes) as a static data structure—in effect, a subpersonal analogue of a theory of the inputs. The psychological and behavioral phenomena that arise from the interaction between the I-language and these real-time computations fall under the heading of *performance*.

The competence-performance distinction has been the source of a great deal of confusion, and the topic of much heated debate, both in linguistics and philosophy. One set of issues that frequently arises in this connection has to do with whether, and to what extent, linguistic theory—conceived of as the study of what Chomsky calls "I-language"—is subject to confirmation or refutation by results in psychology and neuroscience. Consider, for instance the following remarks from Eddington (2008).

> [S]ome linguists hold that their analyses reflect the language of an ideal speaker/hearer and not actual speakers. In like manner, many linguistic analyses claim to reflect a speaker's competence—that is, the system of abstract mechanisms that are thought to underlie the speaker's ability to produce and understand language—and not the speaker's performance, which consists of actual utterances and other behaviors (Chomsky 1980a: 205). Accordingly, hypotheses about ideal speaker/hearers or competence are irrefutable; they do not relate to observable, real-word entities. No tangible evidence of any sort could contradict them because they are hypotheses about fictional entities. The grammar of an abstract speaker/hearer may be an interesting topic of study just as a study of the psychology of the character Jean Valjean in the novel *Les miserables,* or a philosophical treatise on the inherent goodness of man; however, they are not scientific. If all observable manifestations of language relate to performance, they can never be directly relevant to the study of abstract competence that many linguistic studies purport to deal with. Derwing (1983: 66) demonstrates how the competence/performance distinction serves to insulate a theory from possible falsification: "Suppose we find some child who is quite adept at basic arithmetic. One possible hypothesis about the 'competence' thought to underlie this skill might be to attribute the child [sic], not with something so mundane as a learned, laborious, step-by-step procedure for carrying out simple arithmetic operations, but rather with knowledge of number theory. And what if experimental results are found that seem to fly in the face of this hypothesis? Just chalk them up as 'performance errors' and the well-formed theory remains inviolate." (p. 4)[12]

identification led to specific behavioral predictions, collapsing grammatical with processing complexity… Again, this simple first attempt was the natural one. If it had been correct, we would have learned a lot about the parsing device…" (pp. 38–9). The suggestion in Miller and Chomsky (1963) is, however, quite tentative. "The psychological plausibility of a transformational model of the language user would be strengthened, of course, if it could be shown that our performance on tasks requiring the appreciation of the structure of transformed sentences is some function of the nature, number, and complexity of the grammatical transformations involved" (p. 481; quoted in Berwick and Weinberg 1984: p. 39)

[12] See also de Beaugrande's claim: "All data that the grammar cannot treat are shunted off into the domain of 'performance' and are excluded from consideration as nonissues." *Text, Discourse, and Process*, Norwood, NJ: Ablex, 1980. (Quoted in Bever, Carroll, and Miller [eds.], 1984: p. 135.)

Chomsky would surely reply that Eddington has missed an important distinction, between studying idealizations—e.g., the paradigmatic frictionless plane of basic physics—and studying outright fictions, like the character Jean Valjean. Pure fiction involves the description of *nonexistent* entities, whereas scientific idealizations are deliberately simplified descriptions of *actual* entities. True, in neither case do the descriptions accurately and exhaustively portray any real entity. But this resemblance is superficial. Idealizations are inaccurate because they are *incomplete*; fictions are inaccurate because there is nothing for them to be accurate *about*.

In my view, Derwing's contribution to the debate, reflected in the passage quoted above, likewise rests on an error. The line between competence and performance is admittedly quite blurry. (Indeed, I will capitalize on this fact in discussing the identity conditions on I-languages in Chap. 3.) But, *pace* Derwing, it does *not* follow from this that theorists have *no grounds at all* for rejecting certain proposals about the nature of competence. To take an extreme example, no linguist would be tempted to "chalk up as 'performance errors' [*sic*]" an English speaker's inability to comprehend German. This is *plainly* a competence issue. More subtly, no linguist has ever been tempted to propose that English speakers' judgments about the unacceptability of (1) are cases of performance error, despite the acceptability of (2) in a *pro*-drop language like Italian.

(1) *Is raining. [Intended: It is raining.]
(2) Piove. [Italian for 'It is raining']

Other theorists have argued that, while results from psychology and neuroscience *can*, in principle, bear on linguistic theory, they have not, as yet, been *brought* to bear. This is the position adopted by Edelman and Christiansen (2003), who argue that the psychological reality of *Merge* and *Move*—the central formal operations in Minimalist syntax (Chap. 9)—has not been empirically addressed, let alone demonstrated.

> Unfortunately, to our knowledge, no experimental evidence has been offered to date that suggests that merge and move are real (in the same sense that the spatial frequency channels in human vision are). Generative linguists typically respond to calls for evidence for the reality of their theoretical constructs by claiming that no evidence is needed over and above the theory's ability to account for patterns of grammaticality judgments elicited from native speakers. This response is unsatisfactory, on two accounts. First, such judgments are inherently unreliable because of their unavoidable meta-cognitive overtones, because grammaticality is better described as a graded quantity, and for a host of other reasons [Schuetze 1996]. Second, the outcome of a judgment (or the analysis of an elicited utterance) is invariably brought to bear on some distinction between variants of the current generative theory, never on its foundational assumptions. Of the latter, the reality of merge and move is but one example; the full list includes assumptions about language being a 'computationally perfect' system, the copy theory of traces, the existence of Logical Form (LF) structures, and 'innate general principles of economy'. Unfortunately, these foundational issues have not been subjected to psychological investigations, in part because it is not clear how to turn the assumptions into testable hypotheses. (Edelman and Christiansen 2003: p. 60)

Edelman and Christiansen share an important theoretical commitment with Eddington and Derwing—namely, the claim that the communicative behaviors and linguistic intuitions of competent speakers are *not legitimate sources of empirical data*. This belief underlies Eddington's striking claim that some conceptions of linguistics render the inquiry in this field "nonempirical."

> Some would argue that according to their conception of linguistics there is no need for experimental verification. These researchers would probably agree with Itkonen (1976:15–16) that the nonempirical linguist's goal is "to generate all and only intuitively valid formulae: insofar as they fail to do this, their systems are (non-empirically) falsified … not by reference to some specific spatiotemporal occurrences, but showing that it does not capture the concept which it tries to capture." (See also Carr 1990:66, Kac 1992:39, Linell 1976:84–85) Linguistic analyses in this sense belong to a metaphysical or philosophical realm of inquiry that deals with axiomatizations about linguistic structure which 'make it possible to deduce all true statements about the system from a small set of prior assumptions about its nature' (Kac 1974:44), much in the same way arguments in philosophy or logic do. However, a serious concern in linguistics is determining whether the claims in an analysis are philosophical or scientific. Of course there are those who do clearly define their position. For example, Marantz (2005:440) believes that 'the categories and operations of generative grammar are hypotheses about the representations and computations in the minds and brains of speakers.' Bromberg and Halle also adopt a realist stance to phonology: 'Do speakers *really* retrieve morphemes from their memory, invoke rules, go through all these labours when speaking? We think they do' (2000:35). According to Marantz and Bromberg and Halle, linguistics belongs in the realm of scientific empiricism. At the opposite end of the spectrum, Bradley claims that 'grammars do not (and moreover, are not intended to) dictate the ways in which the computation of speaking and listening proceed …' (1980:38). Her definition of linguistics is clearly nonempirical. (pp. 6–7)

This passage suffers from a number of serious misunderstandings. First, as I argued in Chap. 1, the reasons that are traditionally given for denying that philosophy and logic are every bit as empirical as physics and chemistry are actually quite weak. Second, the quotation from Dianne Bradley at the end of the passage aims to distinguish theories of the competence grammar from theories of the computational performance mechanisms that draw on that grammar. But, *pace* Eddington, insisting on this distinction is *not* tantamount to endorsing the bizarre view that linguistics is "nonempirical." At bottom, this is because Eddington's view about the methodology of gathering linguistic intuitions from reliable informants is simply mistaken. As Phillips and Lasnik (2003) point out in their reply to Edelman and Christiansen, this methodology is not only extremely fruitful, but is *empirical* in a quite straightforward sense.

> Gathering of native-speaker judgments is a trivially simple kind of experiment, one that makes it possible to obtain large numbers of highly robust empirical results in a short period of time, from a vast array of languages. Any good linguistics study involves carefully constructed materials, appropriate control items, and robust and replicable results. It is only because the technique is so easy and requires no more than a notebook that it is not usually described as an "experiment". Note that when 4-year-olds are involved, the same task calls for a quiet room, toys, and various clever ruses, and then everybody agrees that it is an experiment. Outsiders would surely be puzzled by the attitude that seeks to deny the psychological relevance of easy, robust results, while insisting on other, far more subtle measures, such as 20-ms differences in reaction times, or 1-s changes in how quickly babies get bored, or 2% changes in regional cerebral blood flow. The variability that one observes in

> native-speaker judgments is real, but very small relative both to the agreement among speakers and relative to the variability in other measures… (p. 61)[13]

For his part, Chomsky has likewise endorsed the view that native-speaker judgments serve as an important and fertile source of empirical evidence for linguistic theory, alongside others, including fine-grained behavioral and neurocognitive data from psycholinguistics.

> In the real world of actual research on language, it would be fair to say, I think, that principles based on evidence derived from informant judgments have proved to be deeper and more revealing than those based in evidence from experiments on processing and the like, although the future may be different in this regard. If we accept—as I do…—[the] contention that the rules of the grammar enter into processing mechanisms, then evidence concerning production, recall, and language use in general can be expected (in principle) to have a bearing on the investigation of rules of grammar, on what is sometimes called "grammatical competence" or knowledge of language. (*Rules and Representations*, pp. 200–201)

There is, of course, a great deal more to say about the role of intuitions in linguistics.[14] My purpose here, however, is only to point out that the linguists' heavy reliance on intuitions as a source of data neither renders their inquiry unempirical nor precludes them from recognizing the legitimacy of other types of data.

The Chomskyan conception of language, sketched above, serves as a useful benchmark for the discussion ahead. It is dominant in both linguistics and philosophy, and informs much of the work on the psychological reality issue.[15] Nevertheless, it has seemed to many that Chomsky's proposals regarding the nature of language are, in one way or another, unclear or mistaken. Others have argued that his views concerning the role of "mental representations" in linguistic theory is open to substantive challenge. In what follows, I outline a number of qualms that Chomsky's critics have had with his methodological and ontological assumptions—several of which have already received some discussion in the pages above.

To begin with, there is a lack of clarity concerning a host of key notions in Chomsky's theoretical framework. He holds that competent speaker/hearers have "tacit knowledge" of the generative procedure that "outputs" SDs—i.e., tacit knowledge of a grammar. But in what sense do adult language users *know*, *represent*, or "cognize" the grammar of their language? Such worries about the notion of tacit knowledge tend to go hand-in-hand with broadly philosophical concerns regarding Chomsky's persistent reliance on the notions of computation, information, and representation, illustrated in the quotations above. What is it, one might wonder, for an I-language—rather than a parsing mechanism—to perform "computations" defined over linguistic rules, principles, or categories? If the I-language is a static knowledge-structure or, even weaker, an "abstract condition on unknown mechanisms," then

[13] See Sprouse and Almeida (2013) for detailed studies of the consistency among native speakers.

[14] See, e.g., Devitt (2006a: ch. 7), Culbertson and Gross (2009), Devitt (2010a, c), Culbertson and Gross (2011).

[15] See Devitt (2006a: chs. 1 and 4) for a number of quotations from linguists, philosophers, and psychologists who seem to endorse the basic tenets of Chomsky's ideology.

why characterize the "generation" of linguistic expressions, or the "derivation" of their associated SDs, as a *computational* procedure? Finally, and perhaps most importantly, what conceptions of "information" and "representation" are at play in Chomsky's work?

Underlying all of these questions is the suspicion that there may well be a compelling alternative to viewing linguistic theorizing as an exclusively psychological enterprise. Perhaps, as many philosophers have suggested, it is better to see linguists as engaged in a systematic inquiry into the nature of what Chomsky calls "E-language." On such a view, the "generation" of SDs by a grammar would not be computational, in the literal sense—a claim that accords with Chomsky's insistence that I-languages need not, and perhaps *ought* not, be seen as algorithms that the brain runs in real time. Perhaps the generation and derivation of SDs should be seen as being somehow *potential*, as when the axioms of a mathematical theory are said to "generate theorems," though the axioms exist only as static ink-marks on a page.[16]

One starting point for addressing these issues is a published exchange between Chomsky and the philosopher, Georges Rey.[17] The questions that Rey poses for Chomsky are, in effect, the ones raised above—the very questions to which philosophers of mind have sought answers from the very outset of the generative linguistic enterprise. Importantly, Rey's line of questioning is agnostic about which theory of representation, computation, information, or intentional content is correct, and which, if any, is applicable to the present case. Rey is also relatively neutral on whether Chomsky's theory does indeed employ or require any of the notions in question. Finally, in contrast to a number of philosophers who have engaged with Chomsky,[18] Rey does *not* mount an outright attack; his essay constitutes more of an inquiry than a critique.

In his reply, Chomsky argues that the notions Rey and other philosophers employ in characterizing his view—paradigmatically, the notions of "representation" and "intentional content"—are both degenerate and theoretically idle. According to him, philosophers have given no precise meaning to their talk of reference, representation, and intentionality, and insofar as such notions are at all clear, they are in no way relevant to the sort of empirical pursuit that linguists have undertaken in the past several decades. Chomsky also reiterates his contempt for the theoretical construct of an E-language, and insists once more that there is nothing more to the psychological reality of a syntactic rule or principle than its being mentioned in a true linguistic theory. The suggestion, it seems, is that while there is no coherent alternative to a cognitivist conception of linguistics, the notions that render it a psychological inquiry are not those that animate philosophical debates. Those who find it difficult to reconcile these claims with Chomsky's persistent use of intentional idioms may find it comforting to note that he takes these uses to belong to merely

[16] For two rich and engaging discussions of this analogy, see Levin (1979) and Soames (1984).

[17] See Rey (2003a), and Chomsky's reply in the same volume. Rey (2003b) follows up, raising largely the same issues, though in a more refined manner.

[18] E.g., Harman (1967, 1969) and Quine (1975a, b).

"informal" presentations of his theory. In its purest form, he insists, the theory eschews intentional notions altogether.[19]

As previously noted, an opposing view of the relation between a competent speaker and the grammar of her language can be found in Devitt (2006a, b). In contrast to Chomsky's position, the alternative that Devitt develops is accepted by a scattered few (Quine 1953, 1970), Black and Chiat (1981), Gazdar, Pullum, Klein, and Sag (1985). Nevertheless, many aspects of it seem to be both plausible and illuminating. Below, I give a compressed account of the view, and rehearse some of the reasons for accepting it over its Chomskyan rival.

2.3 Devitt's "Linguistic Conception"

Michael Devitt's book, *Ignorance of Language* (2006a, b), is a study of the galaxy of issues that jointly comprise the debate concerning the psychological reality of language. Devitt begins by criticizing Chomsky's view that grammars are theories of a psychological faculty and, hence, are to be regarded as psychologically real to the extent that they are true. He takes this argument for the psychological reality of grammars to be "not only fast but dirty" (9). He argues that grammars are *not* theories of the mind and that a "powerful psychological assumption" must be made in order to ascribe psychological reality to grammatical rules or principles. One such assumption might be what Devitt calls the Representational Thesis (RT).

RT "A speaker of a language stands in an unconscious or tacit propositional attitude to the rules or principles of the language, which are represented in her language faculty" (4)

According to Devitt, Chomsky and others may very well be relying on RT (perhaps tacitly) when they claim that principles of grammar are psychologically real. Consequently, Devitt devotes much of *Ignorance* to arguing against RT. Chapters 8 and 9 of the present work examine the case for RT. For the moment, let us keep to ontological issues and explore Devitt's conception of what grammars are theories *of*, if not psychological states or mechanisms.

In contrast to Chomsky's cognitivist conception, Devitt (2003, 2006a, b, 2008a, b) holds what he calls the Linguistic Conception, according to which token linguistic expressions are typically external to the skin of any competent speaker. When spoken, a token expression is an acoustic blast, scattered over time. When written, it is made of ink, lead, chalk and the like.[20] Token expressions have a definite loca-

[19] Chomsky (2000) develops an "internalist" account of representation. We will examine this in Chap. 7.

[20] It is sometimes suggested that no sense can be made of the claim that a physical object, state, or event can have a syntactic structure. Notice, however, that all the parties to the debate over the Language of Thought Hypothesis (LOTH) are committed to the coherence of this proposition—

tion in space-time, and are concrete, in a quite ordinary sense: we can touch them, burn them, photograph them, and so on.[21] Devitt holds that stating the individuation and persistence conditions of linguistic expressions involves making ineliminable reference to the psychological states of competent speakers.[22] A token physical entity or process counts as belonging to a particular type of linguistic expression in virtue of three factors: (i) the communicative intentions of its author, (ii) the psychological state that it is used to convey, and (iii) its relations to other linguistic expressions.[23]

To further flesh out Devitt's view, it will be useful to organize the discussion around four issues concerning the metaphysics of language and the ontological commitments of linguistic theory. These issues concern (i) the individuation of linguistic expressions, (ii) the irreducibility of linguistic properties, (iii) the infinitude of natural language, and (iv) the status of phonologically null constituents ("empty categories"). I will refer to (i)–(iv) as "Significant Metaphysical Issues," to mark the fact that they concern the most hotly debated topics, where the claims that define Devitt's overall position come into sharpest relief and, unsurprisingly, receive the lion's share of his opponents' criticism.

---

though not necessarily its truth. For, they all agree that the debate is about whether straightforwardly physical events—i.e., the neurochemical vehicles of thought—are syntactically structured. Thus, in order to make sense of the debate, one has to first take on board the view that datable occurrences in a physical medium might be syntactically structured. Suppose, then, that an opponent of the Linguistic Conception denies that datable physical events can have syntactic structure. He thereby commits himself to the view that the LOTH debate is strictly incoherent. This strikes me as a *reductio*; the LOTH debate can't be dismissed *that* easily.

[21] In Chomsky's terminology, a set of such expressions constitutes an E-language. It is not clear, however, that Devitt would accept all of the theoretical baggage with which the expression 'E-language' has been loaded. (See his (2003: pp. 120–21).) I discuss the issue at some length in Chap. 3. It is also worth pointing out that Chomsky, himself, occasionally slips into treating linguistic expressions as concrete, physical events. In his paper, "Explanatory Models in Linguistics" (1962), he writes: "What [the child] accomplishes can fairly be described as theory construction of a non-trivial kind … a theory that predicts *the grammatical structure of each of an infinite class of potential physical events*' (p. 528, emphasis mine). Even if Chomsky did not intend, here or elsewhere, to claim that physical events are literally the bearers of grammatical structure, the remark illustrates how natural it is to think and speak of them in this way.

[22] The relevant states, however, need not be metalinguistic intuitions. Such intuitions are, of course, fairly reliable and relatively easy to elicit, given the right setting. But, according to Devitt, they are dispensable in principle. The psychological states that are crucial for the purpose of individuating linguistic expressions are *thoughts* and *intentions*.

[23] There are, of course, difficulties in making the individuation conditions precise in this way. I shall pass over these here, mentioning only that the account of convention in Lewis (1969/2002) may be a good starting point for those wishing to make some headway on this problem. See also Millikan (2004).

2.3.1 Objections and Replies

The first Significant Metaphysical Issue has to do with the individuation and persistence conditions of linguistic expressions. As noted above, Devitt concedes that these conditions are spelled out in terms of the psychological states of competent language users. He points out, however, that this does *not* warrant the identification of linguistic expressions with the psychological entities that Chomsky calls 'structural descriptions' (SDs). The identity and persistence of a linguistic expression surely *supervenes* on psychological states, but the same can be said of any number of things that are plainly *not* psychological entities, e.g., the United States government.[24] Furthermore, recognizing the supervenience relationship between linguistic expressions and psychological states does *not* warrant the methodological reductionism that is encapsulated in the oft-repeated slogan that "linguistics is a branch of psychology" (Chomsky 1972: p. 1). Extending the analogy with governments, we note that political science is not a branch of psychology, though the two fields may be intimately related, and may well provide one another with valuable data. By the same token, biology is not, in any useful sense, a branch of quantum physics, despite the fact that all biological entities, mechanisms, and processes supervene on the activity of sub-atomic particles. Devitt (2006a) puts the point as follows:

> Even if symbols had their properties in virtue of certain mental facts that would not make the theory of those symbols about those facts and so would not make the theory part of psychology. Indeed, consider the consequences of supposing it would, and then generalizing: every theory—economic, psychological, biological, etc.—would be about physical facts and part of physics because physical facts ultimately determine everything. A special science does not lose its own domain because that domain supervenes on another. (p. 40)

This point is crucial, as it forestalls an all-too-common objection to the view that linguistic expressions are non-psychological, concrete denizens of space-time. Plainly, if there were no people, or no psychological states, then there would be no language. Therefore, the objection goes, language must be a psychological entity. Though certainly tempting, this line of reasoning misses the mark. Supervenience—a relation weaker than both identity and constitution—is all that is needed to account for the connection between psychological states and the persistence or identity of linguistic expressions.[25] If Devitt's view is correct on this point, then a reduction of linguistic expressions to psychological states is simply not in the offing.

The second Significant Metaphysical Issue has to do with the lack of success that phonologists have had in discovering robust acoustic commonalities between tokens of what is, intuitively, the same phone type. As is well known, there are no strictly acoustic criteria—no necessary and sufficient conditions stated in the language of physics—for the identity of a phone. This is so even in the case of a single speaker, *even in a fixed acoustic environment*, let alone for phones as they occur in a realistic

[24] See also Rey (2003b), pp. 170–71, where essentially the same point emerges.

[25] No particular account of the supervenience relation is being assumed. To my knowledge, all of the going accounts are consistent with the claims made above.

community of speakers. Consequently, the structuralist dream of articulating "objective discovery procedures" for phonological categories has been abandoned. The situation is, moreover, not confined to phonology. Taxonomic linguistics seems to run into decisive difficulties at *every* level of analysis.[26] From this, it has been inferred that phones, morphemes, and phrases must be psychological entities.

Devitt's Linguistic Conception points to a difficulty with this inference. Though the demand for purely acoustic individuation and persistence conditions cannot be met, the psychological criteria that do the necessary work do *not* substantiate the claim that linguistic expressions are, in the first instance, mental or psychological entities. Rather, phones, morphemes, and phrases are sets of concrete tokens that resemble one another primarily in respect of the psychological states by which they are caused, and the states that they cause in turn. This account lends no support to the *identification* of linguistic expressions with psychological entities. As noted above, such entities *supervene* upon distributions of psychological states, but they are not *identical* with such states. Black and Chiat (1981) take essentially the same line when they write:

> Talking about the grammar or language as abstract systems is no more problematic than talking about society or the economy as abstract systems. The concepts of many social sciences are not reducible to physical reality, and yet they do not have to be interpreted as psychological objects either. We clearly would not want to say that concepts like 'kinship' or 'circulation of money', which cannot be correlated with particular physical entities, correspond to something 'in the mind'. Nor does a psychological interpretation of grammar solve any philosophical problems that might arise with respect to its ontological status. (p. 41)

Devitt's frank recognition of the fact that the categories of linguistic theory are not reducible to those of physics is just one of the features of his view that separates him from the traditional nominalism of structuralist linguists, such as Bloomfield. He writes:

> [Jerry Katz] calls Chomsky's view "conceptualism" and my sort of view "nominalism". He takes nominalism to have been refuted by Chomsky's criticisms of Bloomfieldian structuralism. Yet, so far as I can see, these criticisms are not of the *nominalism* of the structuralists but rather of their *taxonomic methodology*, a methodology in the spirit of positivism. According to Chomsky, this methodology imposed "arbitrary and unwarranted" limitations on linguistics: it insisted on defining "lower levels" before "higher levels"; it was inductive instead of explanatory (abductive); its epistemology was localist instead of Quinean holist. Indeed, despite the explicit nominalism of the structuralists, Chomsky is prepared to take the structuralists as implicitly concerned with the psychological reality of language and hence not really nominalist at all (Chomsky 1975c: 30–6). Yet he still thinks his methodological criticisms stand. In any case, Chomsky's methodological criticisms can be and, in my view, should be embraced by the nominalist. In particular, we should not demand that the linguistic properties of tokens be reduced to "brute-physical" intrinsic properties of the tokens. The linguistic properties that concern us are "high-level" relational properties. (2006: p. 26)

---

[26] Fodor, Bever, and Garrett devote two chapters of their classic 1974 text, *The Psychology of Language*, to a seemingly exhaustive survey of the arguments against the structuralist program of taxonomic linguistics.

Thus, despite sharing with traditional nominalism a focus on physical tokens—inscriptions, acoustic blasts, bodily movements, and the like—Devitt's Linguistic Conception is not intended to inherit Bloomfield's conception of grammar as a taxonomic enterprise, nor the methodology of "discovery procedures" that rely on so-called "substitution criteria." Rather, Devitt intends his ontological position to be compatible with the deeper *explanatory* goals of contemporary generative grammar, and with much (though perhaps not all) of the methodology that is characteristic of Chomskyan linguistics. The pursuit of these goals, he recognizes, requires linguistic theory to "abstract away… from a range of properties of [physical entities]—for example, form of script and pitch of sound—focusing simply on the syntactic properties that we are interested in." And, as Devitt makes clear, such properties are typically *relational*, hence difficult to discern without a prior theoretical grip on what constitutes a language.

> [T]he linguistic structure of an utterance is not obvious and superficial. But this structural property is relational not intrinsic and relational properties are typically not obvious or superficial. Yet objects really have relational properties; for example, some objects really are paperweights, moons, echidnas, Australians, and so on. Sometimes, it is easy to tell that an object has a certain relational property because that property is well correlated with superficial properties. This makes it quite easy to tell an echidna, but not an Australian (if she keeps her mouth shut). And it makes it quite easy to tell many English adverbs, the ones that end in 'ly'. It can also be easy to tell that an object has a certain relational property if learning to identify the object involves learning to identify it as having that property. This makes it quite easy to identify the other English adverbs; identification comes with word recognition. One way or another, it is quite easy to tell the explicit structural properties of utterances although sometimes hard to tell the implicit ones. But utterances still really have both. (2006: 185–6)

The third Significant Metaphysical Issue has to do with the infinitude of natural language, discussed in the previous chapter. Various grounds have been offered for the claim that linguistic theory is concerned with an infinite domain. Chomsky has argued that an observationally adequate grammar for any natural language must be recursive. And, as above, we know of no principled bound on the length of natural-language sentences; one can always add 'Aron said that' (or a suitable translation) to any well-formed declarative sentence, generating a longer one. Moreover, some theorists hold that there exist sentences that will never be uttered. How then can linguistics be concerned with physical tokens, of which there is only a finite number?

As already noted (Chap. 1), Devitt's response is that the infinitude of language is not a *datum*. He denies that there are infinitely many sentences. Rather, the infinitude of language is an artifact of theory construction—a reflection of the fact that we seek maximally wide generalizations, in our effort to accommodate *any* potential datum that we might ever encounter. The linguist takes as her object of theorizing both past and future linguistic tokens. And while the number of these is finite, there is no way of telling in advance which particular tokens we will encounter. The continual novelty of language is a pattern that shows all signs of projectibility; increases in the human life span, motivational set, and cognitive resources may encourage production of novel sentences that are significantly longer than those we observe today. There is no telling in advance what the limits are on either linguistic creativity or sentence length. This alone is sufficient grounds for holding that sentence length is unbounded and that natural languages house denumerably many sentences. No further ontological commitments follow from this.

Turn now to the fourth Significant Metaphysical Issue, which pertains to the ontological status of so-called "empty categories"—i.e., unvoiced ("phonologically null") entities that are posited by syntactic theories in the Principles and Parameters (P&P) tradition.[27] We will explore these entities in more detail in Chap. 5, but, for the moment, we can summarize the motivations that lead syntacticians to posit them: Empty categories allow a grammar to encode various semantically relevant structural relationships between the constituents of a phrase. Let us distinguish four types of empty category. (Following the notational conventions of Government and Binding theory, I use the labels 'trace' and 'PRO'.)

(3) [What instrument]$_i$ do you play *wh-trace$_i$*?
(4) [$_{NP}$ The boat]$_i$ was carried NP-*trace$_i$* by five people.
(5) Jeffrey$_i$ tried PRO$_i$ to record a song.

The traces in (3) and (4) mark the fact that the phrases with which they are co-indexed are the objects of the verbs 'play' and 'carried', respectively. The empty category, PRO, in sentence (5) signals that the co-indexed NP, 'Jeffrey', is the subject of the infinitival verb 'to record'.

While the notion of an empty category helps the syntactician capture significant structural relationships and semantic regularities, it creates trouble for both the psycholinguist and the metaphysician. Postponing our discussion of psycholinguistic matters until Chap. 5, let us remain focused on issues of metaphysics. On a standard view in linguistics, which derives from Chomsky's cognitive conception of language, empty categories are a kind of *psychological* construct. The I-language, viewed as a psychological entity, is home to various types of language-specific mental representations, many (perhaps all) of which do not "correspond" to anything in the physical environment. For instance, the I-language contains phonological representations that, as we saw above, do not correspond in any straightforward way to the purely acoustic properties of any actual speech stream. Moreover, contemporary syntactic theory views phrases as having a hierarchical structure that encodes, at a minimum, the relations between the constituents of a phrase. Such relations are not overtly marked in the linear order of a speech stream or an inscription.

Proponents of the cognitive conception infer that these relations do not hold between items in the physical environment. And if the relations are not, in this sense, "out in the world," then they must be in the mind—where else? Less poeti-

[27] In reality, empty categories were introduced some time before the P&P model was fully articulated. Note also that traces are posited only by grammars that rely on movement operations—e.g., the Government and Binding theory. Other types of empty category (e.g., PRO) do not, however, encode the effects of movement. I will not discuss here whether Devitt's ontology can accommodate the sorts of entities that are posited by formalisms such as LFG and GPSG, which generate syntactic structure without recourse to movement operations. I presume that analogous considerations will hold for those cases.

cally, the conclusion that is often drawn is that a person's I-language generates mental representations of hierarchical structure. In the course of language comprehension, these representations are *imposed* or *projected* on a speech stream or an inscription. This explains how and why these physical entities are perceived by competent speakers as having such properties, *despite not actually having them*.[28] Against this background, the posit of empty categories in the syntactic structure of a phrase is not at all puzzling. Empty categories are just another aspect of the mental representations that ultimately constitute language. And because they are on a par with phonological and syntactic structures, it is no surprise that they are not overtly present in the physical form of speech or writing. For, on the Chomskyan view, no interesting linguistic properties *at all* are to be found there. Collins (2008b) sees this as a profound difficulty for Devitt's view. He writes: "*Prima facie*, perhaps the most serious problem facing Devitt is how the abstractness of syntax might be accommodated in his model of linguistic reality" (p. 20).

In response, Devitt rejects the Chomskyan picture of both language and psychology. He claims that the abovementioned phonological and syntactic properties *are* out in the world, in the sense that they are real properties of the inscriptions and acoustic blasts that, on his view, constitute language. Moreover, he denies that language comprehension requires the deployment of mental representations—metalinguistic descriptions, as it were—in response to linguistic stimuli. Indeed, his break with the cognitive conception is more radical still. For, on the strength of his arguments against the Representational Thesis and the psychological reality of syntax, he speculates that there may well be no such mental representations for a competent speaker to deploy. As we saw above, though, Devitt does recognize that it is impossible to reduce phonological and syntactic properties to the brute-physical features of inscriptions and acoustic blasts. He views them instead as "high-level" relational properties.

> [N]o naïve brute–physical account of the relation between sounds and phonemes is possible. Phonology shows that there are many complicated ways in which sounds can instantiate a phoneme, including relations to other sounds; and that a sound may be able to instantiate more than one phoneme. Similarly, there are many complicated ways in which inscriptions can instantiate a letter; and so on for other linguistic media. But this does not show that the sounds, inscriptions, etc. do not instantiate SLEs [Standard Linguistic Entities]. Quite the contrary. The property of being Australian is instantiated by a vast variety of physical forms; for example, the forms of the capitalist Rupert Murdoch, the runner Cathy Freeman, the horse Phar Lap, the city of Sydney, a bottle of Penfolds Grange, and the many forms of the saying "No worries, mate". The property of being the word 'cat' is instantiated by a much smaller variety of physical forms, a variety of sounds, inscriptions in many different fonts and handwritings, and so on. Just as all the former instantiations really are Australian, all the latter really are the word 'cat'. And note that just as some things do not count as Australian, some things do not count as the word 'cat'. (Devitt 2006a, b: p. 186)

With these remarks in mind, a metaphysician who is puzzled about the phonological and syntactic properties of phrases may well begin to see, if only in dim outline, how Devitt's position can stand as a coherent alternative to the Chomskyan view

[28] For a close examination of this line of reasoning, see Rey (2003a, b, 2006a, b, 2008).

outlined above. Focusing specifically on the case of syntactic properties, Devitt gestures toward a metaphysical position that was a hallmark of the structuralist tradition.

> The outputs of a linguistic competence, physical sentence tokens, are governed by a system of rules…Something counts as a sentence only if it has a place in the linguistic structure defined by these structure rules. Something counts as a particular sentence, has its particular syntactic structure, in virtue of the particular structure rules that govern it, in virtue of its particular place in the linguistic structure. (Devitt 2006a, b: p. 24)

Nevertheless, Devitt's picture seems to run into trouble with regard to empty categories. For, in the context of contemporary linguistic theory, an empty category—i.e., an unvoiced ("phonologically null") constituent of a sentence—is most naturally thought of as a strange kind of *object*, not a high-level relational *property*. For this reason, making clear Devitt's treatment empty categories is pivotal for the project of defending his overall metaphysical position. It is here that Devitt's appeal to public linguistic conventions is brought into sharpest relief.

> The problem [for my view] is thought to be particularly pressing in the case of "non-overt" or "unvoiced" constituents of a sentence. How is it possible for a sentence to have such a constituent? I responded: "The simple answer is: *there is nothing to stop there being a convention of this sort*" (2006b: 599). Consider the string 'Bob tried to swim'. The idea is, roughly, that each word in the string has a syntactic property by convention (e.g. 'Bob' is a noun). Put the words with those syntactic properties together in that order and the whole has certain further syntactic properties largely by convention; these further properties "emerge" by convention from the combination. The most familiar of these properties is that the string is a sentence. A more striking discovery is that it has a "PRO" after the main verb even though PRO has no acoustic realization. There is no mystery here. (Devitt 2008a, b, c: 217–18)

Thus, on Devitt's view, just as *being a sentence* and *having 'Bob' as the first noun* are properties of some physical tokens, so is the property of *having PRO as a constituent c-commanded by the NP*. These high-level relational properties are, metaphysically, all on a par.

Two problems can be raised for this response. The first, due to Collins (2008b) is that it is difficult to see how public-language conventions could "get a grip" on such subtle and fine-grained properties of linguistic expressions. Devitt (2008a) responds:

> I don't know. But I don't need to know to sustain linguistic realism. I have shown that it is plausible that a whole lot of sounds and inscriptions that humans produce form representational systems. Those systems are not fully innate and so must be partly conventional. I have shown how it is *possible* for conventions to yield unvoiced elements. I have indicated in a general way, referring to David Lewis (1969), how linguistic conventions, like other conventions (that are not stipulated), arise from regularities together with some sort of "mutual understanding." The regularities for linguistic conventions are in speaker meanings (pp. 156, 179–80). It would be nice to go much further, giving full explanations of the forming of linguistic conventions—indeed, of the forming of *any* conventions—but the hypothesis that there are such conventions does not depend on giving these. Lewis begins his book by claiming that it is a "platitude that language is ruled by convention" (1969: 1). This is surely right. … Of course, if it could be shown that the required conventions were *impossible* then my proposal would be in trouble. But giving such an "impossibility proof" would be a mighty tall order. After all, the phenomena that lead linguists to theorize that expressions have certain structures must be phenomena that speakers *could be* sensitive to in forming a language.

Collins will no doubt be left unsatisfied by this reply, finding fault with the lack of detail in Devitt's characterization of public language conventions. This, however, is not a damning criticism, for two reasons. First, as Devitt points out, Lewis (1969/2002) has supplied a great deal of relevant detail. Working out the applications of Lewis's general picture to the specific case of *syntactic* conventions is a research project whose eventual success we have been given no reason to doubt. Second, Collins' alternative account of empty categories relies on a fully internalist notion of mental representation, as well as an account of language acquisition in the absence of conventions. It is no stretch to say that these accounts are, at present, in no better shape than contemporary versions of a Lewisian theory of convention.

The second and, in my view, more difficult problem for Devitt's position emerges when we examine his strategy for dealing with empty categories. The problem, recall, was that contemporary linguistic theory treats these as a strange kind of *object*, not a high-level relational *property*. We saw above that Devitt's reply to this worry consists in recasting claims such as (7) as having the form of claims like (8).

(7) In the sentence 'Larisa wants to rest', there is an unvoiced entity, PRO, which is c-commanded by the initial NP.

(8) The sentence 'Larisa wants to rest' has the *property* of having an unvoiced entity, PRO, as a constituent which is c-commanded by the initial NP.

But treating apparent quantification over empty categories ("*there is* an unvoiced entity…") as garden-variety property attributions ("*has the property of having* an unvoiced entity…") is only the first step for Devitt. Talk of properties will eventually have to be paraphrased away as well. Devitt holds that there is no special difficulty here, beyond those that nominalists face in other domains.

> Is my contemplated task appropriately characterized as nominalistic? It takes all the objects that linguistics is about to be concrete tokens, and so to that extent it is nominalistic. Where it stands ultimately on the nominalism issue depends, of course, on what we make of its ascription of … *properties* to those objects. However, it seems unlikely that the nominalist would have any *special* difficulty paraphrasing away this property talk. My contemplated task for linguistics is likely to be as nominalistic as tasks in physics, biology, or economics. (Devitt 2006a, b: p. 30)

It is not clear, though, that this is so. In general, nominalists treat claims of the form 'x has F-hood' as saying no more than that x is F. But in the present case this strategy has the effect of paraphrasing claims like (8) into ones like (7), which appears to make reference to empty categories, construed as particulars that have no physical reality. The whole point of moving from (7) to (8) was to avoid such reference. While perhaps not as worrying as apparent reference to numbers ('There is a number between 1 and 3'), or the ascription of properties to properties ('Humility is a virtue'), this should, I think, trouble a nominalist.

Of course, nominalists have always faced issues to do with regimentation, and no shortage of clever solutions have been proposed.[29] This may well be another outstanding technical problem—different from those in other domains, but not espe-

[29] See Goodman and Quine (1947), Sellars (1963c).

cially worrying. One possible solution is to simply reject grammars that posit empty categories, in favor of ones (e.g., LFG and HPSG) that deal with movement, ellipsis, and similar phenomena in a different way. However, this would be to let ontological issues guide theory-choice in linguistics, and the methodological conclusions of Chap. 1 counsel us to be suspicious of such maneuvers.

2.3.2 Distinguishing Competence from Its Products, and Linguistics from Psycholinguistics

Having set out some of the metaphysical commitments that underlie his position, Devitt goes on to draw a number of distinctions, which, when applied to objects of our concern, elucidate the relationship between language and the mind, and, correlatively, between linguistics and psychology. For present purposes, the most important of these is the distinction between a competence and its products. Devitt illustrates this distinction with a simple and intuitive example. We are asked to consider a blacksmith who has the competence to produce horseshoes. This competence is, without a doubt, a psychological entity, consisting of myriad beliefs, preferences, action scripts, and so forth. Exercising this competence, the blacksmith creates a horseshoe, which, in contrast to his competence, is decidedly *not* a psychological entity. Of course, the fact that the blacksmith's creation is a *horseshoe*, rather than, say, a ceiling decoration or a doorstop, may very well supervene on the psychological states of the blacksmith or of those who go on to use the object for their own purposes. But, for reasons now familiar to us, this does *not* make the horseshoe a psychological entity.

Devitt appeals to the distinction between a competence and its products in his argument against the Chomskyan identification of a language with a speaker's competence. Just as horseshoes are the outputs of the blacksmith's competence, linguistic expressions are the outputs of a linguistic competence. The competence to produce and comprehend language, like the competence to make horseshoes, is a psychological entity. Hence, a theory of either sort of competence is *ipso facto* a psychological theory. But, on Devitt's view, one needs to supplement that theory with a *distinct* theory of the *products* of competence—whether those products are horseshoes or linguistic expressions. One must, that is, provide some relatively independent characterization of the domain that the competence engages with.

Devitt argues that the working syntactician is attempting to construct precisely such a characterization when she formulates grammatical principles, e.g., those constitutive of Binding Theory, Case Theory, empty category licensing, and the like. Other branches of linguistics attempt to specify the phonological and the semantic properties of linguistic expressions. From the perspective of the metaphysician, linguistic expressions have all of these properties in virtue of their interactions with a variety of psychological states—the now-familiar supervenience claim. But, from the perspective of the psychologist, these properties may nevertheless have to be "recovered," "extracted," or simply *recognized* by the hearer in the course of comprehension. Indeed, it is only in this way that we can make sense of a hearer's *misperceiving*, or being *mistaken* about the structure of a sentence in the course of comprehension. Such cases of misperception are, of course, all too common in daily life, and often serve as crucial data in psycholinguistic experiments (Chap. 5).

Devitt's conception of the relation between the grammar of a language and the competence to produce and comprehend linguistic expressions is not without some precedent. In a founding text of psycholinguistics, *The Psychology of Language* (1974), Fodor, Bever, and Garrett criticize a proposal concerning the mechanisms of language production as follows: "We see here a classic example of the problems generated when one fails to distinguish theories of language structure and theories of the computational mechanisms involved in the perception and production of speech" (p. 67). The distinction to which Fodor *et al.* are drawing attention is between what Devitt calls the "structure rules" of a particular language—primarily, the syntactic rules and principles governing the expressions in the language—and the "processing rules" that govern the psychological mechanisms responsible for the perception and production of such expressions. Phillips (1996) notes that this distinction has been implicit in much of the work done in the generative tradition.[30]

> Since the 1960's, work in linguistic theory has focused on characterizing linguistic knowledge in terms of *static* mental representations, and on accounting primarily for grammaticality judgments. The main goal of psycholinguistics, on the other hand, has more commonly been to provide a more or less explicitly *procedural* characterization of how speakers perform linguistic tasks such as comprehension or production. It is standardly assumed that mental systems for grammar and processing are separate, and hence that linguists and psycholinguists are not studying the same thing. (Phillips 1996: p. 3)

Following Chomsky, Phillips places *both* projects in the realm of psychological research. The formal linguists' project is conceived of as a study of the mind, on account of the fact that its domain is a set of "static mental representations."[31] Devitt has provided reasons for being skeptical of this way of construing the work of syntacticians and phonologists. The present point, however, is that, even when viewed in this way, there is *still* reason to distinguish linguistics from psycholinguistics.

[30] Though he ultimately rejects this distinction on principled grounds, Phillips does provide numerous references to other theorists who endorse it, including notable quotations from Crocker (1996), who writes: "This *grammar as parser* approach is not a rational position given the competence–performance division […] which clearly separates the declarative properties of the syntactic theory from any procedural notions" (p. 49), and "The suggestion that the grammar *is* the parser is simply not well-conceived" (p. 51). (From Phillips [1994], p. 16.) A standard textbook in computational linguistics, Jurafsky and Martin (2008) express the distinction as follows: "Syntactic parsing … is the task of recognizing a sentence and assigning a syntactic structure to it. This chapter focuses on the kind of structures assigned by context-free grammars ... [S]ince they are based on a purely declarative formalism, context-free grammars don't specify *how* the parse tree for a given sentence should be computed. We'll therefore need to specify algorithms that employ these grammars to produce trees" (p. 431)

[31] Phillips goes on to develop a proposal according to which the structure rules of a language—i.e. rules of the grammar—do the work traditionally carved out for processing rules. We examine this proposal in Chap. 9.

The domains of the two fields—structure rules and processing rules, the grammar and the parser—are simply not identical. This point holds regardless of whether one takes the structure rules to be psychologically realized as an I-language, or simply a set of rules devised by linguists to capture generalizations concerning what Chomsky would call E-language expressions.

2.4 Conclusion

In this chapter, we surveyed the main features of two competing approaches to linguistic metatheory— Chomsky's cognitivism and Devitt's nominalism. We took note of their metaphysical commitments as well as the rich resources that each of them brings to the task of answering foundational questions about language and linguistic theorizing. We also saw that both approaches face non-trivial difficulties. Chomsky's view awaits the articulation of a clear conception of the relation between grammars and neurophysiological mechanisms. Devitt's view requires a more detailed account of public language conventions and a satisfactory strategy for paraphrasing away apparent references to unvoiced elements of syntactic structure.

Of course, much more can be said—and, indeed, *has* been said—to clarify and defend both approaches. For our purposes, it will important to keep in mind a particularly significant contrast between them. Chomsky's view, unlike Devitt's, has the effect of closing off discussion about whether syntactic rules and principles are psychologically real. For, if one begins with Chomsky's cognitive conception of grammars, then the question of whether a syntactic rule or principle is psychologically real reduces to the question of whether we have good reason to believe the grammar that posits it. And, of course, we typically *do* have such reasons, for it is the working syntactician's business to provide *prima facie* reasons of just that sort. By contrast, if one begins with Devitt's linguistic conception, then the ascription of psychological reality to one or another syntactic principle requires a powerful psychological assumption—e.g., the Representational Thesis (RT).

One of the main goals of the present work is to examine the case for the psychological reality of syntactic rules and principles. Plainly, this project presupposes that the psychological reality issue is not *trivially* true, as it is on Chomsky's cognitive conception of language. For this reason, in the next two chapters, I undertake a critical examination of two well-known arguments for the cognitive conception, having to do with the notions of E-language and explanatory adequacy. My critical task will be a modest one. To substantiate undertaking a sustained inquiry into the psychological reality issue, all that must be established is that the cognitive conception is not "the only game in town." The dialectical situation does not require a decisive refutation; I aim for nothing more than a Scotch verdict. Having secured it, I will devote Chaps. 5–9 to showing that one or another grammar must be psychologically real *even if* Devitt's Linguistic Conception is correct.

Chapter 3
E-Language and I-Language

Abstract Chomsky claims that any theory of public “E-languages” will “surely have to presuppose grammars of I-languages.” Public languages are “more abstract” than I-languages, more “remote from mechanisms”. But can psychological mechanisms be described without reference (tacit or explicit) to social facts? I argue that public languages are indispensable to the study of language acquisition, as practiced by working psycholinguists. The data and explananda of acquisition theory are routinely couched in terms that make ineliminable reference to public languages, which serve as “targets” against which children’s successes and failures throughout development are measured. Though this does introduce a “normative-teleological” element into the science, it does not signal a move toward “prescriptive linguistics,” nor require an appeal to messy socio-political considerations. The normative-teleological element is innocuous, deriving from a theoretically motivated idealization of the child’s linguistic community. Next, I argue that the lack of precision in the individuation of public E-languages is just as much a feature of I-languages. Individuating I-languages requires settling unresolved issues about the competence/performance distinction, dialect mimicry, linguistic change, multilingualism, codeswitching, and cognitive disorders. It is not rational to insist on maximal precision in the individuation of either public E-languages or of I-languages at this stage of inquiry.

Keywords Cognitivism • Noam Chomsky • I-language • E-language • Public languages • Language acquisition • Descriptive vs. prescriptive linguistics • Socio-political considerations • Idealization • Individuation conditions for E-languages • Individuation conditions for I-languages • Dialect mimicry • Linguistic change • Bilingualism • Multilingualism • Codeswitching • Cognitive disorders

3.1 Introduction

In Chap. 1, I argued that abstractism—considered as a package of views concerning the ontology and methodology of linguistics—offers us a radically mistaken picture of the field, of its aspirations, and of language itself. I then went on to examine two alternative views—the dominant cognitivist position of Chomsky (1965, 1995), and the sophisticated nominalist rival developed by Devitt (2006a, b, 2008a, b, c).

D. Pereplyotchik, *Psychosyntax*, Philosophical Studies Series 129,
DOI 10.1007/978-3-319-60066-6_3

Having sketched the main lines of disagreement between them, my aim in this chapter and the next is to defuse several common arguments for cognitivism. My goal is not to refute the cognitive conception of linguistics, but only to show that the arguments *for* it are far from conclusive.

Devitt's Linguistic Conception commits him to the reality of public languages—something akin to what Chomsky has called E-language. He appeals, in this connection, to the notion of a convention, which has gained currency in philosophy after the pathbreaking work of Lewis (1969).[1] In this chapter, I begin by addressing some common concerns regarding the theoretical utility of the notion of E-language, as well as the difficulties that arise in giving identity conditions for E-languages. In Sect. 3.1, I show that E-languages play a useful role in theories of language acquisition. This bolsters Devitt's view that E-language plays a role in formal grammar construction in the straightforward sense that *that is what syntacticians are actually studying*. In Sect. 3.3, I argue that the identity conditions on E-languages are no less secure than the identity conditions on I-languages, contrary to what is often assumed. Along the way, I discuss some reasons to be skeptical of the arguments that Barber (2004), Collins (2008a, b), and Ludlow (2011) give in support of the claim that I-languages, rather than E-languages, are the objects of linguistic inquiry (Sect. 3.2).

3.2 Motivating the Study of "E-Language"

Devitt holds that the objects of linguistic theorizing are linguistic expressions, construed as those physical entities—inscriptions, acoustic blasts, and muscle movements—that are distinctive in having phonological, syntactic, and semantic properties in addition to their brute physical properties. This is the essence of his "Linguistic Conception". Collins (2007) takes Devitt's position to be committed to the existence of what Chomsky (1986) has called 'E-languages'. Devitt is, however, hesitant to commit himself to E-languages outright.

> On my view, a language is composed of the outputs of a linguistic competence, symbols that are governed by a system of linguistic structure rules. That is the reality of a language. And the task we have been contemplating, and that I wish to promote, is the study of the nature of this reality. This is not Chomsky's task (i), the study of the nature of the competence itself. Indeed, at first sight, the contemplated study may seem to be alien to Chomsky's enterprise. It may even seem to smack of studying an "E-language", of which Chomsky takes a dim view: "the concept [of an E-language] appears to play no role in the theory of language" (1986: 26); an E-language has "no corresponding real-world object" (p. 27). But it is not obvious that the outputs of linguistic competence fit Chomsky's description of an E-language. According to him an E-language is "externalized …in the sense that the con-

[1] More recently, in a self-conscious effort to undermine Chomsky's skepticism about the coherence or theoretical utility of the notion 'public language', Millikan (2004) has taken steps to develop an account of public-language conventions, making use of the resources afforded by her well known teleological framework. Devitt's view of such conventions has been criticized by Cain (2010). I will not take a stance here on whether Millikan's or Cain's arguments succeed.

> struct is understood independently of the properties of the mind/brain" (1986: 20). And it sometimes seems as if an E-language for Chomsky is essentially Platonic. In any case, the outputs I have identified, physical sentence tokens governed by a system of linguistic rules, are certainly not divorced from the mind/brain since they are the symbolic outputs of the mind/brain. In studying them our object of study is not the mind/brain, of course, but their linguistic properties are surely largely determined by the mind/brain. Finally, the theory of them is as much concerned with real-world objects as the theories of horseshoes, chess moves, bees' dances, and *wffs*. It is often convenient to talk of the objects posited by these theories as if they were types not tokens, as if they were abstract Platonic objects, but this need be nothing more than a manner of speaking: when the chips are down the objects are parts of the spatio-temporal physical world. (Devitt, 2006a, b: pp. 25–26)

Nevertheless, there is a clear case to be made for seeing Devitt as a friend of E-language. For one thing, Devitt is skeptical of the very reality of I-languages. More importantly, he commits himself (*Ignorance of Language*, Sect. 10.5) to the existence of public languages. Indeed, he argues that linguistic theory is concerned with *these*, rather than with I-languages; his "sixth major conclusion" in *Ignorance* is that "the primary concern in linguistics should not be with idiolects but with linguistic expressions that share meanings in idiolects" (p. 183). So, while the entities that Devitt is concerned with are not "independent of the mind/brain," nor "Platonic," it is safe to say that they are E-languages nonetheless.[2]

As Devitt notes in the passage quoted above, Chomsky has strenuously objected to the coherence and the utility of the notion of E-language. Since the introduction of the E-language/I-language distinction in *Knowledge of Language* (1986), Chomsky has repeatedly claimed that defining a notion of E-language is not only pointless, but that he knows of no successful examples.[3] Collins (2008a) echoes these concerns in the following passage:

> There is a philosophical cottage industry attempting to show that there is a coherent notion of public language or at least a language that is independent of speaker/hearers' brain states. It might be of some curious interest were such efforts to prove successful, but their relevance to linguistics is wholly obscure, for no one has made any serious suggestion how such a public notion might fruitfully enter into our understanding of the empirical questions of the structure and development of linguistic competence. A metaphysics of our commonsense concept of a language challenges no thesis of linguistics. (p. 20)

Collins and Chomsky are, I think, perfectly right to point out that the friend of E-language owes an account of both the utility and the content of this notion. There are problems, however, with the way that Collins couches this demand in the passage quoted above. For, he limits in advance the interesting questions for linguistics to only those that have to do with "the structure and development of linguistic competence." Following Chomsky, he suggests that any other concern belongs to mere

---

[2] Devitt also rejects another aspect of Chomsky's characterization of E-language. Barber (2004) writes: "Folk languages could be regarded as E-languages, but Chomsky does not have the folk in mind so much as philosophers who are explicitly or implicitly behaviorist in their assumptions." As Devitt (2006a: Chap. 5) makes clear, he is no fan of behaviorism.

[3] One might object here to Chomsky's frequent use of the word 'define' in this connection. As Chomsky is surely well aware, the notions employed in empirical pursuits are rarely if ever *defined*, strictly speaking. Rather, they emerge from fruitful theory construction. Chomsky's own characterizations of I-language are not, I take it, definitions in the strict sense.

folk theory. This begs the question against those who, like Devitt, argue that routine practice in linguistics is concerned precisely with dialects, public languages, and families thereof (e.g., the Romance, Slavic, and Germanic families).

> [Public language] classifications seem to be useful in linguistics, for linguistics books and articles are replete with them. The first few pages of Haegeman's GB textbook (1994) have many uses of 'English' to classify shared meanings. Then 'English' is compared with 'Italian'; for example, 'In Italian a subject of a subordinate clause can be moved to the main clause domain across the overt conjunction *che*, corresponding to *that*; in English this is not possible' (p. 20). And then with Spanish and French (p. 23). And so on throughout the book. (*Ignorance of Language*, p. 183, fn 23)

Moreover, Devitt points out that a theory of public languages *does* in fact contribute to our understanding of "the structure and development of linguistic competence." On his view, linguistic competence is an ability to produce and comprehend expressions in a public language. Hence, we can only understand that competence if we have a prior grip on the nature of the language itself.[4]

For Devitt, the point of clarifying the notion of an E-language—or, better, the nature of E-languages—is to show that *this* is what working linguists are actually studying, *pace* Chomsky's ideology, and despite their own protests to the contrary. He writes:

> Turn next to Chomsky's claim [that] conventions have no interesting bearing on the theory of meaning or knowledge of language. Chomsky thinks that this also should be a truism (1996: 48). Yet, if it were true it would seem to be at odds with what most linguists are actually doing. For, what they are mostly doing is theorizing about the largely conventional syntactic and semantic properties of expressions. And they are right to be doing so, in my view. Conventional meaning is important to theory in at least four ways. (p. 181)

Devitt goes on to respond to Chomsky's demand for a specification of the alleged role that E-language plays in empirical theorizing. His answers have mostly to do with conventional shared meanings—the roles they play in explaining successful communication and in the ascription of propositional contents to the internal states that explain our behavior.

> The members of any group that share a meaning of one linguistic expression tend to share meanings of a vast number of others and it is convenient, on the basis of this, to follow the custom of classifying sets of these expressions with shared meanings as English, Spanish, and so on. The classification is bound to be a bit vague but no more so than many scientifically appropriate ones. And such classifications seem to be useful in linguistics, for linguistics books and articles are replete with them. Are these classifications mere manners of speaking that can be paraphrased away when the serious linguistic work is to be done? I

[4] There is another problem with the quoted passage from Collins. His rhetoric suggests that questions of ontology are not empirical, contrary to the conclusions we drew in Chap. 1. It is important, I think, to remember that there is no one particular kind of inferential consequence that a thesis must have in order to count as empirical. The fact that some thesis about the metaphysics of language does not, by itself, yield predictions about syntactic structure or the acquisition process does not mean that this thesis has no inferential consequences whatsoever. It very well might have significant consequences—at present or some future time—when conjoined with other theoretical commitments. In any event, making and defending claims about the history and sociology of a field like linguistics is surely an empirical enterprise.

> think not: they are necessary for the linguist to identify what she is talking about, to identify the subject matter. For, the subject matter is the shared meanings and syntactic properties of linguistic expressions in a certain group of people. (Devitt, 2006a: p. 183)

For reasons that go beyond the scope of the present discussion, I am not convinced that we must appeal to public meanings in explaining communication and the ascription of propositional contents. Nevertheless, Devitt does mention one explanatory project for which reference to conventions and public languages seems to me to be indispensable—viz., the project of giving an account of language acquisition. He writes:

> [T]he syntactic differences between public languages show that much syntax is not innate. These differences are captured, on the received Chomskyan view, by different settings of "parametric values". Very occasionally an idiolect's parameter settings may be eccentric but almost always they will be conventional. Thus most people in the USA participate in parameter-setting conventions that lead them to speak an SVO language; most people in Japan participate in parameter-setting conventions that lead them to speak an SOV language. (Devitt, 2006a: p. 181)

In what follows, I expand on this claim, making clear its implications for how we should regard both acquisition theory and the famous poverty-of-stimulus arguments that Chomsky has popularized.[5]

3.2.1 The Role of E-Language in Acquisition Theory

In his original discussion of the E-language/I-language distinction, Chomsky (1986) claimed that certain aspects of the folk conception of language would have to be abandoned as the study of language goes scientific. One of these is what he called

[5] Devitt (2006b) makes reference specifically to parameter-setting accounts of language acquisition in the course of motivating a commitment to conventionally instituted public languages. "I argue that language acquisition provides one reason for believing in SLEs [Standard Linguistic Entities]: the best explanation of the setting of parametric values in the typical member of a linguistic community is that she comes to participate in the parametric conventions of the community. … My view is that *the best explanation* of our *actual* acquisition adverts to conventions. So that is a good reason to believe that there are conventions. The child sets his parameters in a certain way because it regularly experiences others who have set them in that way. Conventions explain those regularities. Thus, "conventions explain the child's experiences which explain its settings. And the conventions explain why all the children in that community usually come to have the same setting" (2006b: 582). While I myself am happy to endorse parameter-setting accounts of acquisition, I am not sure that such accounts are compatible with Devitt's rejection of the psychological reality of syntactic principles. As it is ordinarily conceived, setting parameters is a psychological process that involves mentally representing the rules or principles of UG. Perhaps Devitt intends his talk of parameter setting as committing him only to the claim that children "come to participate in the parametric conventions of the community." Still, the going accounts of the psychological processes that lead children to their eventual success in participating in such conventions traffic heavily in mentally represented rules and principles. See Devitt (2006a: Chap. 12) for more discussion of this topic.

the "normative-teleological" aspect of the folk concept, which arises in connection with considerations of language acquisition, in both children and adults.

> The commonsense notion [of language] also has a normative-teleological element that is eliminated from scientific approaches. I do not refer here to prescriptive grammar, but to something else. Consider the way we describe a child or a foreigner learning English. We have no way of referring directly to what the person knows. It is not English, nor is it some other language that resembles English. We do not, for example, say that a person has a perfect knowledge of some language L, similar to English but still different from it. What we say is that the child or foreigner has a "partial knowledge of English," or is "on his or her way" toward acquiring knowledge of English, and if they reach that goal, they will then know English. Whether or not a coherent account can be given of this aspect of commonsense terminology, it does not seem to be one that has any role in an eventual science of language. I will follow standard practice in disregarding these aspects of the commonsense notions of language and the associated notions of rule-following and so forth, although the departure should be noted and one may ask whether it is entirely innocent. (Chomsky 1986: p. 16)

A glance at the literature on language acquisition, however, reveals that what Chomsky calls the "commonsense notion" actually plays a major and seemingly indispensable role in studies of language acquisition. Consider, for instance, the following passages, which are taken from a recent and authoritative textbook by Maria Teresa Guasti, a proponent of the Chomskyan paradigm. I have underlined what I take to be instances of "normative-teleological" vocabulary, as well as references to public languages.

> Human language acquisition is an astonishing process. Let us consider what these children have accomplished in about 3 years. Although their language may still not be perfect, they put words in the correct order. Nina produces quite a complex sentence, putting the complements in the right order (first the direct object and then the prepositional complement) and applying wanna-contraction. Scrambling complements is possible in Italian, and Diana shows that she can take advantage of this option, by putting the prepositional complement (*a Luca* 'to Luca') before the direct object (la *bambola* 'the doll'). Eve places the adjective *funny* before the noun clown, as required in English, while Diana places the adjective *lossi* (*rossi*) 'red' after the noun *capelli* 'hair', since she speaks Italian. (Guasti 2002: p. 2, underlined emphases added)

> It is well known that babies born into a multilingual environment can easily pick up more than one language. This shows that infants can distinguish not only between utterances in the language of their environment and utterances in other languages, but also between utterances in two or more languages spoken around them. Learning a language requires discovering the rules of that language—for example, how words are ordered in clauses, and how questions are formed. If infants could not distinguish between utterances from different languages, they might make bizarre conjectures concerning the properties of what they hear. Without this ability, how could a child hearing sentences from, say, French and Spanish ever figure out the properties of French? (Guasti 2002: p. 24, underlined emphases added)

For the purposes of the present discussion, what is striking about these passages is that they are replete with what appear to be clear and direct references to public languages, their properties, the expressions that belong to them, the rules that govern them, their presence in the child's environment, and the degrees to which various children's usages are "correct," "right," or even "perfect" with respect to them.

Chomsky and his followers would surely resist this interpretation, claiming that these remarks belong to the informal presentation of a theory, not in its serious development—a distinction that Chomsky wields with worrying frequency.[6] But this is difficult to square with the fact that Guasti (2002) *continues* to make use of the "normative-teleological" terminology throughout the text, even in the context of what appear to be quite serious statements of empirical generalizations. Consider, for instance, the following passages, taken from those chapters of Guasti (2002) where she reports a variety of experimental and corpus data. Once again, I have underlined the phrases that are remarkable in the present context.

> Italian, Spanish-, and Catalan-speaking children use singular agreement morphemes with the appropriate subject. … Errors are rare and mostly found with plural subjects. (pp. 120-121)

> Agreement errors are rare, about 3-4% in early Italian (Guasti 1993/1994; Pizzuto and Caselli 1992), 1.72% in early Catalan and Spanish (Torrens 1995). These rare errors mostly consist of using a singular third person morpheme with a plural subject or a third person morpheme with a first person subject. These findings have been replicated for early German: Poeppel and Wexler (1993) found that their child subject, Andreas (2;1), used the first and third singular agreement morphemes accurately; the three plural morphemes and the second singular morpheme were rare or absent in his speech; and errors were rare. (p. 121)

> Italian learners whose speech has been studied use postverbal subjects in about 30% of their sentences with overt subjects and make no agreement errors… (p. 125)

> examples … found in the CHILDES database show that children do not make mistakes with coordinate subjects. (p. 125)

Guasti is a card-carrying Chomskyan linguist, so it cannot be that her use of the underlined phrases signals a rejection of Chomsky's ideology. One might, of course, resort to the claim that her usage in these passages (and countless others throughout the book) is merely "informal" or perhaps even "sloppy."[7] Putting aside the fact that such claims run a serious risk of insulting Guasti and many other acquisition theorists, one simply finds no warrant for them in the actual text. Indeed, consider the kinds of empirical findings that animate inquiry in acquisition theory. As illustrated

---

[6] See, e.g., his exchange with Georges Rey in Barber (2000).

[7] Collins (2008a) avoids this charge by inserting occasional scare quotes around the offending terms: "Children exposed to a degraded language, such as a pidgin or late-acquired sign, do not learn the 'errors' but develop a language that is consistent in its syntax and morphology. … The development of language is also marked by very little 'error' outside of narrow parameters. Familiarly, children have trouble learning irregularity in number and past tense and they make other notable errors, such as leaving wh-words in medial positions. Viewed from the perspective of what children could get wrong, however, their competence is remarkably consistent with adult understanding. Indeed, adults often stumble over irregularity. Children's competence is so error free that it has been plausibly suggested (Crain and Thornton, 1998) that children's competence matches the adult speakers; the 'errors' there are performance based" (pp. 105–6, underlined emphasis added)." Plainly, though, this typographical change does not settle the substantive issues. It remains unclear whether Collins can avoid commitment to the existence of a public language—an "adult understanding"—to which the child's idiolect might conform or fail to conform.

in the passages quoted above, these typically concern quantitatively described patterns of error—e.g., the percentage of agreement errors that children make in Italian (apparently 3–4%). It is difficult to see how such findings might be "formally recast" as claims about the relation between the child's usage and some specific I-language. Against *whose* I-language would a child's usage be quantitatively compared?

It's true, of course, that the acquisition theorist can claim to be comparing the child's grammar to an *idealized* I-language. But it's not at all clear what import the appeal to idealization has in this context. If we press on the notion of an *idealized* I-language, we find that it amounts to no more than a consistent setting of parameters (assuming a P&P grammar). But there are many such settings, most of which fail to match the language of the community. Thus, reference to the grammar of a public language seems unavoidable in singling out the language that the theorist identifies as the child's "target grammar." If there is an idealization in the vicinity, it is one that abstracts away from the variation *within the community*—i.e., the differences between individual speakers—and yields an *idealized speech community* as the theoretical object of interest.

What this shows, I think, is that the opponent of E-languages owes an explicit account of the child's goal throughout the acquisition process—a formal characterization of what is typically called "the target language," "the ambient language," or (as above) "the language of [the child's] environment." Providing such an account would go a long way toward making sense of what acquisition theorists mean when they claim that a child has made a "mistake" in the course of acquisition, or, equally, that no mistake was made. I am not convinced that an explicit, formal account of this would make no mention of E-language. Whether it would is an open empirical question.[8]

It is important to note, moreover, that the "normative-teleological" terminology is ubiquitous in the various statements of the poverty-of-stimulus argument for the innateness of UG. Consider the following two passages:

> [C]hild errors are rare *tout court*. This would appear to demonstrate that children are making 'decisions' about <u>the target language</u> independent of data. If children were being instructed by the data, one would find a preponderance of early mistakes that would gradually decrease as more disconfirming evidence is accumulated. (Collins 2008a: p. 105)

> Although <u>children make "errors,"</u> they do not make certain <u>errors</u> that would be expected if they generalized from the linguistic input. For example, although children hear sentences like *Who do you wanna invite?* and *Who do you wanna see?*, they do not generalize from these to <u>impossible English sentences</u> like **Who do you wanna come?*; although this generalization would seem reasonable, children never say such sentences. (Guasti 2002: p. 5)

Collins' use of the term 'target grammar' carries the implication that there is a grammar external to the child at the time of acquisition, which the child is

[8] I am grateful to Daniel Harris for pointing out that my main purpose in this chapter—i.e., to argue that Chomsky's cognitivist conception is not the only game in town—does *not* require the strong claim that public E-languages actually play a role in acquisition theory. My conclusion is a weaker one—viz., that we do not have overwhelming grounds for supposing that E-languages will *not* play any such role.

struggling to grasp. Given that he is at pains to reject this implication, it is important that he be able to state the poverty of stimulus argument without using this term or any of its cognates. Again, I am not persuaded that this is feasible. And if it is not, then we have here a clear motivation, supplied by a respectable, up-and-running, empirical research program, for investigating the nature of E-languages. Combined with Devitt's arguments to the same effect, I believe the case for this project is, if not airtight, at the very least non-trivial.

It is easy to see what the outlines of a methodology for such a project might be. We begin with the idea that the child's goal in acquiring a language is to obtain the capacity for fluid linguistic interaction with members of her community.[9] One metric along which the fluidity of communication can be measured—though by no means the most important metric—has to do with the extent to which the grammar of the child's idiolect deviates from the grammar of the "ambient language" in her community. The actual process of constructing such a grammar would proceed by first collecting data from the adult members of that community. Suppose, for instance, that the community initially consists of only two competent adult speakers, A and B. Now, if A and B have a child, C, then the set of their utterances $\{\{u(\mathrm{A})\} \cup \{u(\mathrm{B})\}\} = \{u(\mathrm{A, B})\}$ exhausts C's primary linguistic data (PLD). Now take Γ_{AB} to be the grammar, or set of grammars $\{\gamma_1, \gamma_2, \ldots \gamma_n\}$, that best covers $\{u(\mathrm{A, B})\}$. The abovementioned metric involves making comparisons between Γ_{AB} and the child's grammar Γ_{C}. The extent to which there is some well-defined divergence,[10] we can say that Γ_{C} is a *misrepresentation of the target language*, and, consequently, that C is in *error*.

We can generalize this strategy to the case in which the child is exposed to speech from more than two speakers. Let the number of speakers be n. Now consider the set of utterances $\{u(\mathrm{A, B, \ldots S}_n)\}$ that are perceived by the language-acquirer from *all* of the speakers with whom he or she has been in contact in the course of the acquisition process.[11] A corresponding generalization takes us from Γ_{AB} to $\Gamma_{\mathrm{AB\ldots S}n}$, i.e., the grammar shared by the speakers from whom a person acquires his or her grammar. If we wish, we can assign weights to any member or subset of $\{u(\mathrm{A, B, \ldots S}_n)\}$, to reflect factors such as the amount of attention that the language-learner paid to the utterances, or the extent to which the learner understood those utterances (if these can be independently measured).

It is of course possible—indeed, likely—that no grammar will adequately capture the set of utterances $\{u(\mathrm{A, B, \ldots S}_n)\}$. For instance, speakers F and G might

[9] Chomsky famously denies that the purpose of language is communication, but we can put this puzzling view to one side for the moment.

[10] There are many conceivable ways of formally characterizing this sort of divergence. No general characterization can be given for all possible types of grammar. A fan of context-free systems might count how many context-free rules exist in Γ_{AB} but not Γ_{C}, or vice versa, whereas a proponent of old-school transformational grammars might prefer to count the number of transformations instead. Likewise, if the Γ_is emerge from the P&P approach, then perhaps it will be most natural to compare parameter settings.

[11] For simplicity, I treat only utterances, though, in principle, we can consider metalinguistic judgments as well.

have divergent linguistic practices. The linguist has a number of options in dealing with this phenomenon. She can assign weights to particular speakers, and count weigh more heavily the utterances (or metalinguistic judgments) of those people who were most prevalent in providing the child with PLD during the acquisition process. Alternatively, she can appeal to the notion of *degrees* of grammaticality—a notion that was initially proposed in Chomsky (1961/1964), and which has an analogue in the P&P approach.[12] In the limit, the differences between members of the community might be so dramatic that it would be more expedient to treat them as speaking different languages—i.e., drawing on largely disjoint stocks of phones, phonemes, or lexical items. In such a case, we can consider the child as belonging to two distinct language communities.

The linguist may find that explicitly accounting for variation in the community within her theory significantly hampers her research. In such a case, she has the option of idealizing away from such variation, as the linguists in the structuralist tradition found it expedient to do. Chomsky (1986) describes this strategy in the following passage:

> Modern linguistics commonly … consider[s] an idealized "speech community" that is internally consistent in its linguistic practice. For Leonard Bloomfield, for example, language is "the totality of utterances that can be made in a speech community, regarded as homogenous" (Bloomfield, 1928/1957). In other scientific approaches, the same assumption enters in one or another form, explicitly or tacitly, in the identification of the object of inquiry… Of course, it is understood that speech communities in the Bloomfieldian sense—that is, collections of individuals with the same speech behavior—do not exist in the real world. Each individual has acquired a language in the course of complex social interactions with people who vary in the ways in which they speak and interpret what they hear and the internal representations that underlie their use of language. Structural linguistics abstracted from these facts in its attempts at theory construction… (1986: p. 16)

[12] Chomsky (1988: p. 560) argues that the notion of E-language cannot be defined in any principled way because this would require deciding how to treat strings like 'The child seems sleeping'. For, on the one hand, the string should be included in the E-language on account of its having a clear meaning, but, on the other hand, any attempt to include it in the E-language will inevitably cast the net too wide, potentially bringing into the fold *all* noises and inscriptions. Chomsky's argument overlooks his own work on degrees of grammaticality. The E-language theorist can associate a language, L, with a *multi-tiered* set-theoretic structure that takes such degrees into account. Consider, for instance, the fact that 'The child seems sleeping' only has a clear meaning in English because of its relation to 'The child seems to be sleeping'. This asymmetry warrants a distinction between a "core" set of utterances, and a "peripheral" set. The strings in a core set are the fully grammatical sentences. By contrast, 'The child seems sleeping' would be in the less-than-grammatical set. Other strings would be in the fully ungrammatical set, and still others wouldn't be in any of the sets, perhaps because they aren't made up of elements in the lexicon. (One must, of course, settle on what's in the lexicon—a non-trivial issue, but one that arises for *all* theories of language.) The distinctions between such sets can be made even more fine-grained by reference to a P&P grammar, where the degree of a string's grammaticality can be seen as a function of how many syntactic principles it fails to satisfy. See Berwick (1991b) for an extended discussion of the latter point.

Comparing the structuralist approach with his own, Chomsky makes clear that analogous idealizations must be made in order to facilitate inquiry into the structure of I-languages.

> [Chomskyan linguistics] also abstracts from these facts in posing [the cognitivist] questions [What constitutes knowledge of language, and how is this knowledge acquired and used?], considering only the case of a person presented with uniform experience in an ideal Bloomfieldian speech community with no dialect diversity and no variation among speakers. … We should also make note of a more subtle theory-internal assumption: The language of the hypothesized speech community, apart from being uniform, is taken to be a "pure" instance of UG in a sense that must be made precise…. We exclude, for example, a speech community of uniform speakers, each of whom speaks a mixture of Russian and French (say, an idealized version of nineteenth-century Russian aristocracy). The language of such a speech community would not represent a single set of choices among the options permitted by UG, but rather would include "contradictory" choices for certain of these options. (*ibid.*)

Here I want to stress a point that is commonly overlooked in philosophical discussions of generative linguistics. Given that that the study of *both* E-language *and* I-language requires idealizing away from variations—within a community and within a speaker's idiolect—a theorist who champions the cognitivist conception cannot argue against the neo-structuralist or nominalist approach on the grounds that it requires making such idealizations. Nor can he substantiate the charge that the objects of his opponent's preferred mode of inquiry are "unreal" or "fictional." The two approaches are entirely on a par in these respects.[13]

Note also that the conception of E-language that emerges from the approach outlined above is thoroughly *apolitical*. That's to say that on *this* approach, the concerns of the practicing linguist dictate a delineation of (idealized) language communities that makes no appeal whatsoever to geographical boundaries, ethnic or religious affiliations, power relations among groups, historical conflicts and alignments, and so forth. As Devitt (2006a) points out, "a 'sociopolitical dimension' does occasionally intrude into such classifications but the intrusion can be resisted by linguists; for example, a linguist may think that, for almost all expressions, there is no theoretical point to the politically inspired division of Serbo-Croatian into Serbian and Croatian and can simply refuse to go along..." (p. 184).

Nor is the project in any way intended to contribute to the agenda of *prescriptive* linguistics.[14] The construction of a descriptive grammar is a worthy and noble enterprise, and linguists in the generative tradition are—I think *rightly*—unwilling to be seen as engaging in the construction of a prescriptive grammar. No good can come

[13] Thus, consider the following remark from Collins (2007): "[L]et us grant that linguistic properties *might* arise from a relation between the mind/brain and external stuff, but all the action is at one end of the relation; the other end is noisy, variable and does not submit to independent inquiry" (p. 420). The reply is that *all* natural phenomena are variable and noisy. Idealization is just as necessary in the study of public languages as it is in the study of the mind/brain. The cognitive conception of linguistics derives no advantage from these considerations.

[14] To fix the terminology, a *descriptive* grammar tells us about what *actual* usage is like, whereas the prescriptivist seeks to impose normative "standards of usage" that invariably deviate from the actual linguistic conventions in communities and the systematic behaviors of actual speakers.

from confusing their theoretical goals with the goals of prescriptive grammarians. I want to urge, however, that it is *equally* erroneous to conflate prescriptive grammar with the study of public languages, or E-languages. It is important to see that taking idiolects or I-languages as the objects of inquiry is *not* the only way to avoid doing prescriptive grammar. As I argued above, one can construct a descriptive grammar for a set of utterances that have been produced by multiple people. And choosing which group of people to study need not involve taking into account messy socio-economic or political issues. Theory-internal considerations can guide the choice. These considerations include, *inter alia*, the degree to which the members of a group are disposed (i) to produce strings that can plausibly be seen as having the same syntactic structures, (ii) to comprehend linguistic input in similar ways, (iii) to make the same metalinguistic judgments, and (iv) to communicate in a way that both they and the linguist judge to be maximally fluid.

It must be admitted that the approach that I have recommended in the pages above is rough around the edges, to put it mildly. Subtle issues arise about how much agreement is required in competent speakers' intuitions, to what extent—and in what respects—their linguistic output must be similar, to what degree their comprehension of the same utterances must coincide, how to gauge the success or fluidity of their linguistic communication, and so forth. But while constructing an apolitical notion of public languages is difficult, we must not confuse difficulty with impossibility. Many linguists are confident that we can gain a great deal of insight into comparatively daunting matters—the evolution of language, for instance. It strikes me as simply irrational to have confidence in *that* sort of enterprise but no hope at all for an apolitical construction of a grammar for a set of utterances produced by two or more native English speakers. Accordingly, I am confident that increasingly precise characterizations of E-languages can be filled in as theoretical and experimental advances are made.

In Sect. 3.3, I will argue that the project of individuating I-languages is replete with similar promissory notes. It follows that the lack of precision in our notion of E-language is just as much a feature of our conception of I-languages. Accordingly, one would do well to ask what degree of precision it's really necessary to achieve in order to get on with fruitful theorizing. I suspect that neither in the case of E-languages nor I-languages will any rational theorist insist on *maximal* precision at the outset of inquiry.

3.3 Do Linguists Really Study I-Languages?

Barber (2004) gives voice to some of the most popular arguments against viewing linguistics as a study of E-language, and for avoiding that sort of project in practice. His first argument opens with the innocent-sounding premise that linguistics is a science—a claim that I take to be plainly true, occasional protests notwithstanding (Chap. 1). Barber goes on, however, to claim that genuine sciences do not make normative pronouncements and that individuating public languages involves making normative ("prescriptive") claims. It follows, then, that individuating public languages cannot be part of a genuine scientific practice and, *a fortiori*, cannot be part of linguistics. As a separate argument—though plainly continuous with the first—Barber adds that individuating public languages involves reference to social and political factors. He goes on to take for granted the claim that "any object whose individuation conditions are determined by socio-political factors should not figure in the fundamental ontology of science."

There are several mistakes in this line of reasoning. To begin with, Barber is relying heavily on a sharp distinction between the normative and the descriptive. Putnam (2004) has argued compellingly that this distinction is both mythical and pernicious; whatever currency it had in the heyday of positivist philosophy, we should cast a skeptical eye on it today. Of course, Barber might protest that, by 'normative claim', all he meant was 'a claim that makes ineliminable reference to socio-cultural constructs'. But, then, to say that no science can make normative claims in *this* sense is to preclude any scientific study of social-level phenomena. It would follow that anthropology, sociology, economics, political science, history, and many other fields are all forever barred from enjoying the status of genuine sciences, on the *a priori* ground that their ontologies are, in some sense, second-rate. It is hard to take seriously a conception of science that animates such a claim.[15]

Consider now what Barber (2004) says about the objects of linguistic theory.

> The operating notion of a language is instead that of an idiolect, often in Chomsky's developed sense of this term (an I-language). So although a label like 'Hungarian' will inevitably be used in practice, it is normally intended to be thought of as convenient shorthand for 'The idiolect of some arbitrary but typical inhabitant of present-day Hungary'.

As formulated here, Barber's claim simply cannot be correct. No linguist seriously thinks that Hungarian (or any other language) is, or need be, confined to a particular geographical region.[16] More importantly, it's not at all clear *why* "the idiolect of some arbitrary but typical inhabitant of present-day Hungary" would be any better an object of study than simply *Hungarian*. Barber's answer has to do with his claim, criticized above, that individuating public languages involves making normative claims and ineliminable references to socio-cultural constructs. But it seems that Barber has run afoul of his own constraint here, for his phrase "the idiolect of some arbitrary but typical inhabitant of present-day Hungary" *likewise* makes inelim-

[15] I hasten to point out that Chomsky does *not* fall prey to this error. He writes: "Internalist biolinguistic inquiry does not, of course, question the legitimacy of other approaches to language, any more than internalist inquiry into bee communication invalidates the study of how the relevant internal organization of bees enters into their social structure. The investigations do not conflict; they are mutually supportive. In the case of humans, though not of other organisms, the issues are subject to controversy, often impassioned, and needless" (Chomsky, 2001: p. 41).

[16] Ludlow (2011) makes a similar mistake in taking this strange way of individuating languages to be a feature of folk theory. Of the folk notion, he writes: "Typically, the question of who counts as speaking a particular language like German is determined more by political boundaries than facts about linguistic properties…" (p. 44). It is difficult to believe that the folk would insist that a monolingual German speaker who has, since birth, resided exclusively in France, does not *really* speak German.

inable reference to socio-cultural constructs—viz., *present-day Hungary*. We must conclude, I think, that Barber fails to adequately characterize the objects of linguistic theorizing, both by his own criteria and by independent standards. How might we do better?

For dramatic effect, suppose that we were to conduct an informal survey of linguists working in the generative tradition, in an effort to determine whether they are studying public languages or idiolects. Given the prevalence of Chomsky's ideology, which is often presented to students of linguistics alongside his truly impressive results in syntactic theory, we would expect to find that many of them will claim that they are studying I-languages. But now suppose that we continue the survey by asking the linguists, "*Whose* I-language are you studying?" Some will have little to say in response to such a bizarre question. Others will give a response that makes reference to socio-cultural constructs—perhaps the very same ones that Barber appealed to in the passage quoted above.

Such responses, which can take any of a number of forms, all leave open the possibility that the linguists who make them were, as a matter of fact, *mistaken* when they initially claimed to be studying I-languages. It may be, after all, that they study E-languages (of one stripe or another), but that they are not aware that this is what they are doing, perhaps because they have not reflected on the matter, or because they've accepted Chomsky's ideological remarks on this issue without considering the alternatives. There is nothing incoherent about this. It is both common and, within limits, even rational for scientists to pursue empirical questions without reflecting, in any great depth, on the ontological status of their field's subject matter—or by simply taking for granted the received view among their research community (Kuhn 1962/1996).

A more philosophically inclined linguist might respond as follows: Ultimately, I am studying *my own* I-language. But, while this ingenious response has the virtue of enriching the meaning of the 'I' in 'I-language', it is by no means clear that it is an adequate response, as an extension of our hypothetical survey would reveal. Consider, for instance, expanding the survey to include the following questions: Do you plan to publish your findings? If so, then why do you suppose that *your* I-language is important for anyone else to learn about? Do you ever ask other people for *their* intuitions about various linguistic constructions, after having articulated your own intuitions about them? If so, *why*? What purpose does it serve to know what data other people's I-languages generate?

There are quite natural ways of answering these questions, which do not bode well for the standard Chomskyan line. For instance, it's plausible to say that the purpose of publishing one's results, even those that the linguist takes to be about his or her own idiolect, is to fuel the project of formulating and confirming generalizations that can be projected across a broader population. Indeed, this sort of generalization is a routine practice among linguists.[17] Similarly, asking multiple informants

[17] Cf. Devitt (2006a, b): "[Public language] classifications seem to be useful in linguistics, for linguistics books and articles are replete with them. The first few pages of Haegeman's GB textbook (1994) have many uses of 'English' to classify shared meanings. Then 'English' is compared with

for their intuitions regarding some linguistic construction can easily be seen as a way of gathering information about the *commonalities* between speakers. If so, then the search for generalizations across speakers seems to be the linguist's primary motivation. On the strength of these considerations, one is led to suspect that what the linguist is ultimately interested in is the nature of the public language—perhaps a dialect—not the I-language of an arbitrary individual speaker.[18] Moreover, it should be noted that consensus among multiple informants—particularly among professional linguists[19]—is a sign (though by no means the only nor the best sign) that their judgments reflect the nature of the language, rather than being a species of performance error. For this reason, even a linguist who takes herself to be studying linguistic competence can benefit from paying attention first to the character of a public language. Doing so can help her to draw a rough-and-ready distinction between phenomena that are indicative of competence and those that reflect only the vicissitudes of performance. Needless to say, the friends of I-language may well have alternative, cogent responses to the hypothetical survey questions posed above—responses that accord with their ideology. But this is by no means obvious and remains to be shown.

In connection with this issue, consider the following scenario, described by Eddington in his 2008 Presidential Address:

> It is unfortunate that many linguistic analyses do not submit their hypotheses to experimental test. Instead, they make the critical error of elevating a hypothesis about a phenomenon to the status of an explanation of the phenomenon (Black and Chiat 1981:48, Higginbotham 1991:555, Itkonen 1978:220-221, Ohala 1990:159, Sampson 2001:124, Yngve 1996). Chomsky provides an example of this common fallacy. He claimed that 'perform' may only be followed by a count noun, never by a mass noun, which would render 'perform labor'

'Italian'; for example, 'In Italian a subject of a subordinate clause can be moved to the main clause domain across the overt conjunction *che*, corresponding to *that*; in English this is not possible' (p. 20). And then with Spanish and French (p. 23). And so on throughout the book." (p. 183, fn. 23)

[18] Cf. Devitt (2006a): "There is indeed something a little paradoxical about denying the frequency of linguistic conventions. The linguistic method of consulting the intuitions of linguists and other speakers to discover facts about a language … presupposes masses of conventional regularities among them, even while allowing for some differences in idiolects. Books are written and papers are given about expressions in this or that language, all of which presuppose a great deal of regularity in usage among speakers" (p. 181).

[19] Devitt (2006a) argues that the metalinguistic intuitions of professional linguists are more informed, and hence more reliable, than those elicited from untutored speakers. This claim has been challenged by Culbertson and Gross (2009). The challenge has spawned a fruitful debate; see Devitt (2010a) for a reply, and Culbertson and Gross (2011) for a follow-up. I do not wish to embroil myself in these issues. The claim in the main text is merely that linguists are more likely to have learned about the varieties of performance error and hence more reliable than non-linguists in evaluating whether they are subject to it in some particular case. This is especially true when the judgments are subtle or fine-grained, as they are in the case of, say, degrees of subjacency violation. Though much of contemporary syntactic theory is based on stable and widely shared judgments (see Sprouse and Almeida [2013] for details), there are isolated cases on which the judgments are quite subtle or fine-grained. Occasionally, such cases turn out to be pivotal for settling major theoretical disputes. As such, they deserve further exploration by all parties to the debate over the relative merits of tutored and untutored linguistic intuitions.

> incorrect in English (Hill 1962:29). When pressed for evidence that his hypothesis was correct, Chomsky merely responded that he was a native speaker of English. Chomsky's intuition about the possible predicates of 'perform' resulted in a hypothesis. Hypotheses are indeed borne of intuition. However, his intuition at the same time provided him with the supporting evidence for the hypothesis; he saw no need to consult a corpus to confirm it. [fn.2: A quick search on Google yields thousands of cases of 'perform labor'. Sampson (2001) discusses a similar case in which personal introspection led to the idea that central embedding was impossible, when many cases are found in actual language data.]" (p. 8).

For our purposes, the important aspect of this anecdote is that Chomsky replied to the challenge by pointing out that that he is a native speaker of English. What, we might ask, did he mean to convey with his remark? One interpretive possibility that we should reject immediately is that Chomsky meant that he has his own I-language, and no one could refute him about *that*. This interpretation conflicts directly with the fact that Chomsky takes great pains to emphasize that competent speakers of a language can be wrong about their own language and that introspection is not a good (let alone an optimal) tool for the construction of linguistic theory (Chomsky 1965: Ch. 1.) Charity plainly dictates that we see Chomsky as taking his generalization about the distribution of count nouns and mass nouns to be open to empirical disconfirmation. How might such disconfirmation proceed? Eddington takes the results of a Google search—a very crude sort of corpus analysis—to be evidence against Chomsky's generalization. There are two ways in which such data might play that role, corresponding to two views of the nature of linguistic theorizing.

On one view, Chomsky was offering a generalization about only one person's I-language—his own. In this case, the results of Eddington's crude corpus analysis would disconfirm the Chomsky's generalization only in the event that Chomsky, himself, looked at them and saw that he perfectly well understands what is meant by 'perform labor' and that he finds the phrase either fully acceptable or fully grammatical. The opposing view, by contrast, has it that Chomsky was offering a generalization about a *dialect of English*—possibly, though not necessarily, his own. In this case, the results of Eddington's crude corpus analysis would refute the generalization outright, regardless of Chomsky's intuitive judgments—indeed, regardless of whether he ever makes any. Now, we can ask: What's more plausible? That Chomsky's theories of syntax have always been about—and *only* about—his own I-language, or that the theories are about the dialect of some reasonably large community of competent speakers? While it seems to me that the latter position is more probably true, I cannot claim to have shown this. Again, all I wish to establish here is that the case for the opposite view is not immediately obvious and remains to be made.

It may be objected that Chomsky's insistence on the study of I-language does not carry a commitment to the claim that linguists study *some particular speaker's* I-language. Rather, the insistence on studying I-language should be taken as a commitment to either (i) the study of distinctively *psychological* states and processes or (ii) the propriety of a "linguistic theory … concerned primarily with an ideal speaker-hearer in a completely homogeneous speech-community, who knows its language perfectly" (Chomsky, 1965: p. 3). Let us take these in turn.

Regarding (i), the following point is, I think, crucial: The interest in psychological states and processes is *not in the least bit diminished* by the recognition of the existence of public language, nor by the recognition that the syntacticians can be easily seen as formulating principles of the grammars of dialects rather than I-languages.[20] As we will see in later chapters, the relation between syntactic theory and psycholinguistic models of parsing can be quite complex. A grammar formulated by a syntactician who is either committed to studying E-language or is agnostic about the E-language/I-language distinction and the attendant controversies, can be just as fertile a source of hypotheses for the psycholinguist and the computational modeler as a grammar designed specifically to be "about" some I-language. More generally, as noted earlier, the issue regarding whether syntacticians are studying E-language or I-language is *orthogonal* to the issue of whether some grammar or other is psychologically real. It may well be that syntacticians' grammars are, in the first instance, *about* E-language, but that one or another such grammar can subsequently be shown, on the strength of psycholinguistic evidence to be mentally represented and used for comprehension and production.

Regarding (ii), there is this to say: We must be open to a theorist's wish to study a phenomenon by making idealizations—a strategy that is both fruitful and ubiquitous in the natural sciences. A problem does arise, however, if the theorist either mistakes an idealization for reality or if s/he denies the existence of the very reality that s/he is idealizing away from. In the case of linguistics, the reality is that there are communities of speakers whose linguistic performance is susceptible to powerful generalizations. For instance, we can say truly of a great many speakers that their uses of the pronoun 'him' in the sentence 'Bill said that John loathes him' are *not* references to John (assuming, of course, that John ≠ Bill).[21] Such generalizations may, on occasion fail to apply to specific speakers, on account of one or another quirk of individual psychology, education, or other factors. The purpose of idealization, then, might be to exclude such cases as not being indicative of the nature of the public language shared by the community. It need not be seen as a commitment to the study of each community member's I-language, let alone any one specific I-language—a conception that is neither mandatory nor even a particularly natural way of interpreting the practices of working linguists. Establishing it requires an

[20] This is an echo of the point that Devitt (2006a) makes concerning the compatibility between adopting his Linguistic Conception and studying knowledge of language. Devitt writes: "[M]y contemplated task must be worthwhile if Chomsky's task (i) is. [Task (i) is answering the question: What constitutes knowledge of language? D.P.] For, although we have distinguished the two tasks we have also related them in a way that makes completing the contemplated task necessary for completing task (i). For, the nature of the speaker's competence studied by task (i) involves the nature of the symbols studied by the contemplated task: those symbols are what the competence produces. Indeed, our earlier discussion shows that the contemplated task has a certain epistemic and explanatory priority over task (i). How could we make any significant progress studying the nature of competence in a language unless we already knew a good deal about that language? Just as explaining the bee's dances is a prerequisite for discovering how the bee manages to produce those dances, so also explaining the syntax of sentences is a prerequisite for explaining how speakers manage to produce those sentences" (pp. 28–29).

[21] Needless to say, even more general and theoretically robust claims can be made—generalizations that abstract away from the use of the names 'Bill' and 'John', and away from the fact that this sentence has only one embedded clause. I have in mind the generalizations of Binding Theory and its successors.

argument that I suspect will be difficult to come by. The difficulties, moreover, are not limited to the ones I have adumbrated thus far. In the next section, I will argue that the project of individuating I-languages faces a number of challenges that are often overlooked by philosophers of linguistics.

3.4 Individuating I-Languages

While I-languages are not themselves collections of text or verbal behavior—the outputs or products of linguistic competence—such things must, of course, serve as *evidence* for the ascription of an I-language to any specific individual, as well as for any claims regarding the individuation of I-languages. A linguist who wishes to write down a grammar for a specific I-language has essentially two options: She can either survey a large corpus of the target individual's speech, writing, and signing, or she can ask that individual to provide intuitions about the grammaticality of a large class of preselected utterances, inscriptions, and signs.[22] For expository convenience, let's call the target individual 'Ana', and let's give the label 'utterance' to any item in the class of Ana's spoken, written, and signed expressions.

Suppose the linguist decides to go the corpus route. Immediately, certain decisions have to be made about which of Ana's utterances to keep in the corpus and which to throw out as "bad data"—false starts, utterances made while chewing or yawning, drunken babble, capricious or impulsive mid-sentence switches of topic, interrupted speech, utterances made in the heat of violent passion, careless typos, unforeseen tongue-twisters, spoonerisms, mixed metaphors, and so on. The corpus must also be divided up in a way that reflects the fact that Ana is very likely to switch between a number of dialects, or even be multilingual (as the "unreflective folk" would put it). The linguist would, in addition, have to take into account changes in Ana's speech over long periods of time. For instance, suppose that it's plain that Ana's college education effected a radical change in her speech and writing. One might argue, of course, that the education merely boosted Ana's *performance*, but had no effect on her competence grammar. Though initially tempting, this view raises more questions than it answers. Suppose, for instance that, prior to entering college, Ana consistently rejected sentences of the following form—i.e., she never produced them and, when prompted, judged them strongly unacceptable.

(1) That the man went is a truth.
(2) That we have come to a firm decision is unclear.
(3) That you eat is a fact.

Suppose, further, that Ana gradually became increasingly comfortable with such sentences as she progressed through her college career, balking less frequently or

[22] Tellingly, it's rare to see anyone undertaking the project of individuating an I-language. Linguists typically study individual subjects only when they have some very special property, such as a rare linguistic or cognitive disorder.

intensely at their appearance in textbooks. We can imagine Ana eventually coming to use such sentences in her writing—at first with trepidation, later confidently—and in her speech, though only in the rare cases where the rhetorical register allowed for it. While it's clear that inebriation and temporary inattentiveness can breed performance errors, it's not *at all* clear whether Ana's initial attitudes toward CP-topicalization constitute a performance error.

One might argue that Ana's I-language prior to enrolling in college consistently allowed other types of topicalization, and that a maximally simple grammar of that I-language would also allow CP-topics. But whatever gains in simplicity this move affords must be carefully weighed against the complexities that must then be introduced into the performance model to support it. Specifically, we would now need to posit a mechanism somewhere in Ana's linguistic processing routines that plausibly accounts for her initial inability or reluctance to understand, produce, and otherwise accept only *some* types of topicalization, but not others. Perhaps there is an optimal way of striking a balance between these competing pressures. But it must be admitted that any such judgment regarding relative simplicity of otherwise empirically adequate theories is bound to be extremely subtle.

From this, we can draw two morals. First, the playing field on which the notions of E-language and I-language compete must be a level one. In particular, the theorist who is bent on constructing or refining a notion of public dialects, public languages, or E-languages should be allowed to deploy heavily theoretical resources in the course of her efforts. After all, her counterpart, the I-language theorist, is allowed to appeal to the competence/performance distinction and to employ a largely implicit methodology for drawing that distinction in difficult cases. Second, when the E-language theorist encounters cases in which her theoretical decisions are less than principled—i.e., recognizable as more-or-less arbitrary judgment calls—we should *not* conclude that her project is fundamentally misguided. For, our standards are simply not that high in the case of I-languages. Provisional decisions, even when somewhat arbitrary, are tolerated in isolated difficult cases, if only in the hope that more principled lines can be drawn in the fullness of time, when theoretical progress affords us with conceptual and experimental tools to address increasingly subtle issues.

The friend of crisp I-languages may, at this point, object that we have been unfair to his position. I-languages, he says, are only determinate *at a time*; talk of changes due to education and the like is simply beside the point. But problems loom even for this view. How fine-grained, exactly, is the individuation supposed to be? Does Ana wake up with a new idiolect every day? Does her I-language change from one hour to the next? One minute to the next? One second?

Take a related worry: Suppose Ana is bilingual—a competent speaker of what the folk would unreflectively label 'Russian' and 'English'. Calling Ana bilingual or multi-lingual *hints* at a public conception of language, because without reference to Ana's similarity to other people—that is, without reference to two separate *languages* or *dialects*—it makes more sense to classify her as *mono*lingual (or mono-idiolectal). That is, although Ana would be classified as bilingual by ordinary standards, the position we are now considering entails that her I-language includes *all* of the things that she is disposed to produce, comprehend, and evaluate (again, performance errors aside, whatever that comes to). Thus, given that her I-language includes what we would ordinarily call the English and the Russian language, the crisp-idiolect view has it that Ana's I-language is best thought of as $\text{Runglish}_{\text{Ana}}$. This view, however, faces a challenge from the opponent who asserts that Ana actually has *two* I-languages, simultaneously—call them $\text{Russian}_{\text{Ana}}$ and $\text{English}_{\text{Ana}}$—and that she draws on one of these for some purposes and on the other for others. How to settle this dispute?

Looking at the relevant empirical research, we discover the well-documented fact that polyglots code-switch at the drop of a hat (Gardner-Chloros 2009). The literature on multilingualism suggests that this phenomenon, far from being uncommon, is actually the norm in a vast range of human cultures (Auer and Wei 2007). The prevalence of monolingualism is a relatively recent phenomenon, and even in the present day is confined to social groups that have rather peculiar socio-economic relations to the rest of the human population.

With this in mind, consider asking a speaker of English, Russian, and Hebrew what he thinks of the following sentence.

(4) Я решил definitely принести some wine to the חגיגה משפחתית tonight.
[I <u>decided</u> definitely <u>to bring</u> some wine to the <u>family celebration</u> tonight.]

The trouble starts well before the question is asked, for it is no small matter to decide what the question should even *be*. Asking whether the sentence is acceptable in *his* language presupposes that he has only one. But he may well reject that presupposition, either out of hand or upon careful reflection, on any number of grounds. He may cite, for instance, the fact that *only* Hebrew—not any of his other languages—is used for religious practices. Equally, he may bring up the fact that he remembers phone numbers and performs basic calculations only "in English," because that's what he finds easiest. Finally, he might trot out his distinctive preference for swearing in Russian—a language whose morphosyntax seems positively designed for the productive generation of ever more intricate forms of profanity. He might add that his preference for Russian swear words is especially pronounced when he is embroiled in a heated situation. Having brought to bear a wealth of psychological and socio-cultural considerations, the man would surely be puzzled at a linguist's suggestion (let alone insistence) that, strictly-speaking, he only knows *one* language, "in the technical sense." And yet the corpus of his utterances contains (4) and a great many such mongrels. Are these performance errors? Do we know how to tell?

One might hope to find the answers to these questions in the field of second-language (L2) acquisition—an area of study whose very title betrays a folksy commitment to either public languages or multiple I-languages residing in one speaker. Unfortunately, there is widespread disagreement amongst theorists who study L2 acquisition about the boundaries between internalized grammars. In some cases, it seems that there is a "core" grammar that is being made to do double duty, as the speaker deploys it in the course of using "more than one language"—a phenomenon that falls under the heading of *L1 transfer*. In other cases, it's just not clear *what* to say. The questions are, as always, empirical, which warrants the hope that they will one day be resolved on principled grounds. But, in practice, the data that are brought to bear on such issues are often subject to conflicting interpretations, all of which appear antecedently plausible—a situation that requires suspending judgment pending further clarification.

Again, I urge that conceptions of E-language be allowed the same leeway. At present, they are often summarily dismissed by Chomsky and his followers as being too vague and ill-defined to be fit for serious empirical inquiry, as though the individuation of I-languages is, by comparison, a straightforward and uncomplicated matter. This double standard is, by my lights, both unfounded and obstructive to progress. Making precise the individuation conditions on I-languages is an ongoing research program. We should view the project of individuating E-languages in the same way, rather than expecting that the friend of E-language will be able to provide crisp and clear definitions in advance of sustained inquiry.

Returning to the thought experiment, suppose that another linguist, undaunted by the difficulties adumbrated above, persists in the quest to divide up Ana's utterances into exactly two crisp I-languages. He might begin by sorting her lexical items into those that belong to the I-language $\text{Russian}_{\text{Ana on March 29th, 2016, 5:44:06 PM}}$ and those that belong to the I-language $\text{English}_{\text{Ana on March 29th, 2016, 5:44:06 PM}}$. How might this be accomplished?

One natural suggestion would be to take stock of the similarities and differences in the *phonological* features of the lexical items. But if Ana has made it a point to pronounce "correctly" certain "foreign" words (including names)—in the way that many people did some years ago with the name of Icelandic volcano, Eyjafjallajokull—then this procedure would not yield a binary partition. Suppose Ana, like many of us, regularly uses the following names and expressions: *Amen*, *je ne sais quoi*, *c'est la vie*, *que sera sera*, *gesundheit*, *chutzpah*, *gracias*, *enfant terrible*, *mea culpa*, *verboten, machismo, vis-à-vis, a priori, prima facie, Gödel, and Wagner*. It would be rash to insist that these are not part of her lexicon. At the same time, it would be pointless to increase the number of I-languages that we attribute to Ana, simply on the basis of the relevant phonological considerations. The issue becomes even more acute if we learn that, in the course of pursuing her acting career, Ana has fully mastered over a dozen English accents.

Another proposal involves individuating lexical items by reference to the syntactic frames in which they are licensed. But this, too, runs into problems. Consider (5).

(5) Я решила definitely принести some wine to the party tonight.
[I <u>decided</u> definitely <u>to bring</u> some wine to the party tonight.]

Is the word 'definitely' *licensed* in (13)? Should 'принести' (transl. *to bring*) really be 'to принести', in accordance with the grammar of English infinitives, or is the sentence grammatical as it stands? Is (5) even a *sentence*—i.e., a string generated by a single, consistent grammar? If so, which grammar is that?

The difficulties that we have catalogued thus far can be multiplied *ad nauseam*. Ludlow (2011: pp. 44–6) makes passing reference to many of the considerations that I have raised here. However, he comes to the conclusion that they constitute problems *only* for the friends of E-languages—particularly E-*idiolects*. Having argued that these problems are devastating to the enterprise of E-linguistics, he claims that the very same problems constitute a "research program" for I-linguistics (p. 46). The suggestion seems to be that an analogous research program in the case of E-language would be unprincipled or theoretically fruitless. But this double standard begs the question against a position like Devitt's, according to which contemporary linguistic theory—a fruitful research program—is actually *about* E-languages. Moreover, Ludlow seems to ignore the fact, stressed earlier, that a theorist who wishes to study E-languages is entitled to the same sorts of idealizations that animate inquiry into I-languages. Such idealizations serve precisely to abstract away from the problematic types of inter- and intra-speaker variation that Ludlow has in mind.

Nothing I have said entails that there is no way of constructing individuation conditions that take into account all of the subtleties that constrain the identity of an I-language. Perhaps some of the difficulties raised above can be resolved on principled grounds. My aim, rather, has been to point out that, for many theoretical pursuits, the precision may not be worth the effort. One might well simply idealize away from the messy variation. And, as always, the nature of the idealization will depend on one's theoretical purposes—in this case, the purposes to which a notion of I-language will be put.[23]

For some purposes, the identity conditions on I-languages can be made strict enough that a suitable proportion of Ana's lexical items and utterances are treated as though all they fall neatly into exactly two classes and obey exactly two sets of grammatical principles. Ana can thus be said to have exactly *two* I-languages (on March 29th, 2016, 5:44:06 PM). Questions would remain, of course, about whether the project of individuating those I-languages can really avoid making reference—even *tacit* reference—to larger social groups or cultural constructs (Sect. 3.1). Likewise, the exact import of the competence/performance distinction would need to be specified, in more fine-grained terms than one often finds in the literature on I-languages.

On the other hand, the identity conditions on I-languages can be left so loose that Ana would count as having only *one* I-language. Perhaps we say this simply because she has only one body, or one brain, or—in the limit—only one *self*. (Of course, individuating *these* entities is no small feat either, as generations of philosophers have come to realize.) Another possibility is to adopt the formula: One I-language per language faculty. But this invites still other kinds of trouble. Individuating lan-

[23] This point has an analogue in the study of E-language. Devitt (2006a) points out that linguistic theory—conceived as an inquiry into the nature of E-language—can proceed at various grains of analysis. In identifying the group of speakers that she wishes to study, the linguist can, for the most part, make do with a term like 'English', "but sometimes [she] need[s] a less precise term like 'Romance' and sometimes a more precise one like 'Australian–English' or 'the Somerset dialect of English' (p. 184).

guage faculties is itself something of a dark art, as cases of brain damage, bilingual aphasia, agraphia, and alexia serve to highlight.[24]

3.5 Conclusion

In this chapter, I addressed the cluster of issues surrounding Chomsky's distinction between E-language and I-language. Chomsky (1988) claims that any theory of E-language will "surely have to presuppose grammars of I-languages." In his view, whatever an E-language might be, it is "more abstract" than I-language, on account of its being "more remote from mechanisms" (p. 561). But one might reasonably challenge Chomsky's presumption that the psychological mechanisms underlying any individual's language use are describable without reference to social facts—without, e.g., a prior statement of the grammar that best captures the conventional language of a particular linguistic community.[25] My argument in this chapter is intended to render plausible the idea that references to E-language play a role in individuating I-languages, and in explaining how they are acquired.

In Sect. 3.2, I argued that the notions of E-language and public language are indispensable to the study of language acquisition, as actually practiced by working psycholinguists. A close look at the literature on language acquisition illustrates that the grammars of public languages serve as "targets", against which acquisition theorist measures a child's successes and failures throughout development. I showed that the data and explananda of acquisition theory are routinely couched in terms that make seemingly ineliminable reference to public languages. In particular, a child's deviations from the public language of his or her linguistic community are regarded, in mainstream acquisition literature, as errors. Though this does introduce what Chomsky calls a "normative-teleological" element into the science of language acquisition, there is no obvious reason to regard this as a move toward prescriptive linguistics, nor as a tacit appeal to socio-political considerations. The normative-teleological element is innocuous, deriving from a theoretically motivated idealization of each child's linguistic community.

It may be, then, that the notion of E-language has primacy over that of I-language—the exact reverse of the situation that Chomsky envisages. Alternatively, it might be that the two notions are, as it were, coeval and hence on a par with one another. The arguments that I have offered here do not settle the matter. To do this, one would have to, at a minimum, distinguish between a variety of primacy theses, including those having to do with the temporal, ontological, phylogenic, ontogenic,

[24] See also the discussion in Chap. 10 of *Ignorance of Language*, where Devitt summarizes the results of various neurocognitive studies and draws on them in casting doubt on the unity of the language faculty.

[25] Needless to say, an anatomist might be able to discuss the underlying neurophysiology of these mechanisms. But that is not what Chomsky has in mind. It's the *psychological* level of description that concerns him, also the level to which the E-linguist's argument is addressed.

explanatory, evidential, and conceptual relations between E-language and I-language. Clearly, the primacy issue deserves a great deal more discussion.

The position that I have developed is incomplete without a rigorous statement of the identity conditions on E-languages. Achieving this would require settling issues about how much agreement is required in competent speakers' intuitions, to what extent—and in what respects—their linguistic output must be similar, to what degree their comprehension of the same utterances must coincide, how to gauge the success or fluidity of their linguistic communication, and so forth. But while constructing an apolitical notion of public languages is difficult, we must not confuse difficulty with impossibility. I am confident that increasingly precise characterizations of E-languages can be filled in as theoretical and experimental advances are made.

In Sect. 3.3, I argued that the project of individuating I-languages is in much the same boat. The lack of precision in our notion of E-language is just as much a feature of our conception of I-languages. To provide plausible individuation conditions for I-languages, one must settle a variety of issues pertaining to the competence/performance distinction, the mimicry of "other" dialects, change over time, multilingualism, code-switching, the infusion of "foreign-language" lexical items, brain damage, cognitive disorders, and related phenomena. I have suggested that meeting these challenges requires making fine-grained decisions that we do not, at present, know how to make on principled grounds. Bearing this in mind, we should ask what degree of precision it's really necessary to achieve in order to get on with fruitful theorizing. I suspect that neither in the case of E-languages nor in the case of I-languages will any rational theorist insist on *maximal* precision at the very outset of inquiry.

Chapter 4
Language Acquisition and the Explanatory Adequacy Condition

Abstract I examine John Collins' reconstruction of the cognitive revolution in linguistics, showing that one of the main arguments for cognitivism is simply not compelling. While there is a convincing case for aiming to achieve "explanatory adequacy" in linguistics, over and above mere observational and descriptive adequacy, this aim need not be underwritten by a cognitivist conception of language. A unified theory of all human languages is desirable whether or not cognitivism is correct. Next, I point out that, although cognitivism entails that grammars are psychologically real, the reverse entailment does not hold; a grammar can be psychologically real even if the objects of the formal syntactician's concern are public, conventional E-languages. Chomsky's view entails that psycholinguists should seek a relatively transparent relation between the syntacticians' grammar and the "knowledge-base" that constitutes competence—a "natural" grammar-parser combination. Progress toward this goal has been slow, in part because syntacticians are not as concerned with psycholinguistic data as a cognitivist would expect them to be. In the mainstream syntax literature, psychological reality is a distant, dimly understood, and rarely invoked desideratum. Nevertheless, a parsing model that makes direct use of independently plausible syntactic principles is the simplest and strongest theoretical option.

Keywords Explanatory adequacy • Observational adequacy • Descriptive adequacy • The cognitive revolution • Cognitivism • Language acquisition • John Collins • Binary branching • Compactness • Psychological reality • Parsing • Innateness • Principles and parameters • Triggering • X-bar theory

4.1 Introduction

I devote this chapter to examining John Collins' reconstruction of the cognitive revolution in linguistics. My aim is to show that one of his main arguments for "taking the cognitive turn" is simply not compelling. While Collins makes a convincing case for aiming to achieve *explanatory* adequacy in linguistic theory, over and above mere observational and descriptive adequacy, I will argue that this aim need *not* be

D. Pereplyotchik, *Psychosyntax*, Philosophical Studies Series 129,
DOI 10.1007/978-3-319-60066-6_4

underwritten by a cognitivist conception of language. A unified theory of all human languages is desirable whether or not the cognitivist conception is correct.

4.2 Does Pursuit of Explanatory Adequacy Require the Cognitive Conception?

First, we'll need some concepts from linguistic meta-theory. A grammar, considered as a theory of a specific natural language, must meet three distinct adequacy conditions. In order of increasing strength, these are: *observational*, *descriptive*, and *explanatory* adequacy. Observational adequacy consists in the grammar's merely generating each of the surface strings of a language, irrespective of the structures that it assigns to them.[1] Descriptive adequacy, by contrast, requires the grammar to generate the correct syntactic structure—or, in cases of ambiguity, multiple correct structures—for each string. These conditions are typically thought of as "external," in the sense that they deal with the interface between linguistic theory and empirical data. There is a question about whether descriptive adequacy is really best characterized as an external condition, in this sense. Let us pause to clarify this issue, with an eye toward developing some claims that will assume prominence later in the discussion.

When two distinct grammars, G_1 and G_2 are observationally adequate, it is often the case that one of them—say, G_1—achieves observational adequacy by positing vastly more rules than the other, G_2, or by offering distinct explanations of phenomena that G_2 treats in a unified manner (Larson 2010). In such cases, the choice between G_1 and G_2 is *not* made by appeal to some data that G_2, but not G_1, can predict or explain; *ex hypothesi*, no such data exist. Rather, the choice is made on the

[1] There are important ambiguities in the term 'surface string', which bear on issues concerning the coherence and utility of the notion of E-language. On the one hand, a surface string might be a linear ordering of words, morphemes, or phonemes. These categories are picked out by technical notions from linguistic theory and hence might be argued by a Chomskyan theorist to be I-linguistic—that is, psychological—entities. On the other hand, a surface string might be a linear ordering of inscriptions, acoustic waveforms, or muscle contractions. In this case, surface strings are plainly E-linguistic entities. But it is difficult to make sense of the idea that a grammar might *generate* such things. As I see it, the friend of E-language must argue (i) that the notions 'word', 'morpheme', and 'phoneme' actually have their home in E-linguistics and (ii) that an adequate grammar will include some sort of systematic mapping from these theoretical constructs to the observational categories that include inscriptions, acoustic waveforms, and muscle contractions. A formal statement of such a mapping would license talk of a grammar generating observable E-linguistic entities. (Analogously, the cognitivist must specify *psychological* mechanisms that compute phonological, morphological, and syntactic representations on the basis of causal encounters with inscriptions, acoustic waveforms, and muscle contractions.) Note, finally, that even when this ambiguity in 'surface string' is resolved, there remains a further issue: must an observationally adequate grammar generate only those strings that have actually been observed—e.g., recorded in a corpus—or must it also generate strings that either *will* or *might* be observed? In what sense of "will" and "might"? See Quine (1953) for reflections on this and related matters, and Chap. 2 for a discussion of its bearing on the infinitude issue.

grounds of what philosophers of science sometimes call "super-empirical virtues," which include, *inter alia*: formal simplicity, generality of coverage, and explanatory unity (Quine and Ullian 1978; Kuhn 1977; Churchland 1985/1989).

It is on *these* grounds that Chomsky (1957) argued against pure phrase-structure grammars—what Fodor et al. (1974) called "Immediate Constituency (IC) grammars"—and in favor of transformational grammar. Similar considerations played a significant role in motivating the shift away from early transformational grammar (the Standard and Extended Standard Theory) toward the Principles and Parameters (P&P) framework, which dispensed with the growing stock of language-specific and construction-specific transformations (Chap. 9). If pursued further, the early transformational grammars might have turned out to be observationally adequate. But the P&P framework was superior precisely in that it posited fewer formal principles and unified the explanation of phenomena that, from the earlier perspective, seemed heterogeneous. These advantages of P&P grammars are, of course, quite real and significant, but they are, in the end, "theory-internal," having little to do with specific data that went unexplained by earlier theories. Importantly, they also have little or nothing to do with the psychological status or implications of formal grammars—a point I develop in the next section.

Characterizing the third adequacy condition—i.e., explanatory adequacy—is a difficult matter, for, as we will see below, it raises a number of issues surrounding the very motivation for a cognitivist conception of language. One mainstay among the battery of arguments in favor of the cognitivist conception relies on the claim that only *this* conception can supply us with principled grounds for aiming at explanatory adequacy. Indeed, it is sometimes suggested that only the cognitivist is able to so much as draw a principled distinction between the "external" adequacy conditions—observational and descriptive adequacy—and the "internal" condition of explanatory adequacy. Let us take a moment to briefly review how it accomplishes this.

On Chomsky's view, the initial state of the language faculty encodes a Universal Grammar (UG)—roughly, a kind of "template" for the grammars of specific languages, such as French, Russian, or Hindi. As such, it encodes the structural properties that are *universal* among human languages—i.e., the features common to all of them.[2] Furthermore, Chomsky takes the initial state of the language faculty to be centrally involved in language acquisition, in the sense that it provides the child with the innate resources necessary for structuring and organizing linguistic experience. In his early writings, Chomsky conceived of the child's innate endowment as having three components: (i) a specification of all possible human grammars, (ii) a simplicity metric, along which those grammars are ranked with respect to observational data, and (iii) a device or a computational mechanism for taking in the observational data and ranking the grammars along that metric, eventually ranking

[2] The term 'feature' is to be understood in its generic sense, though there is a way in which this claim is true even if the term is taken to have its more technical meaning from phonology and syntax, e.g., in Minimalism (Chap. 9).

highest what is sometimes called the "target grammar"—the grammar of the "ambient language" in the child's speech community.

Chomsky's view of the matter has evolved considerably since the publication of *Aspects of the Theory of Syntax* (1965). In his *Lectures on Government and Binding* (1981) and *Knowledge of Language* (1986), Chomsky introduced and elaborated the Principles and Parameters (P&P) model, according to which the structure of any possible human language is determined by a handful of core syntactic principles, each of which is parameterized.[3] Language acquisition, on this model, is a matter of setting parametric values on innately given principles. One total consistent setting yields an idiolect of, say, English, while another yields an idiolect of, say, Tagalog. The values of the parameters are, in theory, *triggered* by specific inputs, which the learner is assumed to perceive as structured, presumably in accordance with some prior grammar (J. D. Fodor 1998b).

As is well known, this model of acquisition faces a broad range of difficulties.[4] Whether these can be adequately addressed is not a question that concerns us here. For present purposes, all we need is an understanding of how the P&P framework—considered as an approach to *both* formal syntax *and* acquisition theory—animates the "internal" condition of explanatory adequacy. Within the P&P framework, a grammar of a particular natural language is said to be explanatorily adequate when it can be derived from UG, which is assumed to be encoded in the initial state of the language faculty.

We are now in a position to see how the aim of achieving explanatory adequacy in syntactic theory is thought to underwrite an argument for the cognitivist conception of language. For, on the cognitivist conception, language acquisition is one of the primary explananda of linguistic theory.[5] As such, the project of explaining acquisition becomes intimately bound up with the project of finding linguistic uni-

[3] The parameters are often taken to be binary, but this is not a core claim of the P&P model. It is, instead, an additional claim, to be established on independent grounds. In this respect, it is like the claim that the parameters refer solely to the features of functional heads—a condition imposed by grammars in the Minimalist tradition (e.g., Chomsky 1995).

[4] For a decidedly unsympathetic evaluation, see Tomasello (2005). Tomasello raises a serious objection to the P&P model of acquisition, targeting specifically the assumption, highlighted above, that the linguistic input does not come in the form of a labeled syntactic structure, but, rather, in the form of a messy acoustic and visual stream. For a child to assign *any* syntactic structure to pieces of the primary linguistic data, it must have somehow *already* bridged the gap between an acoustic description of the data and the more abstract description couched in some or other syntactic formalism. Tomasello's point is that whatever cognitive resources allow the child to do *this*—in his view, sophisticated statistical reasoning, coupled with innate mind-reading abilities—are, in principle, powerful enough to both generate a rudimentary syntactico-semantic framework, and then to refine that framework in accordance with the needs engendered by increasingly complex social interactions. The objection is, I think, a powerful one, but by no means decisive. Impressive efforts to meet it have been made, e.g., by J. D. Fodor (1998b).

[5] Needless to say, opponents of the cognitivist conception can, should, and do allow that language acquisition is one of the primary explananda of *some* theory. What they deny is that the theory in question is the formal grammar of the language, or of human language more generally. The acquisition theorist will doubtless *appeal* to such a grammar, but acquisition of the grammar will be seen by the anti-cognitivist as a separate matter.

versals, or at least patterns of commonality among seemingly diverse languages. Once found, such universals can be assumed to be innate in the mind of the prelinguistic child, and to contribute to his or her acquisition of the local language. With this picture in mind, it's difficult to see how any *other* conception of language can license serious inquiry into linguistic universals or, perhaps equivalently, into the structure of human language *as such*, over and above the structure of particular dialects or idiolects. If language acquisition were *not* an explanandum for linguistic theory, or if the explanans did not make reference to a psychologically real Universal Grammar (UG), then what other role could UG possibly play in linguistics? And if it plays none, then why spend time devising an account of it?

Tempting as it is to let these rhetorical questions stand as an argument for the cognitivist conception, I think we can provide reasonable answers to them, thus blunting their rhetorical force. This is what I propose to do below. Adopting the cognitivist conception, I shall argue, is *not* the only way of drawing the distinction between "external" and "internal" adequacy conditions, *nor* the only way of motivating the condition of explanatory adequacy. Put differently, the eminently reasonable aim of constructing grammars that meet *an* explanatory adequacy condition is *quite neutral* as regards the cognitivist conception and its rivals. As such, the cognitivist conception is—pending further argument—strictly optional.

4.3 Defusing an Argument for the Cognitivist Conception

A natural place to begin is with Chomsky's earliest works, *Syntactic Structures* (1957) and *The Logical Structure of Linguistic Theory* (LSLT).[6] Here, the demand for genuinely *explanatory* theories of language is already clearly in place—indeed, serves as one of the hallmarks of the profound shift that Chomsky wrought in linguistics. But, crucially, this demand is in no way motivated by cognitivist concerns. In a comprehensive exegesis of Chomsky's work in linguistics, Collins (2008a) writes:

> Paired with external adequacy conditions is an 'internal' adequacy condition (*Logical Structure of Linguistic Theory*, pp. 80–1, *Syntactic Structures*, p. 50). The external conditions relate a given grammar to the facts (data); the internal condition relates a given grammar to GLT [general linguistic theory—a theory of *all* natural languages, not just a specific one]. This notion is something more than the innocent idea that a given grammar may be evaluated in relation to one's ongoing general conception of a grammar as determined by the set of grammars under construction. General linguistic theory, Chomsky (LSLT, ch. 4) hopes, may provide a simplicity metric by which one could choose the simplest grammar among a set that that all meet external conditions of adequacy. (p. 37)

It is plain, I think, that this approach is vastly superior to the opposing methodology that stressed quasi-botanical "discovery procedures" at the expense of explanation.

[6] Chomsky penned *The Logical Structure of Linguistic Theory* in 1955–1956, prior to publishing *Syntactic Structures* (1957), but the work only appeared in print in 1975.

However, what Collins fails to make clear is how the idea of using General Linguistic Theory (GLT) to provide the relevant kind of simplicity metric is supposed to go *beyond* "the innocent idea that a given grammar may be evaluated in relation to one's ongoing general conception of a grammar as determined by the set of grammars under construction." Later in his discussion, Collins motivates this "innocent idea" as follows:

> It *might* be that language is essentially heterogeneous, but it would be absurd to assume that constructing grammars for L_1 and L_2 are wholly independent enterprises. If one inquiry can usefully constrain the other, then it would be obtuse to eschew the discipline. Otherwise put, far from *assuming* that all languages are cut from the same cloth, Chomsky is suggesting that it is a matter of empirical inquiry. We construct individual grammars and test them against one another in terms of their shared resources. The extent to which the grammars all conform to the same general conditions and are externally adequate is the extent to which we are discovering something about language as such. Inquiry into particular languages thus goes hand in hand with the general inquiry into language, each constraining the other (Collins 2008a: pp. 83–4, emphases added)

The crucial point to note here is that the motivation for constructing a General Linguistic Theory is, on the one hand, quite cogent, but, on the other hand, has nothing to do with language acquisition, nor, indeed, with *any* aspiration of psychological theorizing.

To motivate a shift away from the structuralist methodology, according to which language should be studied as though it were "heterogeneous"—i.e., indefinitely variable from one speech community to the next—one need only stress this: It's a virtue of *any* theoretical inquiry that it set the bar high, imposing stringent evidential constraints and adequacy conditions. Generality of theoretical coverage is to be valued *for its own sake*. And, while Chomsky deserves credit for urging this point against the structuralists, we should guard against attempts to treat it as a proprietary feature of the cognitivist conception of language. For, again, *anyone*—even the staunchest behaviorist—can fit a demand for theoretical generality into her methodological outlook.[7] It follows that an insistence on explanatory adequacy *cannot* serve as grounds for an argument in favor of the cognitivist conception. In what follows, I elaborate this line of reasoning in response to Collins' paradigmatic statement of the rationale for taking what he calls the "cognitive turn."

In what I take to be one of the most important passages of his book, *Chomsky: A Guide for the Perplexed* (2008a), John Collins puts forward precisely the kind of

---

[7] In the main text, I have stressed the notion of theoretical generality. But there is also something to be said for the super-empirical virtues of explanatory simplicity and, perhaps more importantly, explanatory *unity*. As noted above, the P&P model was motivated in large part by the fact that it can be used to explain a wide range of seemingly disparate phenomena within a particular language by appeal to a handful of interacting syntactic principles. (See, e.g., Ludlow 2011: Chap. 1 for detailed examples.) Moreover, the model uses a small number of parameters to account for seemingly unrelated differences *between* languages. A prime example of this can be found in the unification of a great many linguistic properties by reference to the *pro*-drop parameter. (See Haegeman 1994: pp. 19–25 and Berwick 1991b.) For our purposes, the important point is that these virtues are enjoyed by the P&P model *irrespective* of whether that model is construed as a theory of the cognitive states involved in acquisition.

argument that I am here seeking to defuse. It will be instructive to quote him at length.

> The cognitive turn, then, essentially arises via a consideration of how language is acquired. The internal justification of general linguistic theory becomes a condition of 'explanatory adequacy', where a grammar is adequate only if it is explicable as arising from what all humans share qua an organism that can acquire any human language. Following Chomsky (ATS, ch. 1), we will call whatever this shared property as [sic] *Universal Grammar* (UG). Chomsky (EML, CILT) also refers to UG as a Language Acquisition Device (LAD). This is simply a label for whatever it is about the human mind/brain that maps from exposure to language to a grammar. However it is perfectly natural to construe general linguistic theory as UG, what all humans share. After all, every human acquires a language and so somehow represents a grammar for their language qua creative users of the language. Since, further, general linguistic theory describes the common resources of the grammars and there must be something common to all humans as acquirers of language, it looks as if general linguistic theory was all along an account of a universal human cognitive feature, that is, UG. In line with this cognitive construal of the old methodological notions, 'external' justification becomes a desideratum for 'descriptive adequacy', that is, a grammar should map a structural description to each sentence such that the set of descriptions explains our intuitions as to the structure within and between sentences. In general then, given a grammar G and language L, we may take G to be descriptively adequate of L if it explains the various systematic features of our knowledge or understanding of L via an assignment of structural descriptions to L sentences. G itself cannot be explanatorily adequate, but G is not adequate unless we are able to show that it is an acquirable grammar from UG and data to which the child is likely to be exposed. Explanatory adequacy relates UG to particular grammars; the condition distinguishes between, as it were, possible languages from 'arbitrary symbol systems' (ATS, chapter 1). In sum, linguistic theory becomes an account of human *knowledge of language*. Particular descriptively adequate grammars constitute theories of what speakers know when they know a language, and UG is what speakers know in virtue of them being able to acquire any given language, that is, to represent a given grammar for their language. It should be noted that Chomsky (LSLT, p. 62) has always been happy to treat language as an object of knowledge and so a psychological phenomenon. What is new with the work of the early 1960s is the explicit construal of linguistic theory as having psychological states of knowledge as their object. (pp. 85–6)

This passage contains a *move* from a seemingly neutral conception to one in which "linguistic theory *becomes* [my emphasis] an account of human *knowledge of language*." Collins may not have intended this passage as a stand-alone *argument* for the cognitive conception of linguistics. He may have intended only a description of what he sees as the historical facts concerning the cognitive conception's rise to prominence. But in the broader context of his book, Collins' remarks leave one with no doubt that he is an ardent supporter of the cognitive conception. He takes the "cognitive turn" to be a turn in precisely the *right* direction. And, given what he goes on to say in the remainder of the chapter, it is difficult to credit the suggestion that the passage quoted above is intended to carry no argumentative force whatsoever.

On the other hand, if the passage *is* indeed an argument, then it plainly begs the question against a position like Devitt's. For, Devitt takes great pains to argue that the Representational Thesis is unmotivated and probably false. Yet Collins simply *assumes* (some version of) the Representational Thesis when he writes that "the human mind/brain … *maps from exposure to language to a grammar*" and, more obviously, that "every human acquires a language *and so somehow represents a* *grammar for their language*" (emphases mine). As far as I can discern, nothing that Collins says prior to this passage warrants these claims.

It is, of course, possible that all Collins means by 'represents' here is something quite weak, like 'behaves in accordance with'. In this sense, any device that computes a function thereby represents that function. To use Devitt's terminology, the device *respects* the structure-rules of the grammar.[8] Unfortunately, as Devitt convincingly argues, this wouldn't be sufficient to show that grammars are theories of the mind, any more than the rules of a card game such as poker are theories of the mind. Devitt would insist that, like the rules of poker, grammars do no more than specify in detail what sorts of inputs and outputs a person will take and emit. Without further argument—specifically, without what Devitt (2006a) calls a "powerful psychological assumption"—the rules themselves cannot be taken as specifying the psychological mechanisms by means of which these input-output relations are effected.

Lacking such an assumption, Collins' claim reduces to the banal observation that whatever "all humans share *qua* an organism that can acquire any human language" must produce languages that we antecedently know to have various structural features. On this interpretation, the "cognitive turn" is far less exciting than it has been made out to seem, both in academic journals and in the popular press (e.g., Pinker 1994). For, it is no surprise whatsoever to learn that there is *something* inside of each human being that accounts for whatever universal linguistic features there are (if, indeed, there are any). The claim is even less exciting when, upon reflection, it emerges that this shared trait—what Collins dubs the "trivial" property P—is not (yet) constrained in any independent way.[9] Thus, for all we have been told thus far, P may be a very long way away from a Universal Grammar (UG)—i.e., the rich structure of highly specific, abstract, innate linguistic principles that are the hallmarks of the cognitive turn. For this reason, Collins' equation of P with UG is unmotivated. Specifically, the consequent of the following inference is a *non sequitur*.

> Since, … general linguistic theory describes the common resources of the grammars and there must be something common to all humans as acquirers of language, it looks as if general linguistic theory was all along an account of a universal human cognitive feature, that is, UG. (p. 86)

It's of course true that "general linguistic theory describes the common resources of … grammars" and it's likewise true that "there must be something common to all humans as acquirers of language." (Devitt [2006a: p. 267] calls this the "boring

[8] "[W]e can say that a competence and its processing rules must 'respect' the nature of the appropriate output in that, performance errors aside, the processing rules must produce outputs that have that nature" (Devitt 2006a, b: p. 22)

[9] "Trivially, then, there is some property P of the human cognitive design that allows each of us to acquire any given language. If all of this is so, then it would appear that inquiry into what each language shares is a substantial pursuit, for we must all share something given that we can acquire any language, notwithstanding the many apparent differences between any two languages" (Collins 2008a, b: p. 85). See also Chomsky (1986: p. 17).

innateness thesis.") But it simply *does not follow*, and it probably is not true, that every linguistic universal that we ever discover—if, indeed, we discover any[10]—must automatically be seen as being encoded in the human genome, represented in the minds of competent speakers, or involved in language acquisition. Regarding the latter claim, Cowie (1997) writes:

> [I]t is in general false that theorizing about a thing's essential properties is the same enterprise as theorizing about what we need to know in order successfully to learn about that thing. Biologists may worry about what makes cats cats, but a child's grip on cathood predates her excursions into zoology. Philosophers attempt to uncover the essential properties of persons, but it is not your knowledge of that essence that is helping you deal with your boss. Mechanics, perhaps, theorize about the properties shared by all internal combustion engines, but no cognizance of those universals, thankfully, is necessary for our learning to drive. And linguists, finally, theorize about UG, the essence of natural languages. But just as was the case with respect to cats and persons and cars, it is a very substantial additional claim that UG describes not just languages' essential properties, but also what we need to know in order for language acquisition to occur. (pp. 27–28)

The fact is that, for any putative linguistic universal, there are numerous possible explanations. Methodologically, we must first get clear about what universals there are—or even whether there are any—and then examine each one, case by case, proposing and (dis)confirming competing genetic, environmental, social, and psychological explanations.[11] An *a priori* commitment to the cognitive conception is, as Devitt argues, simply out of place in an empirical discipline.

It is, moreover, important to note in this context that, at present, we don't have any *independent* idea of what exactly it is that "all humans share qua an organism that can acquire any human language." Pending some independent grip on this prop-

---

[10] It is noteworthy that compelling arguments have recently been advanced for the conclusion that the sustained search for linguistic universals has come up entirely empty-handed. See Evans and Levinson, "The myth of language universals: Language diversity and its importance for cognitive science," in *Behavioral and Brain Sciences* (2009).

[11] Collins provides only a couple of examples of actual linguistic universals (p. 85). First, he cites the possibility of ambiguity, and the need for transformations to explain it. Another universal is "creativity"—the capacity for indefinitely novel production and comprehension. These are notably less sexy than the universals whose discovery is sometimes touted as a shining achievement of the Chomskyan approach—e.g. X-bar theory or the Head-Movement Constraint (both of which, incidentally, are now under fire from recent developments in the Minimalist program). The universals that Collins mentions are quite general, and thus impose very modest constraints on grammars. Correspondingly, they provide very weak support for Chomsky's innateness thesis. That these constraints rule out some very anemic grammars is, I suppose, somewhat interesting, but it shouldn't blind us to the fact that the constraints tell us very little about the innate endowment that the child brings to bear on the acquisition task. Indeed, if our cognition of *non*linguistic domains can *also* be captured by recursive formalisms—a not wholly implausible proposition (consider our understanding of kinship and other social relations, our mathematical competence, our tonal and rhythmic competence, etc.)—then this particular constraint doesn't even militate in favor of independent *language* faculty. In such a hypothetical-but-not-inconceivable case, the constraints would tell us only about general cognitive development, not language acquisition *per se*. More generally, only the existence of very *specific* universals can be marshaled as evidence for a robust innateness thesis.

erty, P, it is vacuous to claim that some grammar "is [or is not] explicable as arising from" P. Consider again Collins' claim:

> In general, then, given a grammar G and language L, we may take G to be descriptively adequate of L if it explains the various systematic features of our knowledge or understanding of L via an assignment of structural descriptions to L sentences. G itself cannot be explanatorily adequate, but *G is not adequate unless we are able to show that it is an acquirable grammar from UG and data to which the child is likely to be exposed*. (p. 86; emphasis mine)

If we don't know what P (or UG) *is*, then what sense can we make of the assertion that G_1 is explicable as arising from it, but G_2 is not? Until such time as something substantive is known *independently* about our innate endowment—in distinctly psychological or computational terms, rather than merely biochemical or anatomic terms—the only way we have of choosing between G_1 and G_2 is by comparing how they fit with the ongoing research into the grammar of *other* languages. This vindicates the aforementioned "innocent" methodological approach of aiming at explanatory adequacy—i.e., at a maximally simple, general, and unified theory of human language *as such*—but it does nothing to warrant a cognitive conception of language.

To put the point slightly differently, consider the fact that the only evidence that one might produce for the claim that UG has structure S rather than S* would be a discovery of some linguistic universal. Consider, furthermore, that we have no way of determining what linguistic universals there are except by constructing grammars for a variety of languages and checking whether the constructs employed by the grammar of L are applicable to another language, L*. Given that there is no evidence for the structure of UG besides what we garner from the innocent methodology of devising and comparing the grammars of various languages, and given that this methodology does *not* presuppose or require a cognitivist conception of language, Collins's argument for taking the "cognitive turn" does not work. What Collins needs, but does not have, is some *independent* ground—beyond the reasonable aspiration toward explanatory adequacy—for thinking that UG is *in the mind*. Pending a reason to think that UG is a distinctly psychological faculty, mechanism, process, or state, we are warranted in conceiving of UG as simply the linguists' working hypothesis about the structure of human language as such.[12]

I have argued that the "innocent" methodological approach is innocent precisely because it is equally consistent with both the cognitive construal of linguistics and its denial. My point, to reiterate, is that this fact alone is sufficient to undercut

[12] See also Devitt (2006a: Chap. 12) for a discussion of various arguments for the innateness of a mentally represented UG. On Devitt's view, an adequate grammar must be *true of* the external representational system that is a language. Moreover, he holds that human languages, like the representational systems of bees, prairie dogs, and other creatures, are acquired on the basis of innate constraints. It follows that an adequate grammar will not ascribe properties to a language that would make it impossible to acquire on this basis. Hence, an independently confirmed theory of the innate constraints can help the linguist narrow down the space of adequate grammars for a language. Devitt claims, however, that our present-day understanding of the innate constraints is poor. I echo this claim in the final pages of this discussion.

Collins's strategy of presenting the virtues of that methodology as somehow *supporting* the cognitive construal of generative grammar. Collins cannot make this move without begging the question against those who, like Devitt, are antecedently skeptical of the psychological reality of the linguists' theoretical constructs. Adopting the aim of explanatory adequacy and the search for linguistic universals does not force upon us a cognitive interpretation of linguistics, precisely because that methodological aim is common ground to both cognitivists and their contemporary opponents.

Before leaving this topic, I must note Collins' gesture, at the end of the long passage quoted above, toward precisely the sort of argument that I am claiming he needs.

> Explanatory adequacy relates UG to particular grammars; the condition distinguishes between, as it were, possible languages from 'arbitrary symbol systems' (*Aspects of the Theory of Syntax*, ch. 1). In sum, linguistic theory becomes an account of human *knowledge of language*. Particular descriptively adequate grammars constitute theories of what speakers know when they know a language, and UG is what speakers know in virtue of them being able to acquire any given language, that is, to represent a given grammar for their language. *It should be noted that Chomsky (LSLT, p. 62) has always been happy to treat language as an object of knowledge and so a psychological phenomenon.* What is new with the work of the early 1960's is the explicit construal of linguistic theory as having psychological states of knowledge as their [sic] object. (p. 86, emphasis added)

The italicized inference is, I claim, a bad one. The fact that something is an "object of knowledge" has *no bearing at all* on whether that thing is a cognitive state, or a psychological faculty, mechanism, or process. If I know a great deli in New York—i.e., if the deli is an object of my knowledge—it simply does not follow that the deli is a psychological entity.

4.4 In Practice, Do Acquisition Results Constrain Syntactic Theorizing?

What should we make of the fact that so many linguists in the generative tradition would balk at the mere suggestion that their object of study is something other than the mind/brain? My hunch is that syntacticians working in the Chomskyan paradigm are sometimes inspired to formulate hypotheses about the structure of language while taking themselves to be tackling questions about language acquisition. Such hypotheses are typically fruitful, lending support to the idea that the object of their inquiry is the acquisition process. Nevertheless, if we focus instead on *justification* of such hypotheses, rather than their genesis in the mind of a theorist, then we find that what we currently know about language acquisition is neither necessary nor sufficient to establish their truth. Not sufficient, because so much remains unknown about the neurocomputational mechanisms underlying acquisition; not necessary, because the super-empirical virtues of a grammar typically suffice.

Moreover, the theoretical constructs of formal syntax would, I submit, be revelatory even if (contrary to fact) we were certain that they play little or no role in the acquisition process. Take, for instance, the core "modules" in many contemporary P&P grammars—the projection principle, X-bar theory, θ-theory, Case theory, and the like. The explanatory value of these theoretical tools would remain undiminished if (again, contrary to fact) we *knew* that language acquisition were a matter of purely statistical pattern recognition in a connectionist network, or something of that nature. They would still be illuminating *to the formal syntactician*, offering an insight into the nature of a language, even if not the means by which it is acquired.

Focusing on a specific example will help us here. Consider the motivations for preferring binary-branching phrase structures, like those we find in X-bar theory. Haegeman (1994) surveys a large number of purely syntactic arguments for the X-bar formalism—i.e., arguments that appeal to the results of the familiar battery of constituency tests, such as pro-form replacement, topicalization, and coordination. And she notes that binary-branching grammars rank high in simplicity and generality. But she goes on to argue that binary-branching phrase structures are *also* preferable from the perspective of the acquisition theorist.

> In the course of this chapter, the change from [flat] structure to [binary-branching X-bar] structure was motivated on empirical and theoretical grounds, but there are further advantages to adopting a grammar which allows only the second type. The reader may notice that such a grammar is more aesthetically satisfying, though aesthetics may be a minor preoccupation for linguists. A grammar which allows only binary branching nodes is more constrained than a grammar which freely allows any type of branching node: in the former type of grammar lots of imaginable representations are ruled out in principle. A more constrained grammar is preferred for reasons of economy and elegance and it will also be preferred if we think of the ultimate goal of linguistic theories in the generative tradition. Remember that linguists wish to account for the fact that children acquire language very fast and at an early age. In order to explain their fast acquisition we posit that children are genetically prepared for the task, that they have an innate set of principles which enable them to construct the core grammar of their language on the basis of the data they are exposed to. One component of the child's internalized knowledge of the language, the internal grammar, will concern phrase structure. Theories of phrase structure such as X'-theory attempt to represent the native speaker's internal knowledge of phrase structure. … A child equipped with a UG that implements only binary branching will have fewer decisions to make when assigning syntactic structure to the data he is exposed to than a child equipped with a less constrained UG which allows ternary or four-way branching. … If the ultimate goal of our grammar is to account for language acquisition, then it will be natural to aim for the more restricted type of grammar in which fewer decisions have to be made by the child. Fewer choices will automatically mean more speed in the construction of the core grammar of the language acquired. Nowadays most linguists working in the generative tradition tend to adopt some version of the binary branching framework. (Haegeman 1994: pp. 138–143)

Haegeman's argument in this passage assumes that we know quite a bit about the processes underlying language acquisition. Unfortunately, the fact is that what is currently *known* about this is less than the argument requires. In particular, we do not know whether, given the child's *actual* learning strategy, a binary-branching grammar presents a substantially easier or more tractable computational problem for the learner. Our present-day understanding of the computational architecture of the brain is far too limited for us to be justified in making firm pronouncements on the matter—much less invoking such pronouncements as premises in arguments for substantive conclusions about the structure of a language. This is not, of course, to suggest that substantial work has not been done in acquisition theory, or that important results have not been garnered (Guasti 2002). It is merely to point out that acquisition models do not, at present, dictate a choice of grammar; the situation is quite the reverse.[13]

While many linguists in the generative tradition are unimpressed by these considerations, there *are* a handful of theorists who take them quite seriously, and pitch their inquiry and their results accordingly. The most prominent example, both in terms of impassioned rhetoric and substantial theoretical results, is the group of linguists who first proposed the feature-based unification formalisms of Generalized Phrase-Structure Grammar (GPSG) and Head-Drive Phrase-Structure Grammar (HPSG). In laying out their metatheoretic commitments, these linguists make absolutely clear that, while the results of their inquiry may well inspire or suggest specific hypotheses for psycholinguistic research, this would be but a serendipitous bonus. The suggestion seems to be that the viability or superiority of GPSG (and later HPSG) is established on grounds that have only a tenuous connection with psychology. I quote their remarks on this topic at length.

> In view of the fact that the packaging and public relations of much recent linguistic theory involves constant reference to questions of psychology, particularly in association with language acquisition, it is appropriate for us to make a few remarks here about the connection between the claims we make and issues in the psychology of language. We make no claims, naturally enough, that our grammatical theory is *eo ipso* a psychological theory. Our grammar of English is not a theory of how speakers think up things to say and put them into words. Our general linguistic theory is not a theory of how a child abstracts from the surrounding hubbub of linguistic and nonlinguistic noises enough evidence to gain a mental grasp of the structure of a natural language. Nor is it a biological theory of the structure of an as-yet-unidentified mental organ. It is irresponsible to claim otherwise for theories of this general sort. It may even be incoherent… Thus we feel it possible, and arguably proper, for a linguist (*qua* linguist) to ignore matters of psychology. But it is hardly possible for a psycholinguist to ignore language. And since a given linguistic theory will make specific claims about the nature of languages, it well in turn suggest specific kinds of psycholinguis-

[13] There is a strange double standard in Berwick and Weinberg's (1984) treatment of this issue. In the opening pages of Chap. 2, they make the following reasonable point about how parsing theory typically fails to constrain grammar construction: "We ought to be able to recruit sentence processing results to tell us something about what the grammar should look like. If we had some independently justified parsing model, we could reject grammars that were incompatible with it. In practice, though, because very little is known about the details of the syntactic parser, confidence in constraining the choice of grammatical theory via this route must be correspondingly weak. If the parsing theory has no independent motivation, we can always change it to suit the grammatical format" (p. 36). However, they go on to suggest that our understanding of the acquisition process is more detailed and hence provides more firm constraints on grammar constructions. But surely, their observation about the lack of fixed points in parsing theory holds just as well with regard to acquisition. Indeed, the only characteristics of acquisition that B&W bring to bear in their introductory chapter are these: (i) acquisition is accomplished in about 5 years, and (ii) the input to it is variable, partially ill formed, and rarely (if ever) provides negative evidence. This is not more than what we know about parsing. Yet these considerations were marshaled, in their first chapter, as the justification for profound shifts in syntactic theory.

> tic hypotheses. Stephen Crain and Janet Fodor have argued that [Generalized Phrase-Structure Grammar] does have implications for psycholinguistic concerns. Nonetheless, it seems to us that virtually all the work needed to redeem the promissory notes linguistics has issued to psychology over the past 25 years remains to be done. If linguistics is truly a branch of psychology (or even biology) as is often unilaterally asserted by linguists, it is so far the branch with the greatest pretensions and the fewest reliable results. The most useful course of action in this circumstance is probably not to engage in further programmatic posturing and self-congratulatory rhetoric of the sort that has characterized much linguistic work in recent years, but rather to attempt to fulfill some of the commitments made by generative grammar in respect of the provision of fully specified and precise theories of the nature of the languages that humans employ. Even when that is done, the psychology of language will doubtless have a vast amount of work to do before we have a scientific understanding of how the human species acquires and uses language. After all, geometrical optics long ago provided us with a fairly clear and stable means of characterizing the objects of visual perception, but the psychology of visual perception still has many problems to solve. So far, linguistics has not fulfilled its own side of the interdisciplinary bargain (Gazdar et al. 1985: p. 5).

These ideas are strongly echoed in the remarks of Black and Chiat (1981).

> Nothing would be changed if instead of distinguishing between different types of knowledge, we simply distinguished between different systems of rules and representations without making any assumptions about their psychological status. If we agree that it is legitimate to talk about abstract structures and properties, though they may not be identifiable with any particular physical property of strings of symbols or sounds, then the addition of terms like 'knowledge of' to grammar and 'mental' to rules and representations is simply a matter of theoretical goals and interests. There is no evidence that the practice of theoretical linguistics would be changed if a 'realistic' interpretation of the grammar was assumed, rather than a 'psychologically realistic' one (cf. Milner 1979). (p. 40)

Other such statements are scattered throughout the literature, suggesting that there is a contingent of linguists who have reflected on the aims and claims of linguistic theory and have come to conclusions that diverge sharply from the main tenets of Chomsky's cognitive conception. This, too, bolsters the contention that the cognitive conception is not the only game in town.

4.5 Conclusion

In this chapter and the last, I considered a few of the main arguments that constitute the debate between cognitivists like Chomsky and opponents like Devitt. One major difference, we've seen, is that the cognitivist takes grammatical rules and principles to be psychologically real—indeed, trivially so—whereas Devitt argues (2006a, b: Chap. 11) that they are not. While this difference is impressive, I have taken pains to distinguish the psychological reality debate from the debate about how to conceive of the objects of linguistic theorizing. Though the two plainly bear on one another, they are logically distinct. If the cognitive conception is correct, then grammars are psychologically real, but the reverse entailment does not hold. A grammar can be psychologically real even if Devitt's "linguistic realism" is correct and the objects of the formal syntactician's concern are public, conventional E-languages. A grammar designed for the purpose of describing the structure of an E-language can nevertheless find a home in psycholinguistic theories, provided that the latter enjoy independent empirical support.

With regard to language comprehension and production, Chomsky's view entails that psycholinguists should seek a relatively transparent relation between the syntacticians' grammar and the "knowledge-base" or "data-structure" that is used by the human sentence processor. True to this vision, the methodological hope of those working on sentence processing is to find a natural grammar-parser combination. The progress toward this goal has been rather haphazard, in part because syntacticians are frequently not as concerned with fine-grained psychological data as one would expect them to be if their goal was to construct psychologically real grammars. A survey of the mainstream literature in generative syntax reveals, I think, that psychological reality is a distant, dimly understood, and rarely invoked desideratum. Indeed, it is a notorious gripe in psycholinguistics that syntacticians do not construct grammars with the intent of making life simpler for parsing theorists (Frazier 1988). For one thing, formal grammars come in and out of fashion by the decade. Each minor revolution forces the psycholinguist to revise her assessment of the relevance of prior experimental results, and to devise novel experimental techniques that address issues raised by each new formalism.[14] Moreover, syntacticians rely heavily on formal simplicity as a criterion for selecting a grammar. But, as Stabler (1984) argues, it is rather unlikely that this criterion will converge with the psycholinguist's aim of psychological plausibility.

Still, while Devitt and many others are deeply skeptical of the psychological reality of grammar, there are, I believe, *methodological* grounds for pursuing this hypothesis. As Berwick and Weinberg (1984: pp. 38–40) point out, a parsing theory that makes direct use[15] of the syntactic principles shored up by the formal syntactician is the simplest and hence the strongest theory that one could adopt. Hence, while we should not expect, let alone presume, that we will find it to be true, we should do our best to subject it to empirical scrutiny *first*, so as to disconfirm it before moving on to weaker hypotheses. As I will report in subsequent chapters, this effort has been underway for over four decades and the results that are now available have not, I think, *contravened* this strong hypothesis, though they have necessitated a much more careful statement of it. All in all, though, the hypothesis that the syntactician's grammar is psychologically real has fared rather well.[16] There is now

[14] The rolling revolution in syntactic theory also makes trouble for *syntacticians*. Ludlow (2011: p. 29) discusses how fans of the Minimalist Program were forced to sacrifice the "sweet results" that issued from the PRO theorem, when that theorem was abandoned by Chomsky (1995)—a case of what Ludlow quite aptly labels "Kuhn loss."

[15] The term 'direct use' is deliberately vague, eliding the distinction between the declarative representation and the procedural embodiment of a grammar. This distinction will take center stage in Chaps. 7, 8 and 9.

[16] Talk of "*the* syntactician's grammar" is misleading. There are many substantively distinct grammars on offer, e.g., Principles and Parameters grammars, Lexical Functional Grammar, and Generalized Phrase Structure Grammar, to name just a few. And there is even more variety if we consider the theoretically irrelevant but computationally significant notational variants of each of these grammars. See Chaps. 8 and 9 for examples.

more reason than ever to believe that developments in psycholinguistics will reveal the role that grammatical rules and principles play in the psychology of a competent language user.

In the chapters to come, I examine the way in which detailed models of parsing and comprehension appeal to one or another grammar in explaining a broad range of behavioral and neurocognitive data. The remainder of the present work can be seen as an exploration of the relevant findings in this area and a sustained argument for the psychological reality of grammar. I begin in the following chapter by reviewing the major lines of evidence for the psychological reality of phrase-structure representations.

Chapter 5
Mental Phrase Markers in Sentence Processing

Abstract I marshal several lines of empirical support for the claim that the human sentence processing mechanism (HSPM) constructs representations of the syntactic structures of linguistic stimuli—what I call "mental phrase markers" (MPMs). Powerful neurocognitive evidence for this hypothesis is drawn from recent EEG and MEG studies. Further support comes from studies of structural priming and garden-path processing, which provide insight into the structure of MPMs. Structural priming involves modulating the speed of behavioral responses by exciting certain MPMs prior to a task. In the case of garden-path processing, the HSPM encounters a locally ambiguous input and resolves the ambiguity in a way that turns out to be incorrect. The principles of ambiguity resolution that are operative in such cases all seem to make direct ineliminable reference to MPMs. Finally, I discuss various attempts to demonstrate the psychological reality of so-called "empty categories". The available evidence suggests that *wh*-traces are psychologically real and that the HSPM employs sophisticated strategies in searching for them, making use of both grammatical constraints and cues provided by their antecedents. In the concluding section, I discuss whether the contents of MPMs can help us decide which grammar the HSPM employs in on-line processing.

Keywords Electroencephalography • EEG • Magnetoencephalography • MEG • P600 • N400 • Early left anterior negativity • ELAN • Mental phrase markers • Neurolinguistics • Violation paradigm • Structural priming • Garden-path sentences • Garden-path processing • Syntax-first models • Ambiguity • Ambiguity resolution • Minimal attachment • Late Closure • The Minimal Chain Principle • Eye-tracking • Cross-modal priming • Empty categories • *wh*-traces • NP-traces • PRO • Filler-gap processing • The filler-gap effect • Human sentence processing mechanism • HSPM • D-structure • Structure • Active vs. passive structures • VP-elision • Accusative case • Nominative case • The empty category principle • Incremental processing • Click experiments • Fodor Bever and Garrett (FBG) • Double-object dative constructions • Reanalysis • Consciousness • Lexical functional grammar • LFG • Head-driven phrase-structure grammar • HPSG • Island constraints • Visual probe recognition • VPR • Syntactic processing • Semantic processing • Government and binding theory • GB

D. Pereplyotchik, *Psychosyntax*, Philosophical Studies Series 129,
DOI 10.1007/978-3-319-60066-6_5

5.1 Introduction

My primary goal in this book is to assess the bearing of syntactic theory on the psychology of language. In the previous chapters, I've argued that there are substantive questions about whether the theoretical constructs of formal linguistics play any role in the psychological processes underlying language use—and, if so, which. In the remainder of this book, I will argue that the rules or principles of a grammar are indeed psychologically real, though perhaps not in the way that some theorists have assumed. Defending this claim will be the burden of Chap. 9. But the rules and principles of grammar are not the only theoretical construct of interest. In this chapter and the next, I aim to show that the human sentence processing mechanism (henceforth, the HSPM) constructs what I will call "*mental phrase markers*".

As I use the term, a *mental phrase marker* is a representation of the syntactic structure of a linguistic stimulus. The claim that the HSPM constructs mental phrase markers in the course of comprehension underlies a great deal of work in psycholinguistics. Here are three typical statements of it by leading figures in the field:

> [L]et us suppose (as we surely should, until or unless the facts dictate against it) that the human sentence processing routines compute for a sentence the very structure that is assigned to it by the mental "competence" grammar. (Fodor 1989: p. 157).

> [W]e assume in the first place [the] [w]eak competence hypothesis: Human syntactic analysis typically involves the explicit recognition of all grammatical relations. (Stabler 1994: p. 303)

> Most models of human language comprehension assume that the processor incorporates words into a grammatical analysis as soon as they are encountered. … We assume that sentence processing involves the computation of dependencies between the words and phrases that are encountered. For example, in the sentence *The troops found the enemy spy*, the relations include information that *the troops* is the subject of *found*. Often, words are incorporated directly into the representation without breaking existing dependencies. For example, when the main verb *found* is encountered, the processor forms the dependency between *the troops* and *found*, and does not need to break any other dependencies (e.g., that between *the* and *troops*). (Sturt et al. 2001: p. 283)

While certainly popular, the assumption that the HSPM constructs mental phrase markers is not without its detractors. Consider the following passages, drawn from work in philosophy, psycholinguistics, and artificial intelligence.

> [L]anguage use [is] a fairly brute-causal associationist process … rather than a process involving metalinguistic representations of the syntactic and semantic properties of linguistic expressions. (Devitt 2006a, b: p. 220)

> [N]o independent level of syntactic representation is constructed, operated on, or output by the language analysis process. (Schank and Birnbaum 1984: p. 220)

> [I]t isn't even clear that we do parse, in the sense of constructing an explicit and complete representation of the structure of a sentence in the course of comprehending it. The only language tasks that we really know humans are able to perform are those we can directly observe, including comprehension, production, repetition, and so forth. That mapping from an acoustic signal to a message occurs via deep structure, or some other hierarchical syn-

> tactic representation, is merely a convenient assumption—a pervasive and often useful one, but an assumption none the less. In advancing our understanding of the human language processing system, we may do well to question this assumption. (Rohde 2002: p. 6–7) … There seems to be little direct evidence that we construct a representation of a syntactic parse tree while comprehending. (p. 18)

In what follows, I present a number of arguments for the psychological reality of mental phrase markers (henceforth, MPMs), relying on behavioral and neurocognitive data from psycholinguistics. As Fodor et al. (1974) put it, "detailed empirical constraints can and should be placed upon [parsing] models. Any theory of sentence recognition must be committed to an array of testable predictions" (p. 319). In the standard case, such predictions will be concerned with the speed and accuracy of verbal reports and nonverbal behaviors (button presses, eye movements), as well as the neuronal activity that underpins language processing. Many ingenious experimental designs have been used to study how written or spoken language is processed.[1] A subject in a typical psycholinguistic experiment is given a task designed to test whether some controlled variable affects her ability to recognize and categorize an expression, to disambiguate it, to recall it, to comprehend its meaning, to make inferences from its meaning to other information, and so on (Fernandez and Cairns 2011). Such tasks can be performed in non-optimal conditions, with degraded or highly ambiguous stimuli, and in conjunction with ancillary task demands. The completion of a task is signaled either by a behavioral response or a by a neurophysiological signal. The results of such experiments constrain models of comprehension; a model that stays faithful to the available behavioral and neurocognitive data has, *ipso facto*, a stronger claim to psychological plausibility.[2]

Since the inception of the cognitivist study of language, untold numbers of experiments have been performed, many of them geared toward establishing the psychological reality of one or another aspect of grammar. Plainly, a comprehensive review of the relevant findings will not be possible in the space of a chapter. In what follows, then, I examine a small but suggestive sample of the results, focusing in large part on studies of comprehension.[3]

---

[1] Studies of sign language processing are quite rare, but they do exist. See, e.g., Bavelier, D., Corina, D. P., and Neville, H. J. (1998). "Brain and Language: A Perspective from Sign Language Properties of ASL," *Neuron*, Vol. 21, pp. 275–278, and references therein.

[2] It perhaps goes without saying that an insistence on the primacy of behavioral and neural data is not an expression of behaviorist or crypto-eliminativist principles, particularly as my conclusion is that certain processes and representations are *psychologically real*. If insisting on the primacy of third-person data were a sign of an underlying philosophical agenda, it would be anti-Cartesianism—a methodology that de-emphasizes wanton intuition mongering and the deliverances of consciousness and introspection. The fruitlessness of Cartesian theorizing is apparent not only from the historical failure of nineteenth-century introspectionist psychology (Lyons 1986; Kukla and Walmsley 2006), but also from the pitfalls that psycholinguistics encounters when relying on introspection (Kartunnen and Zwicky 1985: p. 22; Tanenhaus et al. 1985).

[3] Reflecting a bias that pervades the present work—indeed, the entire field of psycholinguistics—I focus largely on comprehension. In production, the inputs to real-time computations are presumably what philosophers call "propositional attitudes"—judgments, desires, intentions, and so forth. Chomsky often refers to the inputs as "states of the conceptual-intentional system," leaving open

I begin by taking a brief look at what is currently known about the neural processes underlying language comprehension. In the 1990s, results emerging from a number of labs suggested that a phase of distinctly syntactic processing can be identified via neurophysiological measures. Subsequent work has shown that the matter is considerably more complicated. The present state of the field is characterized by healthy debates about the proper interpretation of various neural markers, particularly the P600 and the N400. It is too early to tell how such debates will pan out, but I shall argue that the psychological reality of mental phrase markers is either consistent with, presupposed by, or supported by the available findings.

Turning to behavioral studies in Sect. 5.3, I discuss the results of a number of experiments designed to examine a phenomenon known as structural priming. While these studies provide strong support for claim that the HSPM constructs MPMs, I also review some evidence that casts doubt on the claim that the HSPM constructs the D-structure representations posited by the Government and Binding theory (Haegeman 1994). The evidence suggests, rather, that the HSPM computes S-structures directly.[4]

In Sect. 5.4, I turn to the psycholinguistic experiments that study what are known as "garden-path" phenomena in sentence processing. These are cases in which the HSPM encounters a locally ambiguous input and resolves the ambiguity in a way that turns out to be incorrect relative to the completion of the sentence or phrase. I discuss several principles of ambiguity resolution, such as Minimal Attachment and Late Closure, which play a role in many psychologically plausible parsing models.

Section 5.5 is devoted to a survey of the fascinating attempts to demonstrate experimentally the presence of empty categories—specifically, *wh*-traces and NP-traces—in the mental phrase markers that the HSPM constructs. The available evidence suggests that *wh*-traces are psychologically real and that the HSPM employs rather sophisticated strategies in searching for them, making use of the cues provided by their antecedents, as well as considerable knowledge of grammatical constraints. I discuss the implications of this fact for the psychological reality of syntactic rules and principles—in particular, whether the contents of MPMs can help us decide which grammar the HSPM employs in on-line processing. Finally, I review some evidence for the presence of NP-traces in mental phrase markers.

---

the relation between these states and propositional attitudes. At present, the issues surrounding linguistic production are relatively murky, though work by Kay Bock and others has yielded important advances in our understanding. It is, moreover, an open question what exactly constitutes the *output* of comprehension. In general, empirical issues concerning non-peripheral cognitive states—the propositional attitudes or states of the "conceptual-intentional system"—are less tractable than one would have hoped.

[4] Many theorists were persuaded by the arguments in Fodor et al. (1974) that the transformations posited by a variety of syntactic theories are not psychologically real operations. Fodor, Bever, and Garrett argued against what they called the derivational theory of complexity (DTC), on the basis of a number of experimental findings. Following Berwick and Weinberg (1984: ch. 2), Phillips (1994, 1996) and Phillips and Lewis (2013), I doubt that the findings do in fact disconfirm the DTC. (Phillips and Lewis write: "We do not mean to claim that the DTC was substantiated, but the reports of its defeat strike us as somewhat stylized history.") Nevertheless, I am aware of no empirical grounds for positing D-structure representations in the course of sentence processing.

5.2 Neurocognitive Findings

Neurolinguistics is evolving rapidly, so we must be tentative in drawing conclusions from the available findings. Nevertheless, it would be unwise to ignore a range of data that strongly suggests that MPMs are psychologically real. In arguing for the even stronger claim that "real-time processes assemble syntactic representations that are *the same as those motivated by grammatical analysis*" (emphasis mine), Phillips and Lewis (2013) write:

> [S]tudies that use highly time-sensitive measures such as event-related brain potentials (ERPs) have made it possible to track how quickly comprehenders are able to detect different types of anomaly in the linguistic input. This work has shown that speakers detect just about any linguistic anomaly within a few hundred milliseconds of the anomaly appearing in the input. Different types of grammatical anomalies elicit one or more from among a family of different ERP components, including an (early) left anterior negativity ('(e)LAN'; Neville et al. 1991; Friederici et al. 1993) or a P600 (Osterhout and Holcomb 1992; Hagoort et al. 1993). Many questions remain about what the different components reflect and what determines which components are evoked in any individual situation (Hagoort 2003; Friederici and Weissenborn 2007; Lau et al. 2008; Gouvea et al. 2009), but for current purposes the most relevant outcome from this research is that more or less any grammatical anomaly elicits an ERP response within a few hundred milliseconds. If the on-line analyzer is able to immediately detect any grammatical anomaly that it encounters, then it is reasonable to assume that it is constructing representations that include sufficient grammatical detail to detect those anomalies.

In fact, it is no exaggeration to say that virtually *all* recent work in neurolinguistics presupposes not only that the HSPM constructs MPMs, but that it does so in consultation with a mentally represented grammar. In a text that provides a comprehensive review of the existing work on the neural underpinnings of language processing, Bornkessel-Schlesewsky and Schlesewsky (2009) begin by outlining the conditions that must be satisfied by any psychologically plausible model of comprehension.

> Properties of the processing system required by incremental comprehension: The processing system must impose a structure on the input as quickly as possible. … [It must then use] the structure/meaning assigned to generate predictions about the incoming input. … Structure building must clearly be constrained by the grammar assumed for the language being processed. Thus, the grammar constrains the types of structures that can be constructed from the input. (Bornkessel-Schlesewsky and Schlesewsky 2009: p. 90)

In the next several sections of this chapter, we will review the evidence for the claim that comprehension is "incremental," in the sense these authors intend.

A number of techniques are currently used in obtaining recordings of neural activity. These include electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and transcranial magnetic stimulation (TMS). Bornkessel-Schlesewsky and Schlesewsky (2009) provide an in-depth analysis of the strengths and weaknesses of these techniques. The most salient contrasts between them concern the accuracy that they achieve in measuring the speed of processing and in discerning the precise location at which processing occurs. EEG and MEG recordings are time-locked to the presentation of a stimulus and are capable of yielding impressively precise measurements of the time at which a neurophysiological response occurs, with an accuracy of up to 4 milliseconds. Unfortunately, the cost is a marked *im*precision with respect to the location of the source of any such signal. By contrast, recordings from fMRI devices yield quite precise data concerning the location of a neurophysiological response, but leave open a window of up to several seconds during which the response might have occurred. Our primary concern here is to establish the existence of a processing phase at which the HSPM is sensitive only to syntactic properties. As such, we do well to focus on the EEG data, which has provided the lion's share of the evidence bearing on this issue.

EEG devices measure what are known as *event-related potentials* (ERPs), also sometimes called *evoked potentials*—small voltage differences between electrical activities in the brain, recorded by electrodes placed on the scalp. ERP data are multi-dimensional, providing information about the degree, polarity, and location of neuronal activations. Because ERPs are time-locked to a stimulus event, they admit of functional interpretation, by reference to the properties of the stimuli that evoke them.

In studies of linguistic comprehension, a typical piece of EEG data will specify the latency, degree, polarity, location, and distal cause of an ERP, in a graph like that in Fig. 5.1. Here, we see a comparison between the electrical activity evoked by two German sentences, one grammatical, the other not. At the top left, the gross location of the activity is specified ('F7'). The values on the y-axis represent the degree of the signal and its polarity (positive or negative). The values on the x-axis represent the signal's latency—i.e., time at which it occurs, relative to the onset of the stimulus.

The most widely used experimental paradigm in ERP studies is known as *the violation paradigm*. Participants are shown a variety of sentences, some of which contain violations with respect to one or another linguistic property. For instance, in an early ERP study, Neville et al. (1991) used the following materials:

(1)	***The man admired Don's <u>of</u> sketch the landscape.**	**syntactic violation**
(2)	***The man admired Don's <u>headache</u> of the landscape.**	**semantic/pragmatic violation**[5]
(3)	**The man admired Don's sketch of the landscape.**	**control sentence (no violation)**

As subjects read such sentences, an EEG device monitors their brain activity, particularly at the crucial regions, underlined in (1) and (2). The logic of the paradigm is straightforward: The ERPs associated with the anomalous stimuli differ significantly from those associated with stimuli that are free of linguistic violations. This in turn provides information about where and when in the brain specific kinds of

[5] In view of the notoriously shaky status of the semantics/pragmatics distinction, I simply slur over it in what follows, labeling various properties of sentences 'semantic' regardless of whether they would be classified as semantic or pragmatic by a theorist who insists on drawing the distinction.

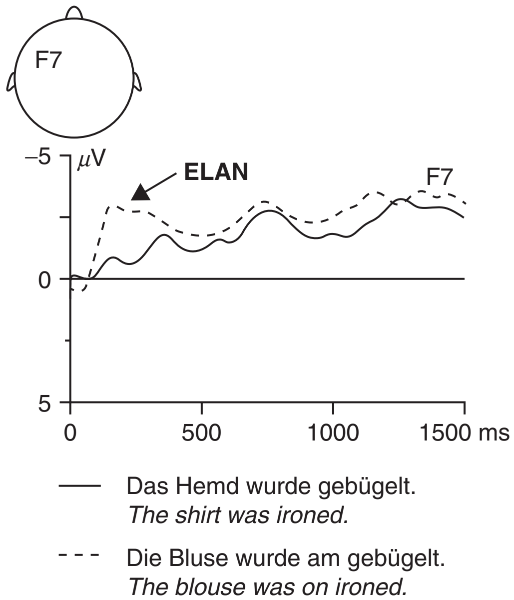


Fig. 5.1 A typical display of ERP data, showing the latency, degree, polarity, location, and distal cause of the neuronal signal. In this case, the signal is an ELAN—an early negativity in the left anterior region of the brain. As the graph shows, the ELAN occurs roughly 125 milliseconds after the onset of the critical stimulus. Like the studies discussed in the main text, the experiment from which this data was derived used two German sentences as stimuli. The first, indicated by the *unbroken line*, is a grammatical sentence. The second, indicated by the *broken line* is *un*grammatical—i.e., it exhibits a basic phrase structure violation. The critical stimulus is the word 'ironed'. The ERP associated with the grammatical sentence are significantly different from the one associated with the ungrammatical sentence. The graph shows a distinct negativity (conventionally plotted upward on y-axis) (*Source*: Bornkessel-Schlesewsky and Schlesewsky 2009: p. 110)

violation are represented. For instance, Neville *et al*. found that sentence (1) evokes a negative-polarity response in the left anterior region of the brain approximately 125 milliseconds after the onset of the word 'of', followed by a positive-polarity response several hundred milliseconds later. This fast negative-polarity response, depicted in Fig. 5.1, has come to be known as the ELAN—early left anterior negativity. By contrast, the non-syntactic violation in sentence (2) evokes a negative-polarity response approximately 400 milliseconds after the onset of the word 'headache'. This has been dubbed the N400.

In another early study, Friederici et al. (1993) found the same pattern—a replication that is especially striking given the fact that, unlike Neville *et al*., Friederici *et al.* used German rather than English sentences and presented them auditorily rather than visually. An sample of the materials from this study appear in (4)–(6).

(4)	***Der Freund wurde im besucht.** **the friend was in-the visited**	**syntactic violation**
(5)	***Die Wolke wurde begraben.** **the cloud was buried**	**semantic/pragmatic violation**
(6)	**Der Finder wurde belohnt.** **the finder was rewarded**	**control sentence (no violation)**

Virtually the same pattern has been observed in subsequent studies. The natural interpretation is that incoming words are incrementally incorporated into a mental phrase marker, with syntactic information being accessed quite early—125 milliseconds after stimulus onset—while other properties of the stimulus are recovered several hundred milliseconds later.

Bornkessel-Schlesewsky and Schlesewsky (2009) discuss a series of further studies designed to test whether the integration of syntactic information—i.e., the construction of a mental phrase marker—is, strictly speaking, a *prerequisite* for the integration of semantic information. This can be determined by constructing sentences in which the semantic violation is *combined* with the syntactic violation. Consider, for instance, sentence (7).

(7)	***Das Gewitter wurde im gebugelt.** **the thunderstorm was in-the <u>ironed</u>**	**combined syntactic *and*** **semantic violation**

What Hahne and Friederici (2002) found was that (7) evokes an ELAN, which is characteristic of syntactic violations, but not the N400, which seems to be correlated with semantic violations. It appears, then, that the (presumably syntactic) ELAN is capable of *blocking* the (presumably semantic) N400. Crucially, it has also been discovered that the reverse is *not* true—a "semantic" N400 evoked *prior* to a "syntactic" ELAN cannot "block" the ELAN.[6] Researchers have thus concluded that "existing ERP findings provide strong converging support for the assumption that constituent structure information *hierarchically dominates* other information types such as semantics/plausibility" (Bornkessel-Schlesewsky and Schlesewsky 2009: p. 113; emphasis in the original).

This kind of work also bears on a key issue that will resurface in Sect. 5.4, where we turn to the question of how the HSPM assigns structure in *locally ambiguous* circumstances—as in the famous "garden-path" cases, such as (8)–(12).

---

[6] The sudden rash of scare-quotes reflects the extreme caution with which we must proceed here. Further studies may well show that this categorization of the phenomena is in need of revision. For instance, Stroud and Phillips (2010) report an array of findings that cast doubt on the initial picture of the N400. None of the findings they report, however, seem to affect the argument here. For, they do not challenge the claim that the ELAN is an unambiguous reflection of distinctly syntactic violations. Furthermore, the refinements they suggest to our interpretation of the ERP signal known as the P600 do no violence to the analysis presented above.

(8) Jake tells students he intrigues to stay.
(9) Benny knows the boy hurried out the door slipped.
(10) The soldier persuaded the radical student that he was fighting in the war for to enlist.
(11) Aron gave the man who was eating the fudge.
(12) The old train the children.

It has been found such sentences cause various kinds of processing difficulty for ordinary hearers and readers. The work reviewed below will show that this is due to isolated failures of the HSPM's heuristics for resolving local ambiguities. When the HSPM's preferred structural assignment in the middle of the sentence turns out to be incorrect, additional computational load is incurred. On some processing models, the HSPM recovers syntactic information at the outset, and then performs an operation of *reanalysis* in light of more sophisticated semantic information. There is ongoing debate about whether this is correct. Bornkessel-Schlesewsky and Schlesewsky (2009) put the matter as follows:

> While the need for constituent (phrase) structure is undisputed in both theoretical linguistics and psycholinguistics, its precise role in real time comprehension is somewhat more controversial. A number of positions on this issue have been advocated, the two extremes of which can be described as follows: (a) constituent structuring is the prerequisite for all further processing steps, both syntactic and interpretive in nature, and (b) constituent structure is not accorded any special role but rather interacts with all other available information types (e.g. animacy, frequency, thematic roles) to determine sentence structure and interpretation. (p. 107)

The experiments discussed above bear on this issue. Indeed, early ERP studies were designed for the explicit purpose of testing such hypotheses, thus adding to the pool of behavioral data that we discuss in Sects. 5.4 and 5.5 below. Some key findings support those models on which purely syntactic information plays a privileged role in determining the HSPM's initial parsing preferences. For instance, Friederici (1995) conducted a series of studies in which the ERP data was found to conform to the predictions of so-called *syntax-first* models.

> The combined findings discussed here suggest that on-line structuring processes are subserved by brain systems located in the anterior part of the left hemisphere [DP: the site of the ELAN], whereas processes of structural reanalysis seem to involve different brain systems. … [W]e may take the available data to suggest a parser with two subcomponents, a first subcomponent responsible for the early structuring of the input seemingly working in a highly time-dependent procedural manner and a second subcomponent responsible for syntactic integration and reanalysis *consulting grammatical knowledge* which may be represented in a less time dependent form. With such an architecture, the syntactic processing system would (a) be fast in assigning structure to the incoming information and would (b) be most flexible in selecting the valid structure for adequate thematic role assignment and the ultimate interpretation. (Friederici 1995: p. 278; emphasis added)

More recently, Friederici and Brauer (2009) emphasize the continuing success of this model. Having reviewed the currently available homological comparisons with

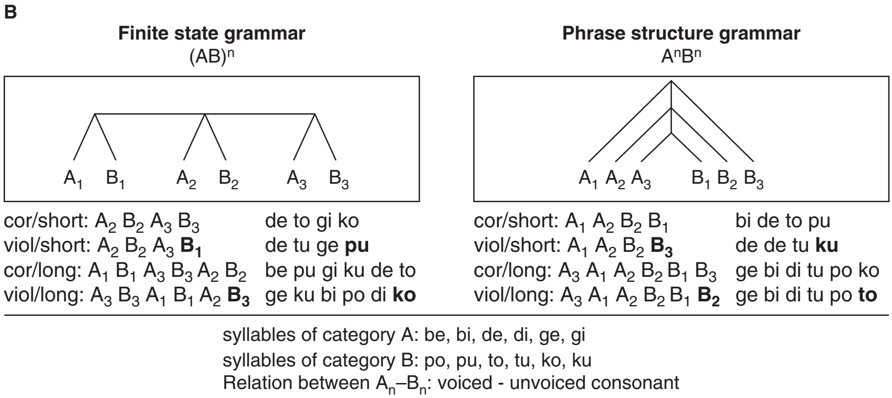


Fig. 5.2 The artificial grammars used for some of the ERP research reported in Friederici and Brauer (2009). Note that the examples marked ‘cor’ (which is short for ‘correct’) illustrate what each grammar can generate, while those marked ‘viol’ (short for ‘violated’) illustrate what they cannot generate

animal neuroanatomy as well as a number of lesion studies and child language processing studies, they conclude that

> [s]o far, empirical evidence supports the assumption of two separate functional brain networks underlying syntactic processes. Local phrase structure building involves the frontal operculum and the anterior part of the STG [i.e., the superior temporal gyrus], while the processing of complex hierarchical structures requires the participation of Broca’s area and very likely the posterior part of the STG. These functional networks correspond to structural connections in the brain between associated areas within each of these circuits. Relevant data from the non-human brain and from children support the view that the network responsible for complex syntax might have evolved late during phylogeny and only develops late during ontogeny. (p. 502)

Some of the studies they report strongly suggest that the relevant brain circuits are processing specifically *syntactic* information. This can be seen from the fact that the regions are activated in the course of processing sentences of *artificial* languages, such as the ones depicted in Fig. 5.2. There is no possibility of semantic processing in such cases, as the “words” are all nonsense syllables.

The latest development in this area of neurolinguistics came about with the publication of a study by Ding et al. (2016), which provides powerful evidence for the existence of “neural representations of abstract linguistic structures that are internally constructed on the basis of syntax alone” (163). The authors used MEG recordings to show that the brains of competent speakers respond to sequences of words in ways that track not only their acoustic, syllabic, and prosodic features, but also phrase-level groupings.

> … during listening to connected speech, cortical activity of different timescales concurrently tracked the time course of abstract linguistic structures at different hierarchical levels, such as words, phrases and sentences. Notably, the neural tracking of hierarchical linguistic structures was dissociated from the encoding of acoustic cues and from the pre-

> dictability of incoming words. Our results indicate that a hierarchy of neural processing timescales underlies grammar-based internal construction of hierarchical linguistic structure. (158)

As the authors' mention of "predictability" indicates, the studies controlled not only for the superficial properties of the stimulus stream, but also for more abstract statistical information about the transitional probabilities of both natural and artificial word sequences. The dissociation of hierarchical structure-building from statistical analysis does not imply that the latter plays no role in language processing, only that it is not the exclusive method employed by the HSPM. The authors clearly acknowledge this: "Although linguistic structure building can clearly benefit from prosodic or statistical cues, it can also be achieved purely on the basis of the listeners' grammatical knowledge" (158). More strongly, they point out that statistical cues "are not always available, and even when they are available, they are generally not sufficient" (163). (This topic will arise again in our discussion of probabilistic approaches to parsing in Chap. 8). Reflecting the gold-standard quality of these studies, Ding *et al*. conclude confidently that "[t]hese results underscore the undeniable existence of hierarchical structure building operations in language comprehension" (162), which they claim "relies on a listeners' tacit syntactic knowledge" (163).

5.2.1 Summary

Although the discussion above paints a tidy picture of the available data, it would be wrong to assume that there are no competing interpretations. For instance, Bornkessel-Schlesewsky and Schlesewsky (2009: ch. 12) discuss an array of data that suggests the need for a revision in our view of various ERP results. Ongoing debate among neurolinguists about the status of the ERP known as the P600 (a late positivity signal) serves as a sobering reminder of the need to be tentative in one's conclusions.[7] Still, we have seen reasons for thinking that the HSPM constructs mental phrase markers in the course of comprehension, and a powerful confirmation of this in the recent Ding et al. (2016) results. In the next section, we review a range of behavioral data in support of the same conclusion. Sections 5.4 and 5.5 expand on these findings to provide a clearer picture of the internal operations of the HSPM. Throughout the discussion, I will occasionally note empirical support for the claim, expressed in the above quotations from Friederici (1995) and Ding et al. (2016), that the HSPM performs its syntactic analyses in consultation with a mentally represented grammar.

[7] See also Batterink and Neville (2013); Featherstone *et al*. (2013); van Gaal *et al.* (2014).

5.3 The Argument from Structural Priming

Early behavioral studies of the psychological reality of syntactic constituency were rather primitive. Some amounted to little more than simply asking subjects which strings of words in a sentence seem to them to constitute natural groupings.[8] In their classic text, Fodor et al. (1974) reported a range of studies that they interpreted as demonstrating the psychological reality of mental phrase markers. Among these were the famous click experiments, in which subjects monitoring a speech stream misheard short clicks as if they occurred at constituent boundaries. However, as Jurafsky and Martin 2008: pp. 424–5) note, much of this evidence was weak.[9]

One issue is that many of these studies failed to control for semantic biases that correlate with syntactic structure. After all, many effects that can be explained by the hypothesis that subjects group words into a *syntactic* perceptual unit can equally well be explained by the hypothesis that the grouping is a *semantic* one. Convincing arguments for the psychological reality of syntactic constituency must, therefore, control for semantic confounds. Jurafsky and Martin point out, furthermore, that "since there are many non-constituent-based theories of grammar based on lexical dependencies, it is important to find evidence that cannot be interpreted as a *lexical* fact; that is, evidence for constituency that is not based on particular words." Recent evidence from priming studies fits the bill.

The logic of priming experiments is this: A mental representation, once activated—say, in response to a stimulus—continues to be active for some time, raising the likelihood of its influencing cognitive processing as long as it persists.[10] Pickering and Ferreira (2008) discuss the importance of this phenomenon to recent research in sentence processing.

> In the past couple of decades, research in the language sciences has revealed a new and striking form of repetition that we here call *structural priming*. When people talk or write, they tend to repeat the underlying basic structures that they recently produced or experienced others produce. This phenomenon has been the subject of heavy empirical scrutiny. Some of this scrutiny has been because, as in other domains in cognitive psychology (e.g., priming in the word-recognition literature; e.g., McNamara 2005), the tendency to be affected by the repetition of aspects of knowledge can be used to diagnose the nature of that knowledge. [T]he tendency to repeat aspects of sentence structure helps researchers identify some of the representations that people construct when producing or comprehending language. As we shall see, much structural priming is unusually abstract, evidently reflecting the repetition of representations that are independent of meaning and sound. This is therefore informative about how people represent and use abstract structure that is not directly grounded in perceptual or conceptual knowledge. One possibility is that the repre-

[8] See, for instance, Levelt, W. J. M. (1970). "A Scaling Approach to the Study of Syntactic Relations," in d'Arcais, G. B. F. and Levelt, W. J. M. (eds.), *Advances in Psycholinguistics*, pp. 109–121. North-Holland, Amsterdam.

[9] Clark and Clark (1977) discuss the methodological problems of the early experiments.

[10] This formulation is somewhat loose. For a fine-grained distinction between priming and persistence, see Pickering and Ferreira (2008: p. 428). The additional details do not affect the present argument.

sentations that it identifies can be equated with the representations assumed in formal linguistics.

An early and influential study in this vein is reported in Bock and Loebell (1990). The researchers were careful in eliminating the semantic and lexical confounds mentioned above, by constructing their stimulus materials in such a way as to vary syntactic structure independently of lexical and semantic structure, and vice versa. This was made possible by the fact that some verbs in English are ditransitive—capable of being used in semantically identical but syntactically distinct expressions. Examples of ditransitive verbs include ‘give’, ‘sell’, and ‘send’. The examples in (13)–(15) illustrate their alternations between a double-object dative construction, i.e., a V-NP-NP structure, as in the first in each of the pairs, and a prepositional dative construction, i.e., a V-NP-PP structure, as in the second in each of the pairs.

(13) a. Quentin [$_{VP}$ gave [$_{NP}$ Oliver] [$_{NP}$ a car]]. / b. Quentin [$_{VP}$ gave [$_{NP}$ a car] [$_{PP}$ to Oliver]].
(14) a. Quentin [$_{VP}$ sold [$_{NP}$ Oliver] [$_{NP}$ a car]]. / b. Quentin [$_{VP}$ sold [$_{NP}$ a car] [$_{PP}$ to Oliver]].
(15) a. Quentin [$_{VP}$ sent [$_{NP}$ Oliver] [$_{NP}$ a car]]. / b. Quentin [$_{VP}$ sent [$_{NP}$ a car] [$_{PP}$ to Oliver]].

Bock and Loebell’s experiment made use of the well-known picture-description paradigm. Participants were first asked to read some sentences out loud. Unbeknownst to them, these sentences served as primes and were selected by the experimenters for having a preposition after the verb (i.e., a V-NP-PP structure), but differing from (13b)–(15b) in their semantics and lexical constituency. For instance, although a sentence like (16) has the same *syntactic* structure as those in (13b)–(15b), it has none of the same words as the sentences in (13b)–(15b) and, crucially, has a different semantic interpretation—e.g., the preposition ‘to’ carries a locative meaning, as against the dative meaning of the prepositions in (13b)–(15b).

(16) IBM [$_{VP}$ moved [$_{NP}$ a bigger computer] [$_{PP}$ to the Sears store]].

Having read out loud sentences like (16), participants were shown pictures and asked to describe them. The pictures depicted events that involve an agent giving something to someone. Such events can be described equally well by sentences that employ the double-object dative construction (e.g., *The boy gave the girl flowers*) and ones that employ the prepositional dative construction (e.g., *The boy gave flowers to the girl*).

What Bock and Loebell found was that participants exhibited a strong priming effect. Those who initially read out loud sentences that employ the double-object dative construction were more likely to employ the same construction in describing the events depicted in the pictures. Similarly, those who had recently read out loud sentences that employ the prepositional dative construction were more likely to employ *that* construction in describing the events depicted in the very same pictures. This strongly suggests that the subjects in the experiment constructed a mental representation of the syntactic properties of the sentences that they were initially asked to read, and then used that representation in repeating those sentences out loud. The representation then remained active in their language processing system, making it more likely to be reused in the production task.

The experiment just described was one of the first in what has become a very long line. Structural priming research is thriving, in large part because the results are so robust and the data so telling. This leads Pickering and Ferreira (2008) to speculate about the "intriguing possibility that *all* levels of processing that occur during production show priming and therefore that the absence of priming suggests the absence of a corresponding level of representation" (p. 429, emphasis added). In their comprehensive review of the literature, they discuss various refinements and extensions of Bock and Loebell's initial conclusions. The more recent studies they survey show that structural priming is not restricted to the constructions mentioned above—e.g., it occurs with active-passive pairs and other constructions. Moreover it is *not* due to the presence of common closed-class words in the stimulus materials—e.g., the preposition 'to' in sentences (13)–(16).[11] Nor is structural priming restricted to a single language; the phenomenon has been observed in German, and bilingual English-German speakers exhibit *cross-linguistic* priming effects, wherein production in one of their languages is primed by structures from the other. Children and aphasics also exhibit structural priming effects, ruling out the possibility that the phenomenon is restricted to some special set of language users. Further studies rule out the possibility that subjects produce forms similar to the prime because they want to stay in the same rhetorical register, e.g., formal speech. Other communication-related effects have also been controlled for. Similarly, structural priming is independent of both prosody and argument structure (i.e., θ-assignment), and can be elicited cross-modally from spoken to written language and vice versa.[12] Finally, the same findings have been replicated using experimental paradigms other than the picture-description paradigm. These include sentence recall, written sentence completion, and spoken sentence completion.[13] Having surveyed and ruled out a range of possible confounds, Pickering and Ferreira write:

> In conclusion, taken together, these results provide compelling evidence for autonomous syntax: The production of a sentence critically depends upon an abstract syntactic form that is defined in terms of part of speech forms (e.g., nouns, verbs, prepositions) and phrasal constituents organized from those (noun phrases, verb phrases, prepositional phrases), and this abstract syntactic form has a large influence upon structural priming.

There remains a question, of course, about precisely *which* structural descriptions the HSPM constructs. Although experiments like those reviewed above show

[11] Bock (1989) finds priming across sentences with different prepositions—e.g., 'to' and 'for'.

[12] This contradicts a contention of Devitt (2006a) to the effect that there may well be no modality-neutral language faculty. Pickering and Ferreira discuss what they take to be "strong evidence that at least those aspects of structural knowledge that underlie structural priming are modality independent—they are used in the same way both when speaking and when writing" (p. 439).

[13] See Pickering and Ferreira (2008: p. 428) for a detailed list of references.

that the HSPM constructs representations of syntactic structure—in addition to representations of semantic, prosodic, and argument structure—this leaves open the possibility that the HSPM constructs only certain kinds of syntactic descriptions and not others. For instance, we might ask whether the HSPM computes not only the surface structures of sentences, but also the D(eep)-structure representations posited by the Government and Binding theory (Chomsky 1986; Haegeman 1994) and earlier versions of transformational grammar (e.g., Chomsky 1965).

The experiments in Bock et al. (1992) provide evidence against the hypothesis that, in the course of producing passive sentences, the HSPM constructs D-structure representations. In these experiments, the materials employed were active and passive sentence pairs, in which the animacy values of the nouns were varied, as in the following examples:

(17)	**Five people carried the boat.**	**(active, animate surface subject, inanimate deep object)**
(18)	**The boat was carried by five people.**	**(passive, inanimate surface subject, inanimate deep object)**
(19)	**The boat carried five people.**	**(active, inanimate surface subject, animate deep object)**
(20)	**Five people were carried by the boat.**	**(passive, animate surface subject, animate deep object)**

Bock et al. (1992) asked participants to read sentences like (17)–(20) out loud. These served as the primes. The participants were then asked to describe pictures in which, e.g., a boy wakes up to the sound of an alarm clock ringing. This event can be described using either an active or a passive form, as in the examples below:

(21)	**The alarm clock awakened the boy.**	**(active, inanimate surface subject, animate deep object)**
(22)	**The boy was awakened by the alarm clock.**	**(passive, animate surface subject, animate deep object)**

In interpreting the results, it is important to understand that the Government and Binding framework treats passive sentences like (18) as having a D-structure representation in which 'the boat' is located in the object position, as depicted in (23).

(23)	**was [$_{VP}$ carried [$_{NP}$ the boat] by five people.**	**Partial D-structure representation of sentence (18)**

This "deep object" is then moved by a transformational step into what appears on the surface to be a subject position, resulting in the surface form shown in (18).

With this in mind, we can appreciate the logic of the experiment. The materials are designed to tease out whether the participants are primed by the surface structures of (17)–(20) or their deep structures. In accordance with the structural priming results described above, it was expected that participants primed with active sentences (17) and (19) would be more likely to describe the event using the active sentence in (21), and that those who were primed with passive sentences (18) and (20) would produce passive descriptions like (22). The data bear this out, thus providing additional confirmation of the structural priming hypothesis. However, what was *not* known was whether participants who heard sentences with, say, animate deep objects—e.g., (19) and (20)—would produce a description with an animate deep object—i.e. the sentence in (21). If they do, then this would provide evidence for the hypothesis that participants are constructing D-structure representations. This, however, was *not* what Bock, et al. observed.

Instead, the data show that participants were more likely to produce (21) after being primed with (18) and (19) than after being primed with (17) and (20). As Pickering and Ferreira point out, the crucial thing to look at is the difference between the priming effects of sentences (18) and (20). Both are passive, so both are less likely to prime for the active sentence (21). However, whereas (18) shares with (21) the property of having an inanimate surface subject, (20) does not. Thus, the fact that (18) primes the production of (21) more than (20) does "reveals that semantic representations—representations with particular animacy values—are mapped onto surface structure positions, not deep structure positions, contradicting an approach to production that is directly translated from transformational accounts of syntactic representation in formal linguistics" (Pickering and Ferreira 2008: p. 433). For this reason and others, today's most promising comprehension models are typically *not* designed to compute D-structure representations. In general, they directly compute S-structure representations, of one stripe or another.[14]

5.3.1 Summary

The phenomenon of priming provides an insight into the nature and effects of the representations we construct in the course of many cognitive processes. Structural priming allows us a glimpse into the character of the representations constructed by the HSPM in the course of language comprehension and use. A number of structural

[14] Below, I discuss the much-studied question of whether the HSPM constructs representations that include empty categories, such as *wh*-traces, PRO, *pro*, and NP-traces. These are posited by the Government and Binding theory, but not by (some versions of) competing grammatical frameworks, such as Head-Driven Phrase-Structure Grammar (Pollard and Sag, 1994) and Lexical-Functional Grammar (Bresnan, 1978, 2001). The Government and Binding framework posits *wh*-traces and NP-traces at S-structure and at the distinct level of representation, LF. As has been widely noted, there is no need for a parser or a generator to construct D-structure representations, given that all of the information in those representations is already encoded by antecedent-trace relations at S-structure. See, for instance, Fodor (1989: p. 178). The most recent incarnation of transformational grammar in the Principles and Parameters tradition—viz., the Minimalist Program—does away with the level of D-structure altogether. I discuss Minimalist parsing models in Chap. 9.

priming experiments shore up strong evidence for the psychological reality of MPMs. They also allow us to rule out substantive proposals that were entertained in the early stages of the cognitivist study of language—particularly, the idea that the HSPM constructs D-structure representations.[15]

5.4 The Argument from Garden-Path Effects

A classic form of argument for the psychological reality of mental phrase markers begins with the observation that competent language users have problems reading and understanding sentences like (8)–(12), repeated below.

(8) Jake tells students he intrigues to stay.
(9) Benny knows the boy hurried out the door slipped.
(10) The soldier persuaded the radical student that he was fighting in the war for to enlist.
(11) Aron gave the man who was eating the fudge.
(12) The old train the children.

From the point of view of formal syntax, all of these sentences contain a local ambiguity, which is resolved at or before the end of the sentence. What could explain the fact that even proficient readers encounter measurable processing difficulties with regard to such sentences? A standard explanation appeals to the on-line construction of mental phrase markers.

A parsing routine that computes phrase markers incrementally will update its representation of a sentence in accordance with the words or phrases that it encounters, up to the point at which the sentence becomes syntactically ambiguous. At that point, the parser has to make a choice among the possible ways of continuing the phrase marker that it has thus far constructed.[16] Any ambiguity resolution strategy will sometimes lead a parser to make *incorrect* choices—i.e., choices that give rise

[15] Phillips and Lewis (2013) discuss an issue raised by the studies reviewed above. Structural priming has been observed between sentences that, from the point of view of a sophisticated syntactic theory (e.g., Government and Binding theory), exhibit fine-grained syntactic differences. That is, for the purpose of structural priming, syntactic structure A might be "the same as" syntactic structure B, even though A and B are, in detail, quite different. But this doesn't show that the MPMs constructed by the HSPM fail to conform to fine-grained grammars. As Phillips and Lewis point out, all this shows is that "the structural priming paradigm is a relatively blunt tool for investigating structure, because relatively coarse-grained similarity between structures is sufficient to cause structural priming. This would leave open the possibility that on-line processes build fine-grained structures" (p. 13).

[16] Here, I assume a serial architecture. Parallel models will build multiple grammatically licensed structures and rank them. Serial and parallel models yield identical observable results in most (though not all) psycholinguistic experiments. When a serial model makes a mistake, it incurs extra computational load by being forced to backtrack. When a parallel model does so, the extra compu-

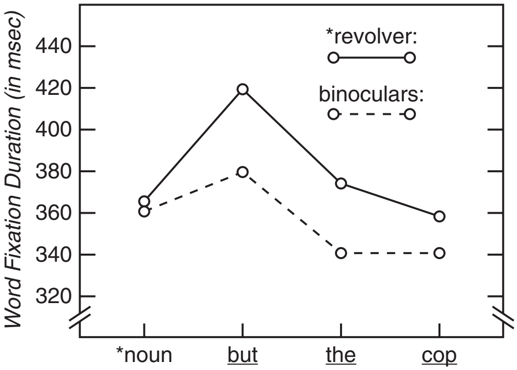


Fig. 5.3 Eye-tracking data from Rayner et al. (1983). The graph depicts a significant spike in fixation times at the critical region of sentence (24), represented by the *unbroken line*. Note the clear contrast between the fixation times at the critical region in sentence (25), depicted by the *broken line*

to expectations that the remainder of the sentence will serve to disconfirm. Herein lies the explanation of the aforementioned processing difficulties. In sentences like (8)–(12), the HSPM's preferred structural assignment turns out to be incorrect in the long run. Additional computational load is then incurred in revising the MPM—a process known as *reanalysis*. This additional processing burden shows up in behavioral and neurophysiological indicators of processing difficulty, such as error rates, delayed reaction times, and the P600 mentioned above. The success of this explanation of the observed processing difficulties constitutes evidence in favor of the claim that the HSPM construct MPMs.

Countless instances of this argument can be found in the sentence processing literature. To take a classic study, Rayner et al. (1983) examined the eye movements involved in reading sentences (24a) and (24b).[17]

(24) a. The spy saw the cop with a revolver, but the cop didn't see him.
b. The spy saw the cop with the binoculars, but the cop didn't see him.

The data in Fig. 5.3 show that readers fixate immediately after the word 'revolver' in (24a) for a significantly longer time than they do after the word 'binoculars' in

tational load is devoted to re-ranking the parses stored in working memory at that point. For further discussion of this issue, see Crocker, Pickering and Clifton (2000). Note also that I am *not* equating parallel models with constraint-satisfaction models (to be discussed later in this section). As a matter of sociology, these tend to be associated. But, as Bornkessel-Schlesewsky and Schlesewsky (2009: p. 106, fn. 5) point out, the question of whether the HSPM constructs mental phrase markers in parallel is logically distinct from the question of what information it uses to resolve local ambiguities or to rank the parses it constructs. (For instance, Gibson, 1991 endorses a parallel parsing model, but not of the constraint-satisfaction variety.) Nor should we simply assume that parallel parsing models will have a connectionist architecture. The best-known parsing models for classical computational architectures employ the CYK algorithm or the Earley algorithm, both of which are parallel, in the sense that they store all possible parses consistent with the grammar in a data structure known as a "table" or a "chart." (See Jurafsky and Martin, 2008: ch. 12.) I discuss these algorithms in Chap. 8.

[17] For an extensive discussion of the eye-tracking paradigm, see Rayner (1998).

(24b). Rayner et al. argue that the increased fixation is a result of the fact that the HSPM initially constructs a representation of sentence (24a) in which the prepositional phrase ‘with a revolver’ attaches to the verb ‘saw’, not the noun ‘the cop’ (Fig. 5.4). Fractions of a second later, this initial attachment preference comes into conflict with the reader’s world knowledge—viz., that one is less likely to use a revolver to see something than to see someone who is in possession of a revolver. This gives rise to a *reanalysis* of the sentence, in the course of which the prepositional phrase is attached to the noun ‘the cop’ (Fig. 5.4a). By contrast, the initial attachment preference in the case of sentence (24b) is consistent with the semantic interpretation of the sentence (Fig. 5.4), so processing is not delayed and fixation times stay low.

Traxler and Pickering (1996) report the results of a similar experiment in which an eye tracker was used to measure the difference between fixation times associated with the critical regions in sentences like (25)–(28).

(25) I recognized you and your family would be unhappy here.
(26) I recognized you and your family right away.
(27) I recognized she and her family would be unhappy here.
(28) I recognized that you and your family would be unhappy here.

Sentence (25) exhibits a classic local ambiguity. The verb ‘recognize’ can take as a complement both a noun phrase, as in (26), and a full clause, as in (27). When a sentence contains a complementizer (e.g., ‘that’) after the verb, as in (28), only a sentential complement can follow. However, the complementizer is optional in constructions like (25), making for a local ambiguity that is resolved relatively late in the sentence. In the case of (25), the HSPM prefers the noun phrase continuation. That is, it initially analyses the string ‘you and your family’ as a noun phrase—the object of the verb ‘recognize’. When the auxiliary verb ‘would’ is encountered, the analysis has to be revised in favor of the sentential complement reading. The revision process exerts a computational load on the system, giving rise to elevated fixation times.

The experiment reported in Traxler and Pickering (1996) was designed to test whether, and to what extent, the HSPM makes use of information regarding the syntactic property known as Case. In a common dialect of American English, only a handful of nouns are overtly marked with Case. The pronouns ‘I’, ‘she’ ‘he’, ‘we’, and ‘they’ take different forms depending on their Case, as shown in the following examples:

(29) I kissed her. / She kissed me. / *Me kissed her. / *I kissed she. / *Her kissed me.
(30) They kissed us. / We kissed them. / *Them kissed us. / *They kissed we. / *We kissed they.

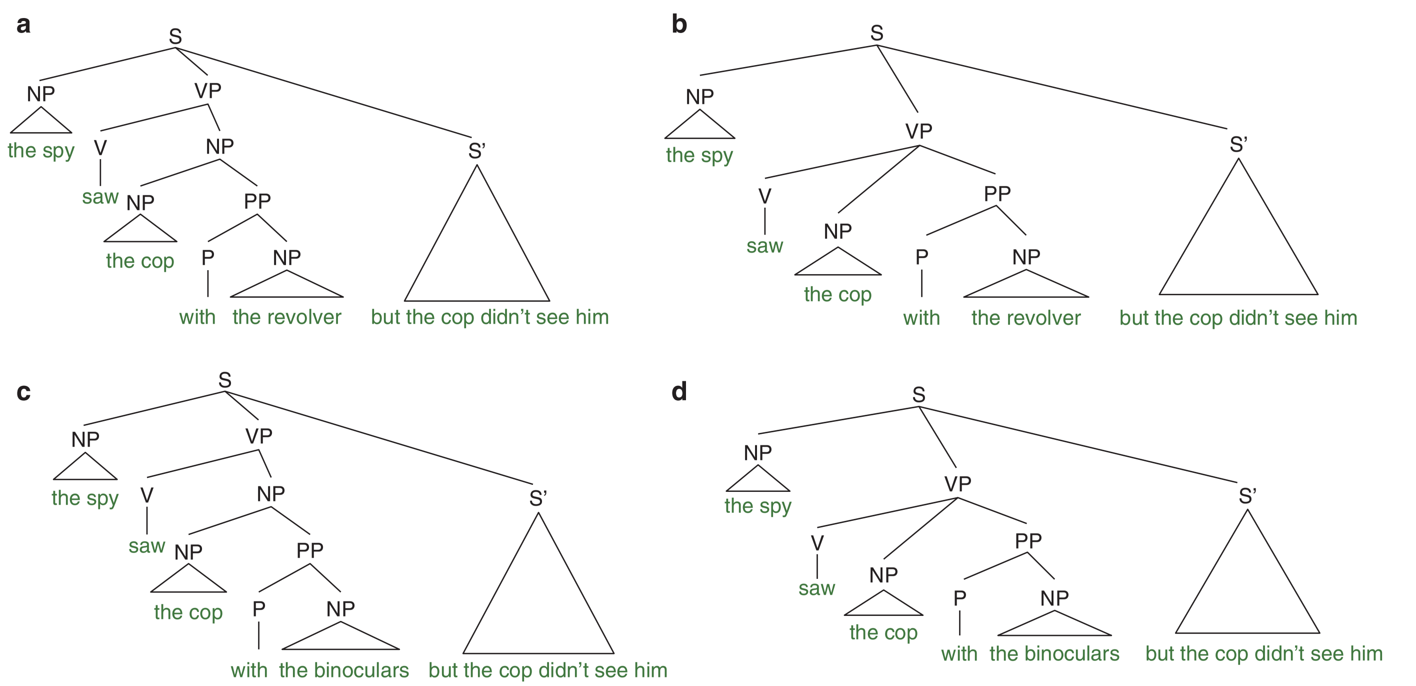


Fig. 5.4 In accordance with the principle of Minimal Attachment, when sentence (24a) is the input, the HSPM will initially fail to build the plausible and grammatical structure in (**a**), preferring instead the semantically implausible structure in (**b**). The reason is that, in (**a**), the verb phrase (VP) dominates *six* nonterminal nodes—one more than it does in the structure depicted in (**b**). The additional nonterminal node in (**a**) is an NP, which is needed to accommodate the attachment of the prepositional phrase 'with the binoculars' to the NP 'the cop'. Minimal Attachment rules against constructing a mental phrase marker with this additional internal structure. When sentence (24b) is the input, Minimal Attachment likewise dictates avoiding the structure depicted in (**c**) and building instead the structure shown in (**d**). In this case, the attachment preference turns out to be semantically plausible. Note that the tree structures in these diagrams are far simpler than would be posited in contemporary syntactic theories, but they are sufficient to illustrate the structural distinctions relevant to the present discussion.

Unlike those above, however, the pronoun ‘you’ does not exhibit Case overtly. Its morphology is the same regardless of whether it has Nominative, Accusative, or Dative Case.

(31) I kissed you. / You kissed me. / I gave a gift to you. / I gave you a gift.

It is in this respect that sentences (25) and (27), repeated here, differ.

(25) I recognized you and your family would be unhappy here.
(27) I recognized she and her family would be unhappy here.

The fact that the pronoun ‘she’ in (27) overtly carries Nominative Case can provide an important clue to the HSPM, for it entails that ‘she’ is the subject of a subordinate clause—a sentential complement—rather than the object of the verb ‘recognize’ (Fig. 5.6). Because the HSPM is sensitive to this information, participants do not fixate on the word ‘would’ in the sentence (27) for an extra period of time, as they do in the case of sentences like (25). Traxler and Pickering conclude that the HSPM makes rapid use of Case information in the early stages of processing.

Similar questions can be raised with regard to information about pronoun gender. Sentences like (32) demonstrate that, lacking a complementizer, the verb ‘know’ has a *very* strong preference for an NP complement, as against a sentential complement.

(32) I know the man who believes the Queen shaves himself.

The sentence in (32) presents enormous difficulties for most readers. Indeed, the effect is so pronounced that readers will not only fail to correctly answer simple questions on the basis of the information it contains (e.g., “What is the speaker claiming knowledge of?”) but they will doggedly insist that the sentence is uninterpretable. That the sentence is, in fact, perfectly grammatical can be seen by attending to sentence (33), which differs from (32) only in the addition of the optional complementizer ‘that’.

(33) I know that the man who believes the Queen shaves himself.

Once it’s made plain that the complement of ‘know’ is sentential, there is no longer anything impeding the binding of the reflexive pronoun ‘himself’ to the NP ‘the man who believes the Queen’. Indeed, given that both the pronoun and the NP are marked with a masculine feature, this is the natural binding relation.

The studies reviewed above, and many others like them, serve to demonstrate three points. First, in the course of processing, phrase structure is imposed on the linguistic input. That is, a mental phrase marker is constructed, in a manner that is sensitive to the syntactic properties of the stimulus. Second, garden-path effects are not limited to cases in which the processing difficulty is *consciously* detected, as it typically is in the course of processing the following sentences:

(34) The ship floated down the river sank.
(35) Fat people eat accumulates.
(36) The cotton clothing is made of grows in Arkansas.
(37) The daughter of the king's son admires himself.

Sentences like (34)–(37) have been dubbed 'conscious garden paths'. Though they are exceedingly rare, one finds them mentioned with some frequency in the literature outside of psycholinguistics—particularly in neighboring disciplines like philosophy. The reason is not far to seek; such cases illustrate vividly the distinction between grammaticality and acceptability.

Unfortunately, the press these sentences have gotten has engendered some confusion on the part of nonspecialists; it gives rise to the assumption that *all* garden-path effects are conscious—i.e., that if an anomaly is not *consciously* detected, then it has not occurred. This is, however, a mistake, due either to overgeneralization or to a conflation of mentality with consciousness. There is simply no compelling reason to think that garden-path effects always give rise to a *conscious* awareness of anomaly.[18] Nonconscious awareness of anomaly can be observed from the third-person perspective by close (theory-laden) attention to fine-grained behavioral data—e.g., the eye tracking data obtained by Rayner, et al.—and, more recently, neurophysiological markers.

The third point illustrated by the studies reviewed above is that the HSPM seems to construct mental phrase markers in accordance with quite general ambiguity resolution strategies. The nature of these heuristics is currently under debate. One popular proposal appeals to a parsing principle known as *Minimal Attachment* (Frazier 1979), which says that the parser will attach incoming material into the existing MPM in such a way as to minimize the number of nonterminal nodes in the resulting structure. Minimal Attachment is a least-effort principle; operating in accordance with it, the parser avoids postulating more structure than it absolutely must at any given point. An application of Minimal Attachment to sentences (24a) and (24b) was illustrated in Fig. 5.4.

Minimal Attachment can also account for the conscious garden path effect associated with sentences that have a local main-clause/relative-clause ambiguity, such as the famous sentence (38) and its cohorts, (39)–(41).

[18] "The fact that hearers are not always conscious of having made a mistake in the analysis of such sentences (as they are for notorious garden-paths such as *The horse raced past the barn fell*) is not, we submit, a good argument against this kind of perceptual complexity" (Frazier and Fodor 1978: p. 296). See Pritchett (1992: p. 33) for illuminating reflections on this point. Briefly, Pritchett argues that although conscious reflection cannot tell us anything about sentence processing, the fact that some processing difficulties are consciously registered while others are not nevertheless constitutes evidence that the former are more serious than the latter.

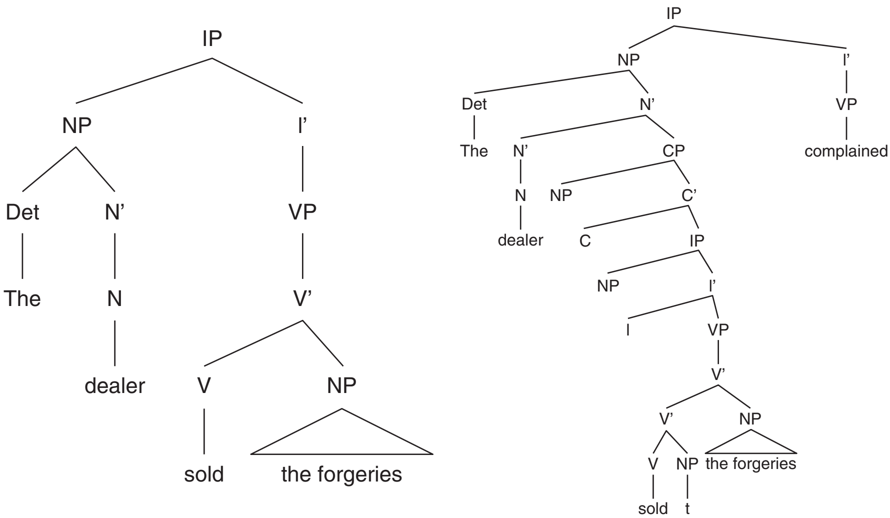


Fig. 5.5 In accordance with the principle of Minimal Attachment, the HSPM assumes that the noun 'the dealer' is the subject of that verb 'sold'. It thus attaches the verb to the existing structure in the manner depicted in the left panel. Subsequent input reveals that the HSPM's assumption was incorrect, thus necessitating reanalysis. The difficulty of reanalysis in this case is a function of the sheer amount of additional structure needed to accommodate a passive participle reading

(38) The horse raced past the barn fell.
(39) The ship floated down the river sank.
(40) The dealer sold the forgeries complained.
(41) The man sent the letter cried.

The verb-forms 'floated', 'raced', 'sold', and 'sent' are ambiguous. They can serve either as past-tense verbs that are part of a main clause or as past participles that introduce a reduced relative clause. (In these cases, the optional complementizer 'that' has been omitted.) The locally ambiguous structures associated with these sentences are illustrated in Fig. 5.5.

Minimal Attachment is one of a handful of principles that parsing theorists appeal to in explaining a wide range of the HSPM's ambiguity resolution preferences. Another such principle is Late Closure, which dictates that the parser will incorporate newly encountered material into the most recent phrase or clause of the mental phrase marker that it has already constructed. Late Closure is invoked to explain the HSPM's preference in cases like the one illustrated in Fig. 5.7.

Both Minimal Attachment and Late Closure are close cousins of a number of other sentence processing principles that have been discussed over the past several decades. In a remarkably prescient work, Kimball (1973) proposed seven principles of sentence processing designed to account for the way in which the HSPM assigns a phrase-structure analysis to ambiguous input. These principles differed in a crucial

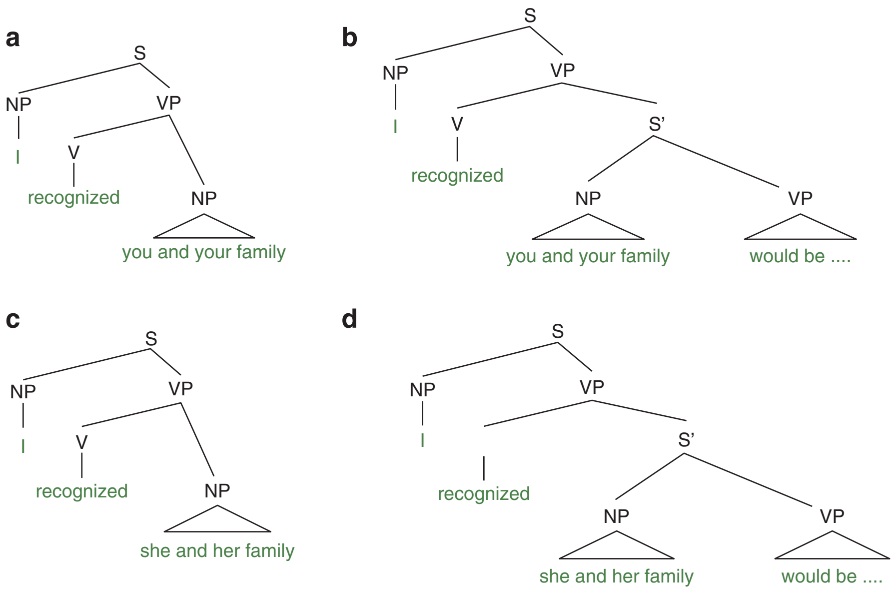


Fig. 5.6 In accordance with the principle of Minimal Attachment, the HSPM assumes that the noun phrase 'you and your family' is the object of the verb 'recognize', thus building the structure depicted in (**a**). This assumption is disconfirmed by subsequent input, giving rise to a process of reanalysis, which results in structure (**b**). However, when Case information is available, as in sentence (27), structure (**c**) is never built. Evidently, the HSPM uses the information about Case *prior* to applying the principle of Minimal Attachment, thus building the structure in (**d**)

respect from the heuristics proposed by Fodor et al. (1974). Unlike those authors, Kimball assumed that the HSPM would build its MPMs in a way that draws on the rules of a surface-structure grammar.[19] On his view, the processing principles that are true of the HSPM would make ineliminable reference to the proprietary notions of an independently motivated syntactic theory—e.g., the notion *number of nonterminal nodes*.

Thus, in addition to explaining a broad range of data, Kimball's work made it possible to see, if only in dim outline, how one might incorporate an independently motivated syntactic theory into a model of sentence processing—an idea that Fodor et al. (1974) rejected, for reasons that have come to be recognized as spurious.[20] Kimball's approach thus offered psycholinguists an alternative to the proposals in

[19] Kimball discussed only the parsing of surface structures without traces. He did not offer a comparably detailed treatment of transformational dependencies (Sect. 5.5). Fodor (1978) discusses an extension of Kimball's approach that deals with transformational phenomena.

[20] See Berwick and Weinberg (1984: ch. 2), Phillips (1994, 1996) and Phillips and Lewis (2013). I discuss the issue in detail in Chap. 9.

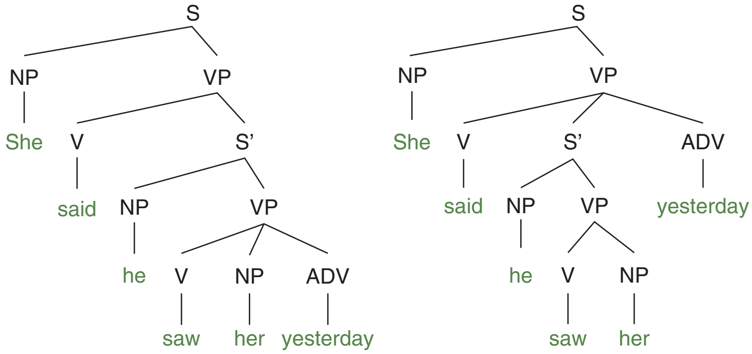


Fig. 5.7 Having built a structure for the input 'She said he saw her', the HSPM receives the adverb 'yesterday'. The grammar licenses two possible attachments, represented by the *left* and *right panels*. In accordance with the principle of Late Closure, the HSPM resolves this ambiguity in favor of the structure depicted in the *left panel*. That is, it attaches the adverb to the most recent phrase of the structure it had already built—in this case, the verb phrase 'saw her'

Fodor et al. (1974), allowing them to seek more concise and predictive principles that accord with the grammars emerging from formal syntax.[21]

In a series of influential follow-ups to Kimball's pioneering work, Frazier and Fodor (1978) and Frazier (1979) refined Kimball's principles into what we now know as Minimal Attachment and Late Closure. Other theorists extended Kimball's ideas in similar directions. For instance, Wanner and Maratsos (1978) describe a nondeterministic Augmented Transition Network (ATN) parser that mimics, in its operations, a recursive phrase structure grammar with built-in attachment preferences (see Chap. 9). Similarly, Marcus (1980) proposed a deterministic parsing model that implemented some of the same principles. Ford et al. (1982) proposed principles that they called "Lexical Preference" and "Final Arguments," which do much of the same work, though in the context of a parser modeled on Lexical Functional Grammar. Pereira (1985) discusses a "shift-reduce" parser, where Kimball's principle of Right Association is implemented as a preference for shifts over reductions, and Fodor and Frazier's principle of Minimal Attachment is implemented as a preference for maximally long reductions. Subsequently, Pritchett (1992) and Frazier and Clifton (1996) proposed refinements of Kimball's principles that appeal to the θ-roles of Government and Binding theory and the related distinction between arguments and adjuncts.[22] In more recent years, computational linguists have proposed a variety of generalizations that subsume MA and LC, and

[21] Frazier (1979) examines in detail the merits and drawbacks of Kimball's approach. A more recent treatment appears in Pritchett (1992: pp. 26–30), where Kimball's work is situated in the context of subsequent theories of sentence processing. Pritchett's own theory of sentence processing is directly inspired by Kimball's work, though it draws on a more recent grammar, to which Kimball did not have access.

[22] I discuss the ATN model, the shift-reduce parser, and θ-roles in some detail in Chaps. 8 and 9.

apply even in attested cases that appear to conflict with those principles (Hale, 2011).

It is important to be clear about the status of such processing principles. To my knowledge, it has never been suggested that Minimal Attachment, Late Closure, or any of their counterparts mentioned above are *represented in* the HSPM, or that the HSPM has *knowledge of* them, however tacit. Rather, these are intended to be *descriptive* principles; they are *true of* the HSPM in much the same way that the principles of celestial mechanics are *true of* our solar system. Admitting the descriptive accuracy of such principles does not, then, automatically commit one to any view regarding the psychological reality of syntactic rules or principles—at least not without additional argument. Still, given that MA, LC, and related principles make ineliminable appeal to the proprietary notions of syntactic theory, it follows that the HSPM works *in accordance* with a particular theory of syntax. In Chap. 7, we shall tackle the question of how to properly cash out the notion of "working in accordance with." I shall argue there that the correct way of cashing it out commits us to a position stronger than what Devitt (2006a) calls "the minimal position (M)."

> (M) A competence in a language, and the processing rules that govern its exercise, respect the structure rules of the language: the processing rules of language comprehension take sentences of the language as inputs; the processing rules of language production yield sentences of the language as outputs. (p. 57)

If we take MA, LC and the like to be true of the HSPM—as the data suggest—then we should *not* rest content with the minimal claim embodied in (M). The relation between a descriptive grammar and the operations of the HSPM is not *simply* that the latter is able to process sentences described by the former. The fact that the HSPM works "in accordance with" a descriptive grammar entails that the rules or principles of that grammar are either represented or embodied in the HSPM.

5.4.1 Summary

In this section we have surveyed a sample of the studies that shed light on the internal workings of the HSPM by examining fine-grained behaviors such as eye movements. The studies uniformly support the claim that the HSPM constructs mental phrase markers. They also provide information about the character of its processing routines, especially with respect to ambiguity resolution. In the next section, we examine another well-studied aspect of the comprehension process—the assignment of filler-gap relations. Studies of this phenomenon shed further light on the syntactic information encoded in mental phrase markers, as well as on the processing routines of the HSPM. Indeed, a number of theorists have expressed the hope that these studies will allow us to discern which descriptive grammar most closely resembles the grammar employed by HSPM. We will also encounter two new experimental paradigms. Like those discussed in Sect. 5.3, the experiments used in studying filler-gap processing also make use of the phenomenon of priming, though not *structural* priming per se.

5.5 The Argument from Filler-Gap Processing

Consider the questions and relative clause constructions, exemplified in (42)–(45).

(42) What instrument do you play?
(43) I play the guitar that my parents bought.
(44) Whom did they consult before buying it?
(45) I don't remember which of their friends they consulted.

The verbs 'play', 'bought', and 'consulted' are transitive—that is, they take noun phrase as a complement, as in (46)–(48).

(46) I play the guitar.
(47) My parents bought the guitar.
(48) They consulted their friends.

In declarative sentences like (46)–(48), the position of the two nouns around the verb makes it clear who is performing the action denoted by the verb and what the action is being performed on. The position of the nouns in (47), for instance, makes clear who is doing the buying and what is being bought. However, in (42)–(45) the noun phrases are not in their canonical positions. Rather, they appear at the front of a clause. Yet the verbs are understood as having those nouns as their objects. The nouns must therefore maintain some sort of long-distance relationship with the object positions of the verbs. Different linguistic theories employ different formal devices to represent this relationship. Government and Binding (GB) theory (Chomsky 1986; Haegeman 1994) uses so-called "empty categories." An empty category is a theoretical posit. While not overtly present in writing and speech, it is assumed to occupy a position in the underlying syntactic structure of a sentence and to behave in many ways as though it *were* overtly present.[23] Sentences (53) and (54) are relatively uncontroversial examples of constructions in which, according to GB, empty categories play this role.

(53) Alia plays the keyboard and Jeffrey the bass.
(54) Igor loves Alia and she him.

Sentences (53) and (54) are understood as having a verb between the two final nouns. And not just *any* verb; sentence (53) cannot be interpreted as meaning that Alia plays the keyboard and Jeffrey *sells* the bass. The missing verb in (53) must be 'play' and in (54), 'love'. This relationship can be represented as follows:

[23] Empty categories are sometimes called "aphonic" or "phonologically null" categories, because they are not spoken or heard (Schiffer, forthcoming). Of course, outside of linguistics textbooks, empty categories are also *orthographically* null.

(55) Alia will play$_i$ the keyboard and Jeffrey [$_{VP}$ EC$_i$ the bass].
(56) Igor loves$_i$ Alia and she [$_{VP}$ EC$_i$ him].

The empty category (EC) in (55) and in (56) is said to be *coindexed* with the overt verb in the prior clause. The subscript 'i' encodes a relationship of identity between the overt verb and its missing counterpart, which in turn constrains the semantic interpretation of the sentence. The overt verb in (55) and (56) is sometimes called a *filler* and the empty category—or the position it occupies—is a *gap*. Returning now to sentences (42)–(45), we see that the same device can serve to encode the relationship between verbs and their objects, as in (57)–(60).

(57) [What instrument]$_i$ do you play EC$_i$?
(58) I play [the guitar]$_i$ that my parents bought EC$_i$.
(59) [Whom]$_i$ did they consult EC$_i$ before buying the guitar?
(60) I don't remember [which of their friends]$_i$ they consulted EC$_i$.

Passive sentences can receive a similar treatment. Recall from Sect. 5.3 that Government and Binding theory posits a level of analysis, D-structure, at which noun phrases are in their canonical positions relative to verbs. Movement operations then apply to these nouns, yielding the surface forms (S-structures) that are then transformed into what we read and hear. S-structures are assumed to contain empty categories that serve to mark the relationship between the displaced nouns and their D-structure positions. Consider (61) and (62):

(61) The boat was carried by five people.
(62) Five people were carried by the boat.

These sentences have the following D-structures:

(63) was [$_{VP}$ carried [$_{NP}$ the boat] by five people.
(64) were [$_{VP}$ carried [$_{NP}$ five people] by the boat.

After the movement operations apply, the resulting S-structures are (65) and (66).

(65) [$_{NP}$ The boat]$_i$ was carried EC$_i$ by five people.
(66) [$_{NP}$ Five people]$_i$ were carried EC$_i$ by the boat.

Government and Binding theory posits empty categories to account for two other syntactic phenomena: raising and control. The first has to do with so-called *raising verbs*, such as 'seem' and 'appear', in sentences like (67) and (68).

(67) Jennifer appears to be reading a book about neuroscience.
(68) Igor seems to enjoy action movies.

In these sentences, the verbs ‘reading’ and ‘enjoy’ are understood as having a subject. For instance, in (67) the subject of ‘reading’ is ‘Jennifer’. However, the subjects in sentences like (67) and (68) are not located in their canonical position; they have been “raised” to a position in the higher clause at the front of the sentence. One way to account for the semantic relationship between the lower verb and its subject is to posit an empty category in the subject position of that verb and co-index that empty category with the subject noun phrase. Here again, the understood subject is taken to be located in the subject position at D-structure and moved to its new position at S-structure. The movement leaves an empty category in its wake. Thus, sentences (67) and (68) are taken to have the following S-structure representations.

(69) Jennifer$_i$ appears EC$_i$ to be reading a book about neuroscience.
(70) Igor$_i$ seems EC$_i$ to enjoy action movies.

Similar considerations motivate the posit of an empty category in constructions involving so-called *control verbs*, exemplified in (80) and (81).

(71) Jeffrey tried to record a song.
(72) Konstantine promised to attend the party.

As with the cases discussed above, the verbs ‘record’ and ‘attend’ are understood as having subjects. This relation is captured by the empty categories displayed below.

(73) Jeffrey$_i$ tried EC$_i$ to record a song.
(74) Konstantine$_i$ promised EC$_i$ to attend the party.

In sum, empty categories allow us to encode various semantically relevant structural relationships between the words in a sentence. Let us distinguish the four types of empty category introduced thus far. Following convention, I use the labels ‘trace’ and ‘PRO’.

(75)	**Igor loves$_i$ Alia and she EC$_i$ him.**	VP-ellipsis marker
(76)	**[What instrument]$_i$ do you play EC$_i$?**	*wh*-trace
(77)	**[$_{NP}$ The boat]$_i$ was carried EC$_i$ by five people.**	NP-trace
(78)	**Jeffrey$_i$ tried EC$_i$ to record a song.**	PRO

Empty categories play an explanatory role in GB theory, but are they psychologically real? Does the HSPM construct mental phrase markers that contain, e.g., *wh*-traces? It is possible, after all, that the HSPM encodes filler-gap relations in some other way. Alternative descriptive grammars—rivals of GB—do not posit traces or PRO. Specifically, Lexical Functional Grammar (LFG) and Head-driven Phrase Structure Grammar (HPSG) encode filler-gap relations without positing empty categories.[24] Hence a demonstration of the psychological reality of empty categories can be used to argue that Government and Binding theory (or the newer Minimalist grammars) more closely resemble the grammar employed by the HSPM than rival grammatical formalisms.[25]

Fodor (1989, 1995) summarizes a number of early experiments designed to address this issue. The most straightforward of the bunch demonstrates what psycholinguists call "the filled-gap effect." This is similar to the garden-path effect that we examined in Sect. 5.4; in both cases, the HSPM makes a prediction that is disconfirmed by later input, giving rise to computationally costly revisions that are, in turn, reflected in longer reading times. What's distinctive about the filled-gap effect is that it has to do with the HSPM's prediction that some position in the syntactic structure of the input will contain an empty category.

Limiting our attention for the moment to *wh*-traces, we can begin by considering sentences (79) and (80), which demonstrate that such a trace can appear in a variety of locations in the input sentence.

(79) Who$_i$ could the little child have forced *wh*-trace$_i$ to sing those stupid French songs for Cheryl last year?

(80) Who$_i$ could the little child have forced us to sing those stupid French songs for *wh*-trace$_i$ last year?

The HSPM is demonstrably sensitive to the fact that the word 'Who' is an antecedent, to which a *wh*-trace will have to be bound—a filler awaiting a gap. Thus, it faces the task of actively searching for legitimate positions at which to posit the *wh*-trace. The filled-gap effect can be observed when the HSPM predicts that the gap will occur after 'forced' in both (79) and (80). In the case of (79), the prediction is correct. But in the case of (80), it leads the parser astray, giving rise to measurable processing difficulties at just the point in the sentence where the word 'us' occupies the predicted position of the *wh*-trace.

Fodor (1989) notes that there is no *obvious* reason why the HSPM should predict a gap where there is none, instead of simply waiting to see whether some overt material occupies that position. She writes:

> Gap anticipation of this kind falsified Fodor's (1978) proposal that the processor, when in need of a certain constituent (say, a noun phrase), will always try the next constituent in the

[24] This is an oversimplification. For details regarding the way in which LFG and HPSG treat wh-constructions and relative clauses see Bresnan (1978, 2001) and Pollard and Sag (1994), respectively. For a briefer treatment of the issue, with a focus on its relevance for psycholinguistics, see Fodor (1989: pp. 177–186) and Featherston (2001).

[25] As noted at the end of Chap. 4, there are methodological grounds for initially favoring the hypothesis that the HSPM employs whatever grammar happens to satisfy the desiderata of formal linguistic theory. For further discussion of this issue, see Steedman 1985: p. 361; 2000: pp. 227–8; Fodor 1989: pp. 177–8; Featherston 2001: pp. 2–3. Still, the hypothesis may well be false; the psychologically real grammar might not meet the compactness constraints imposed by syntacticians (Soames, 1984; Stabler, 1984).

> input to see if it is what it needed, before concluding that the constituent is missing and that an empty category must therefore be assumed. (1989: p. 163, fn. 8)

The HSPM's prediction also cannot be accounted for by Minimal Attachment or Late Closure, as there is no structural ambiguity to resolve in such cases; whether the position after 'forced' in sentences (79) and (80) is filled by on overt lexical item (e.g., 'us') or by a *wh*-trace, the structural relations in the sentence remain the same.

To account for the filled-gap effect, deVincenzi (1991) proposed a third processing heuristic, known as the Minimal Chain Principle (MCP). According to the MCP, the parser will avoid postulating unnecessary gaps/traces and, having identified an item in the input as an antecedent, will posit a gap/trace in the very first position at which it is licensed by the grammar.[26] For present purposes, what is important about the MCP is that it makes ineliminable reference to abstract grammatical notions—e.g., *position at which a trace is licensed*. This reference to grammatical licensing is on a par with the kind we saw earlier, in connection with MA and LC. Hence, we can draw the same conclusions: Even if the processing principles are not themselves represented by the HSPM, but are merely *true of* its operations, the fact that such principles must be formulated in terms of the proprietary notions of a grammar entails that the HSPM employs a grammar, by either embodying or representing it.

To substantiate this line of reasoning, we must show that the HSPM is, indeed, linguistically informed. Fodor (1989) summarizes several studies that suggest that the HSPM draws on information about the subcategorization properties of verbs—i.e., what complements a verb allows or requires—and subsequent experiments have lent further support. For instance, Gorrell (1991) used a reaction time measure on a lexical decision task to show that the HSPM makes a distinction between transitive and intransitive verbs within milliseconds of lexical retrieval—i.e., without any appreciable delay. Frenck-Mestre and Pynte (1995) obtained more fine-grained evidence for the same conclusion by conducting an eye-tracking study.

Stowe (1986) conducted an early experiment that bears on the question of whether the HSPM is sensitive to so-called *island constraints*—i.e., restrictions on the structural configurations that allow for fillers to bind their corresponding gaps. The question of how best to characterize these configurations in a descriptive grammar is a difficult one; the principles appear to be quite abstract. Demonstrating that the HSPM employs such principles in its computations would lend strong support to the claim that they are psychologically real. To address this issue, Stowe (1986) examined sentences like (81).

(81) The teacher asked what$_i$ [$_{NP}$ the silly story about Greg's older brother] was supposed to mean *wh*-trace$_i$.

---

[26] The second conjunct is equivalent to what Frazier and Clifton (1989) refer to as the "Active Filler Hypothesis." I use the locution 'gap/trace' to avoid commitment to grammatical formalisms that have traces and movement operations in their theoretical toolkit.

In the vocabulary of the syntactician, the bracketed noun phrase in (81) is an island. What this means is that the *wh*-trace associated with the word ‘what’ cannot appear within that phrase. Nevertheless, there is a position within the island at which a processor ignorant of the island constraints would be tempted to posit a *wh*-trace. That position is right after the word ‘about’, which requires a noun phrase complement.

The logic behind Stowe’s experiment is this: If the HSPM “knows” about island constraints, then it will not predict a *wh*-trace after ‘about’; if it doesn’t know about them, then it might. Having posited the *wh*-trace after ‘about’, a linguistically ignorant processor would exhibit the filled-gap effect discussed above, elevating the reading times at the word ‘Greg’. Stowe observed no significant rise in reading times at that position, as compared to the analogous position in the control sentence (82).

(82) The teacher asked if [$_{NP}$ the silly story about Greg’s older brother] was supposed to mean anything.

This demonstrates the *absence* of a filled-gap effect. Stowe took this as confirmation that the HSPM has access to a grammar that covers island constraints. Further experimental work has largely borne this out. Phillips and Lewis (2013) summarize the current state of play and provide a number of references to key studies that have been conducted in this area:

> [One] body of on-line studies has examined whether on-line structure building respects various grammatical constraints, i.e., whether the parser ever creates grammatically illicit structures or interpretations. Many studies have found evidence of immediate on-line effects of grammatical constraints, such as locality constraints on wh-movement (Stowe 1986; Traxler and Pickering 1996; Wagers and Phillips 2009), and structural constraints on forwards and backwards anaphora (Kazanina et al. 2007; Nicol and Swinney 1989; Sturt 2003; Xiang et al. 2009). These findings extend to complex cases that present apparent challenges for incremental application of grammatical constraints, such as constraints on backwards anaphora in Japanese, where the constraints must apply before any verb has appeared in the input (Aoshima et al. 2009), and constraints on parasitic gaps inside complex subjects in English, where the parasitic gap precedes its licensor (Phillips 2006). *Findings such as these imply that the structures created on-line include sufficient structural detail to allow the constraints to be applied during parsing*. (Phillips and Lewis 2013; emphasis added)

It is worth examining one of the studies that Phillips and Lewis mention in this connection—Nicol and Swinney (1989). This study initiated a line of research that makes use of an experimental paradigm known as *cross-modal priming* (CMP). To see how this works, consider sentence (83), which contains a relative clause with a *wh*-trace in the object position of the verb ‘accused’.

(83) The policeman saw the boy$_i$ [that$_i$ the crowd at the party accused *wh*-trace$_i$ of the crime].

In (83), there is only one position at which the *wh*-trace is grammatically licensed and only one noun phrase that can legitimately serve as the antecedent of that *wh*-trace: ‘the boy’. However, the sentence contains a number of other noun phrases—‘the policeman’, ‘the crowd’, and ‘the party’—any of which a linguistically ignorant processor might take to be the antecedent. Call these *distracters*.

To test whether the HSPM is temporarily fooled into taking any of the distracters as the antecedent of the *wh*-trace, Nicol and Swinney had participants listen to sentence (83) while fixating on a computer screen. When the word ‘accused’ was spoken, participants saw a word appear on the screen. In some trials, the word was semantically related to ‘boy’, which is the antecedent of the *wh*-trace that appears after ‘accused’. For example, participants saw the word ‘girl’. In other trials, participants saw words that were comparable in length to ‘girl’, but semantically related to one of the distracters. They were asked to read the word out loud and their reaction times were measured. Nicol and Swinney made the following assumption: If the HSPM posits a *wh*-trace after ‘accused’, then it will activate the meaning of the antecedent ‘boy’ at that point, which would in turn prime the recognition of semantically related words, like ‘girl’.[27] By contrast, the recognition of words that are semantically related to one of the distracters would *not* be primed. As expected, this is what they found. The results were robust and have now been replicated a number of times. On the basis of these studies, *wh*-traces had, for some time, been thought to be psychologically real.

As noted above, transformational grammars posit a number of empty categories in addition to *wh*-trace. The question arises, then, whether these, too, have psychological reality—i.e., whether they are explicitly represented in the MPMs constructed by the HSPM. Evidence for a positive answer to this question emerged from a number of studies that employed an experimental paradigm similar to cross-modal priming, known as *visual probe recognition* (VPR). In this paradigm, participants are presented with sentences, one of which is hypothesized to have an empty category (such as PRO or NP-trace) in its underlying structure. In the studies discussed below, the relevant empty category was almost always an NP-trace.[28]

The VPR experiments rest on assumptions similar to those that animate the cross-modal priming paradigm. The presence of an NP-trace is hypothesized to reactivate the meaning of the antecedent, thus priming for faster or more accurate responses. The differences between the two paradigms consist in their dependent variables, as well as in the nature of the tasks they rely on. Whereas Nicol and Swinney (1989) measured response times on reading tasks that were conducted mid-sentence, MacDonald (1989) and McElree and Bever (1989), using the VPR

[27] Priming studies had, by this time, demonstrated quite clearly that the recognition of a word primes the recognition of semantically related words.

[28] Recall that NP-traces are posited by the Government and Binding theory to account for the relation between a verb and its object in passive sentences like (65) and raising constructions like (70), both repeated here:

(65) **[$_{NP}$ The boat]$_i$ was carried NP-trace$_i$ by five people.**

(70) **Igor$_i$ seems NP-trace$_i$ to enjoy action movies.**

paradigm, asked participants to wait until the sentences had been presented and then to judge whether a particular word was present in the input. On many trials, the word was not in fact present, so participants would answer 'No'. But in cases like those shown in (84) and (85), the answer was 'Yes'. The dependent variable was the speed of participants' responses.

(84) The terrorists wanted to disrupt the ceremonies.
[The new mayor at the center podium]$_i$ was shot NP-trace$_i$.
(85) The terrorists wanted to disrupt the ceremonies.
[The new mayor at the center podium] was furious.

Both MacDonald (1989) and McElree and Bever (1989) found that the presence of an NP-trace primed the recall of the word 'mayor'. For instance, participants' responses were faster when primed with (84) than with (85). On the strength of these results, NP-traces were tentatively taken to be psychologically real, on a par with *wh*-traces.

The emerging picture was, however, complicated by the fact that *wh*-traces elicited different effects in the cross-modal priming experiments than did NP-traces and PRO. For instance, in a study that used an improved version of the paradigm, Osterhout and Swinney (1992) found that NP-traces primed for the recognition of their antecedents, but only when they appeared at increasingly longer distances from their antecedents. In the limit, it took a whopping 1000ms for the effect to show up. This contrasts sharply with *wh*-traces, which have priming effects almost immediately. Osterhout and Swinney proposed two alternative explanations for this difference.

According to the first account, the difference has to do with the fact that the antecedents of *wh*-traces always appear in nonargument positions and provide the HSPM with an obvious cue that a gap will appear later in the sentence. The antecedents of NP-traces, by contrast, appear in argument positions, so the HSPM has no warning that the input will contain a gap. According to the second account, the difference is that *wh*-traces are inserted into the mental phrase marker during syntactic processing, whereas the effects associated with passive, raising, and control verbs are due to subsequent *semantic* processing. On this account, NP-traces, unlike *wh*-traces, are not psychologically real.

Although Osterhout and Swinney did not offer any grounds for privileging one of these accounts over the other, they made clear the implications of adopting the second explanation. As noted earlier, a variety of syntactic theories—most notably Lexical Functional Grammar and Head-driven Phrase Structure Grammar—dispense with the theoretical construct of NP-traces and PRO, and deal with passive, raising, and control constructions by alternative means. Hence, if it could be shown that NP-traces and PRO are not psychologically real, this would constitute evidence that the grammar employed by the HSPM is more like these nontransformational grammars than like Government and Binding formalism. For theorists who hold out hope that a single grammar will be both descriptively adequate and psychologically real, this conclusion would carry great significance.

In subsequent discussions, a number of theorists went farther than Osterhout and Swinney, questioning the validity of the studies that were supposed to establish the psychological reality of *wh*-traces. In particular, proponents of Head-driven Phrase Structure Grammar were beginning to claim that their formalism requires *neither* NP-traces *nor* wh-traces. In addition, it became apparent that *all* of the experimental data discussed above can be accounted for by the assumption that the effects are due to semantic rather than syntactic processing. A particularly forceful statement of this view can be found in Sag and Fodor (1994). Questioning the psychological reality of *wh*-traces, they write:

> What would it take to show, on the basis of sentence processing, either that there is, or that there is not, a syntactic entity in an extraction gap? Fodor (1993) noted that no psycholinguistic finding could even in principle qualify as evidence for empty constituents unless it were established that the data pertained to the syntactic processing or representation of a sentence. But for all we know at present, the experimental techniques that have been used to study gap processing to date might be providing information about semantic processing only. If so, they would be completely uninformative about traces, because they wouldn't tell us whether the sentence meaning is computed from a trace or from a traceless gap in the syntax. (Sag and Fodor 1994: p. 8)

We see here another instance of an argument pattern that has emerged several times in the history of psycholinguistics. Recall that the earliest experiments purporting to demonstrate the existence of MPMs were criticized primarily on account of their failure to control for semantic factors (Sect. 5.3). Structural priming studies teased apart syntax and semantics, in the face of deep skepticism concerning a distinctively syntactic level of representation. Careful experiments sufficed to quell such concerns in that case, but we see now that the broader debate regarding the adequacy of explanations pitched at a distinctively syntactic level of processing recurs in discussions of the psychological reality of empty categories. Theorists who seek to eliminate empty categories from both descriptive grammar and from psychological models of sentence can develop alternative explanations of experimental findings, appealing to the semantic level of processing.

Although a decisive resolution of this issue is currently out of reach, recent experimental evidence does bear on it. Before closing this section, then, let us review the results of a series of studies reported in Featherston (2001). As we shall see, Featherston's conclusions favor a view that locates filler-gap effects at a *syntactic* level of representation.

Featherston begins by noting that there are currently four separate accounts of the data from the CMP and visual VPR studies discussed above.[29] Two of these we have already examined. The first is the Trace Reactivation account, which takes at face value the claim that the HSPM posits traces at various positions in the input. On this account, the priming data show that traces are syntactic entities whose presence in the MPM facilitates the reactivation of the meanings of their antecedents—i.e.,

[29] Featherston discusses many experiments that were performed between the late 1980s and his own 2001 studies. For the sake of brevity, I have not mentioned these here. Their implications were taken into account in the design of Featherston's experiments.

that traces are, in this respect, on a par with overt anaphoric elements like pronouns. The second is the Semantic Processing account, which, as the name suggests, construes effects at gap position as being indicative of activity at the semantic phase of processing, rather than the syntactic phase.

The third is what Featherston calls the Depth of Processing account, discussed in Fodor (1995). In reviewing the VPR data presented above, Fodor writes:

> [T]here is known to be an inverse relation between VPR performance on a word, and its expectability and ease of processing in the sentence. Earlier research (not all of it on language) had shown that the greater the effort that goes into processing a stimulus, the more distinctive its subsequent memory representation is. … Let us call this the depth-of-processing effect. It links three things: the less predictable an item is, the greater the effort of processing it, and the easier it is to recall. This provides a good explanation for why passive sentences fare better than adjectival sentences in VPR experiments. Because passive sentences are less common than adjectival sentences, they receive more attention during processing, so they are easier to probe in short-term memory. (pp. 238–9)

To illustrate, let's apply this to examples in (84) and (85), repeated here. The data showed that participants were faster at recalling the presence of the word 'mayor' in (84) than in (85).

(84) The terrorists wanted to disrupt the ceremonies.
[The new mayor at the center podium]$_i$ was shot NP-trace$_i$.
(85) The terrorists wanted to disrupt the ceremonies.
[The new mayor at the center podium] was furious.

The Depth of Processing account has it that the facilitation of participants' recall had nothing to do with the presence of an NP-trace in the MPM for the second sentence in (84), but instead has to do with the fact that this sentence is a passive.

The fourth account that Featherston considers, which he calls the Direct Association hypothesis, is due to Pickering and Barry (1991). According to it, "the effects found are in fact due to the association of the argument with the verb, and that the confusion of this with trace reactivation is due to the position, adjacent to the verb, at which direct object traces are posited" (Featherston 2001: p. 221). It follows from this view that if it were possible to dissociate the location of a trace from the position adjacent to a verb, then priming effects would be observed only at the latter position, not at the location of the trace. As Featherston notes, this dissociation is all but impossible to achieve in English. However, materials from a language like German, which differs in the crucial respects from English, can be used to tease apart the predictions of the Trace Reactivation account and the Direct Association account.

Featherston (2001) reports the results of six separate experiments. I will mention here only the findings from experiments 2–5.[30] Experiments 2 and 3 made use of the

[30] The first experiment was argued to be flawed. The sixth made use of an experimental paradigm known as *sentence matching*, which we have not introduced here.

CMP paradigm, applied to German sentences. With these materials, antecedent reactivation effects are predicted *only* by the account that takes traces to be syntactic entities on a par with overt pronouns, whose presence in an MPM facilitates the reactivation of the meanings of their antecedents. In discussing the results of the experiments, Featherston writes:

> The data clearly suggest that the mechanism of trace reactivation does indeed account for some of the experimental data, since no other account predicts the range of effects observed. This result could lead us to pay greater attention to grammatical and processing models which make use of trace, since their empirical adequacy receives support. (p. 223) … Experiments 2 and 3 are strong results supporting the psychological reality of movement trace since these results are not predicted by any traceless account… (p. 229)

Experiment 4 made use of VPR paradigm. The results of the experiment provided no confirmation of any of the four competing accounts. No depth-of-processing effects were observed and no evidence of *wh*-traces in so-called tough-movement constructions was found.[31] Featherston reasons that the VPR paradigm is simply "blind" to *wh*-traces and concludes, with Fodor (1995), that it may well not be a suitable paradigm for the study of empty categories in sentence processing.

Experiment 5 differed from the rest in that the data consisted of recordings of event-related potentials (ERPs), discussed in Sect. 5.2 above. In addition, it tested specifically for the psychological reality of NP-trace and PRO. The materials were designed to contrast raising, control, and ordinary transitive constructions in German—the latter serving as a baseline measure. Participants were asked to read sentences that, according to Government and Binding theory, contain NP-traces, PRO, or neither.

The results showed significant differences in the ERP signatures of all three constructions. This provides evidence for the claim that the MPM associated with each construction differs from either of the other two. As the sentences were matched for length and other confounding factors, the natural interpretation of the data is that the ERP signatures track the presence of NP-trace and PRO. Featherston draws this conclusion in the following passages:

> The results showed significant differences between all three, but particularly so between raising and equi,[32] where the difference was plain in three successive 200ms time windows. This result is not predicted by the [Head-driven Phrase Structure Grammar] analysis of the difference between raising and equi, since it locates the contrast in the semantic role assignment of the matrix verb. In this way the principled distinction between raising and equi implemented in [the Government and Binding theory and Minimalist grammars] by the use of different ECs is supported. (p. 228)

---

[31] For reasons of space, we have not discussed tough-movement constructions in this chapter. An example of such a construction is 'Larisa is tough to convince', where the adjective 'tough' seems to raise the object of the verb 'convince' to the subject position in the higher clause. There are numerous syntactic analyses of this construction and there is, as yet, no convergence amongst syntacticians as to what its underlying structure is. Here, we simply note that Government and Binding theory treats it as containing multiple empty categories, including *wh*-trace.

[32] To avoid the ambiguity in the term 'control', Featherston refers to syntactic-control verbs as equi-verbs, reserving the term 'control' for the baseline measure in an experiment.

> The pool of evidence for local ECs has lost the visual probe recognition data but gained a new recruit in the form of our ERP result. Since the former is the methodologically least sophisticated and the latter the most sophisticated, we must conclude that the case for local ECs has been strengthened by our studies. (p. 229)

> Our ERP Experiment 5 … offers robust support for the existence of local ECs, since the alternative possible causes of the P600 we found are excluded by our two follow-up experiments. In sum we have a significant body of data suggesting that ECs play a role in on-line processing as well as in the grammar. Since this hypothesis is a strong and testable claim, we feel that any evidence in its favour is a surprising fact and should be treated as a success for linguistics. (pp. 229–230)

5.6 Conclusion

It goes without saying that all of the findings reviewed above can be challenged in a number of ways and may well fail to hold up under further scrutiny. Nevertheless, at present they offer us grounds for the claim that empty categories—including *wh*-trace, NP-trace, and PRO—are psychologically real elements of mental phrase markers. For reasons such as these, Phillips and Lewis (2013) are quite correct to assert that "there is now much experimental evidence [supporting] the notion that real-time processes assemble syntactic representations that are the same as those motivated by grammatical analysis."

Whatever the case about that, one thing is clear: the HSPM is *not* a naïve mechanism. Regardless of the details concerning the structures it builds, the evidence reviewed above establishes beyond reasonable doubt that, in building them, the HSPM has access to sophisticated knowledge of the dependency relations that hold between grammatical antecedents and the syntactic positions to which they are bound. As Fodor (1989) observes,

> long-distance binding of traces should provide many pitfalls for a rough-and-ready processor which relies on informal strategies rather than consulting the information provided by the grammar (except perhaps as an emergency back-up). The hypothesis that the [HSPM] is such a device (Fodor et al. 1974; Bever et al. in press) becomes quite implausible in the face of the speed and accuracy with which the [HSPM] interprets traces. We can conclude, instead, that the [HSPM] is very closely attuned to the grammar of the language. If that is so, then differences in how the processor responds to different (putative) empty categories can be taken seriously as evidence of how they are treated by the grammar. (Fodor 1989: p. 205)

The final remark in this passage is particularly significant for the linguist who seeks to formulate a grammar that is not only descriptively adequate but also psychologically real.

In previous chapters, I argued that descriptive adequacy is a worthy goal, in and of itself, for a syntactician to pursue in constructing a grammar. And, as Devitt (2006a) argues at length, it seems to be the *only* goal that many syntacticians in fact pursue. Nevertheless, it's plain that finding a descriptively adequate *and* psychologically real grammar is a much more exciting prospect. In this chapter, I have laid the groundwork for a defense of the claim that some grammar is, indeed, psychologically real. The results surveyed above make it clear that psychologically plausible parsing models will have to make use of grammatical rules or principles, *in one way or another*, when constructing mental phrase markers. To determine what role, precisely, the grammar plays in real-time processing, we must first lay out the live options and then survey the available evidence in favor of each. I do this in Chaps. 8 and 9. Before turning to these tasks, I will examine some attempts to circumvent the arguments of this chapter and deny the psychological reality of MPMs.

Chapter 6
Two Attempts to Do Without Mental Phrase Markers

Abstract I cast doubt on two proposals for doing without mental phrase markers (MPMs). The first is due to Roger Schank and his colleagues at Yale, who constructed comprehension models that relied almost exclusively on semantic and pragmatic resources. I rehearse the striking and pervasive failures of such models and suggest that similar problems will likely plague newer incarnations in the connectionist tradition. The second proposal for doing without MPMs is Devitt's "brute-causal" conception of language processing, which sees comprehension as a reflex like, associative mapping directly from words and sentences to concepts and thoughts. I argue on empirical grounds that language comprehension, even at the earliest stages, is neither reflex-like nor associative. Setting that aside, I examine Devitt's distinction between responding to a property and representing something as having it, and show that the operations of the HSPM cannot be mere "responses" in the relevant sense. Parsing requires keeping track of prior context, and the clues that are relevant for a successful parse may be arbitrarily far back and in discontinuous regions of the stimulus stream. To explain the HSPM's sensitivity to contextually relevant factors, and the resulting flexibility of its decision-making and behavior, we must posit inference-like transitions between MPMs.

Keywords Roger Schank • Semantic resources • Pragmatic resources • Passive sentences • Passive questions • Embedded passives • Reflex • Associationism • Connectionism • Brute-causal processing • The Language of Thought Hypothesis (LOTH) • Mentalese • Representing vs. responding • Responding vs. responding-as • Filler-gap processing • Sphex • Flexible behavior • Thematic roles • Incremental processing • Frames • Functional-role semantics • Systematicity • "Mostly-semantic" models • Integrated processing models • Linear structure • Hierarchical structure • Binding ambiguities • Syntactic binding • Positional Template Filling (PTF) • c-command • Precedence vs. dominance relations • Embedded/embodied cognitive science • Fodor Bever and Garrett (FBG) • David Hume • Aristotle

D. Pereplyotchik, *Psychosyntax*, Philosophical Studies Series 129,
DOI 10.1007/978-3-319-60066-6_6

6.1 Introduction

In the previous chapter, we saw a number of arguments for the psychological reality of mental phrase markers (MPMs)—explicit representations of the syntactic structure of linguistic input. Neurocognitive data, structural priming experiments, and studies of garden-path and filler-gap processing all point to the same conclusion: Any psychologically plausible model of human language comprehension *must*, at some stage, construct MPMs. My aim in this chapter is to reinforce this conclusion, by casting doubt on two proposals for doing without MPMs.

The first such proposal is due to Roger Schank and his colleagues at Yale, who throughout the 1970s and 80s constructed models that relied almost exclusively on semantic and pragmatic resources in the course of comprehension. Schank et al. hold that "no independent level of syntactic representation is constructed, operated on, or output by the language analysis process" (220). If their "mostly-semantic" models were satisfactory on independent grounds, there would be reason to reevaluate the evidence reviewed in the previous chapter, for we would then have a working example of how natural language can be processed *without* MPMs. As it happens, however, the limitations of Schank et al.'s models are both striking and pervasive. In this chapter, I demonstrate the weaknesses of such models and argue that they are unworkable in light of the available evidence. My broader goal is to survey the *kinds* of problems that such models face—problems that will likely plague newer models, including those implemented in connectionist networks (Rohde 2002).

The models that Schank and his associates developed were rich in semantic and pragmatic resources—representations that encoded not only the argument structures of various predicates, but also a wealth of encyclopedic knowledge. Relying solely on these resources, Schank et al. were able to achieve some promising results. But these achievements turn out to be quite limited, precisely because their models make reference only to the *linear order* of the words in the input string; no representation of hierarchical syntactic structure was used in determining which phrases play which thematic roles. Drawing on Marcus (1984), I show in Sect. 6.2 that such models fail to assign correct interpretations to a wide variety sentences, e.g., those with embedded clauses and gaps. The implausibility of these systems as models of human sentence processing serves as an additional ground for the requirement that any such model make use of MPMs.

The second proposal for doing without MPMs comes from Devitt (2006a), whose "brute-causal" conception of language processing sees comprehension as a reflex-like mapping from sounds to thoughts. The defining claim of this view is that the mapping is mediated not by MPMs, but, rather, by brute associative links between linguistic stimuli and their cognitive counterparts—concepts and propositional attitudes—which are thought of as syntactically structured entities in their own right. Devitt reasons that if public-language expressions have syntactic properties that are largely mirrored by the corresponding "Mentalese" expressions, then language comprehension should be a fairly straightforward process—a kind of cognitive reflex, wherein the HSPM directly activates those thoughts that match incoming linguistic expressions in respect of syntax and semantics.

In Sect. 6.3, I argue that language comprehension, even at the earliest stages, is neither reflex-like nor associative. This places further constraints on what features a psychologically plausible comprehension model must have. I then address more directly Devitt’s skepticism about the role of MPMs in language processing. Though he is, himself, an avowed adherent of a representationalist approach in other areas of cognitive science, the kinds of arguments he offers in the case of language processing are of a piece with a wider anti-representationalist trend in the philosophy of psychology, which attempts to account for cognition by appeal to the notions of embodiment and embeddedness in an environment. Recently, theorists who stress the enactive nature of perception have sought to minimize the role of representations in their models of the mind, or eliminate representations altogether.[1] In rejecting Devitt’s argument for brute-causal processing, my broader aim is to shed doubt on any such proposal. To that end, I examine a distinction between *responding* to a property and *representing* something as having that property, which I suspect drives much recent thinking on this topic. I conclude that the operations of the HSPM cannot be viewed as mere responses, in the relevant sense of the term. This in turn places constraints on the notion of representation that will be operative in subsequent chapters, when we turn to the question of whether the rules or principles of syntax are represented in the mind/brain.

6.2 The “Mostly-Semantics” Models of the Yale School

Schank and Birnbaum (1984) report the results of a nearly two-decade-long effort on the part of linguists and computer scientists at Yale to implement versatile and realistic natural language processing (NLP) systems.[2] The models that emerged from this effort were impressive for their time, and continue to inform work in NLP and in AI more generally. Schank and his associates developed their comprehension models in accordance with a philosophical and methodological outlook that they take to be in direct opposition to the dominant Chomskyan paradigm. They are skeptical of the import of the competence/performance distinction (Chap. 2), and they deny the value of constructing an autonomous syntactic theory, claiming that “it is unlikely that a purely linguistic theory could be *in any sense* adequate” (p. 209, emphasis added). Because they see language as inextricably bound up with thought, they doubt that a purely syntactic analysis is necessary prior to the assignment of meaning.

In laying out more precisely the technical features of their proposal, Schank and Birnbaum draw a helpful distinction between three aspects of a language processing system: (i) the system’s control structure, (ii) the representations it generates, and

[1] For discussion, see Anderson (2003), Clark (2006), Gallagher (2005), Gibbs (2003), Hutto and Myin (2012), Robbins and Aydede (2008), Shapiro (2004), and Wilson (2002).

[2] Schank, R. et al. (1970, 1972, 1975a, b, 1977, 1978, 1979a, b, 1980, 1981, 1984).

(iii) the knowledge-base that it draws on.[3] Specifying a control structure involves saying what inputs the system will receive, what outputs it will produce, and what processes and sub-processes are needed to effect these input-output transitions. Questions then arise about what information is encoded in the representations that the system generates in the intermediate stages of processing, as well as what background information the system needs to have at its disposal in order to compute those representations.

With regard to the system's internal representations, Schank and Birnbaum identify three possible positions.

I. The view of Fodor et al. (1974), according to which "the structural analyses to be recovered are … precisely the trees generated by the grammar."
II. There is no independent level of purely syntactic representation. Syntax and semantics are represented together, in one data structure, but the syntactic aspects of that structure are used for purposes that are distinct from those that require the semantic aspects.
III. There is no independent level of purely syntactic representation. The only legitimate reason to posit syntactic aspects of the representations is if they serve some semantic or pragmatic purpose.

They opt for position (III), which denies the existence of mental phrase markers. This dovetails with their position on control structures. Here, Schank and Birnbaum distinguish four positions.

[a] Syntax and semantics are *completely separable*: Syntactic structure is computed first, without any input from semantic analysis, or from higher-level sources of knowledge.[4]
[b] Syntax and semantics are *nearly decomposable*: Syntactic structure is computed first, but the parser sometimes calls for information from the semantic processor. The semantic processor cannot, on this view, *autonomously* initiate a syntactic process.
[c] Syntax and semantics have a *heterarchical relationship*: They are separate processes, but each can call on the other at any stage of processing. (Schank and Birnbaum cautiously attribute this position to Terry Winograd.)
[d] Syntax and semantics are used in an*integrated control structure*: There's a single control structure that decides whether to use syntactic or semantic resources at any given time. (Schank et al. attribute this position to Marslen-Wilson.)

[3] Following the convention of AI, my use of the terms 'knowledge' and 'conception' is informal throughout. I argue in Chap. 7 that the relevant states are *sub*personal and do *not* require the possession of concepts, at least not in the sense of 'concept' that has become commonplace in the philosophy of mind.

[4] Schank and Birnbaum treat this as a naïve interpretation of Chomskyan grammar as a processing model. In Chap. 2, I provided textual evidence for the claim that Chomsky does *not* intend his syntactic theory as a processing model.

They note that they key difference between [c] and [d] is that the former, unlike the latter, is committed to the construction of distinctly *syntactic* representations in the intermediate processing stages. Schank and Birnbaum opt for position [d], which eschews such representations.

Finally, with regard to the issues concerning the comprehension system’s knowledge base, Schank and Birnbaum note that the range of options depends on one’s prior commitments regarding the need for distinctively syntactic representations. If, for instance, one adopts the position of Fodor et al. (1974)—i.e., position (I), above—then one is thereby committed to endowing the comprehension system with distinctly syntactic rules. On the other hand, if one adopts positions (II) or (III), then the following options open up:

(i) Syntactic rules exist, but they serve no distinct function from any of the other rules.
(ii) There are no specifically syntactic rules at all.

Schank and Birnbaum opt for (i), the weaker of these two claims. In discussing their models, they provide examples of rules that make reference to both syntactic and semantic features of the utterance. It should be noted, however, that their conception of syntax is anemic—restricted, in essence, to precedence relations in linear word order and, in the extreme case, the notions of subject and object. Here is how they summarize the core claims of their “integrated processing models”:

> We are now in a position to state exactly what the integrated processing hypothesis claims. First, it claims that language analysis proceeds as a unitary process, integrating all kinds of knowledge, rather than as a collection of separate processes, one for each kind of knowledge. This is contrast to the models of Fodor, Bever, and Garrett, Woods, and Marcus, among others. Second, it claims that no independent level of syntactic representation is constructed, operated on, or output by the language analysis process. This is in contrast to all of the above models, as well as the model proposed by Winograd. Third, it claims that, although there are rules that are some sense purely syntactic, such rules are not used differentially from other sorts of rule; that is, they are functionally integrated in processing and play no privileged role. This follows from the first two claims. (pp. 220–21)

To motivate the psychological plausibility of their models, Schank and Birnbaum examine several psycholinguistic experiments, which that they take to show that

(i) people do not use the syntactic rules of transformational grammar in the course of language processing, and
(ii) the very strong claim of a completely autonomous syntactic processor is false.

Let us briefly examine two of these arguments.

In support of claim (i), Schank and Birnbaum point out that Schwartz (1980) found no evidence to suggest that hearers use syntactic information to resolve binding ambiguities in pronouns. They take this kind of result to be important because, “processes that might have been thought to depend on explicit syntactic representations can be found not to” (222). But this reasoning relies on a false premise to the effect that the role of syntactic knowledge is to *resolve* binding ambiguities. In actual fact, the role of syntactic knowledge is merely to *rule out* certain binding possibilities, thereby *constraining* semantic analysis. Besides filtering out grammatically impossible binding relations, syntactic representations *leave open* how a pronoun should be bound.

Turning to claim (ii), Schank and Birnbaum adduce the results of a study by Tyler and Marslen-Wilson (1977), who sought to ascertain whether syntactic analysis *within a clause* proceeds by the operation of a completely autonomous syntactic processor. On the model proposed by Fodor et al. (1974), no higher-level knowledge could bear on the process until a clause boundary was reached. Tyler and Marslen-Wilson presented data that suggest that this is not the case. But this does not establish claim (ii), given that both syntactic and semantic analysis are now known to be *strongly incremental*, operating in a morpheme-by-morpheme fashion, even *within* the clause. Compelling evidence for this comes from the neurocognitive studies that we looked at in the previous chapter. Recall how Hahne and Friederici (2002) combined semantic and syntactic violations in one and the same stimulus, by relying on the HSPM to take note of one morpheme just before taking note of another morpheme in the very same word. So even if Fodor, Bever, and Garrett's proposal is incorrect, the question about whether syntactic processing is autonomous (or "modular") remains open.

Putting aside the psycholinguistic data, we can directly assess the models that Schank et al. developed, by looking at how they handle basic linguistic input. As a first pass, let's consider a toy example, where the input is the ordered set of words *Fred ate an apple*, individuated orthographically. The system scans from left to right, activating the concepts associated with each word. In cases of ambiguity, the system activates all of the concepts and attempts to satisfy, in parallel, each of their semantic and pragmatic requirements. When the requirements of a concept cannot be met given the rest of the input, the concept is dropped from the analysis. Here is a walk-through of the system's high-level activities:

1. 'Fred' activates a concept of an animate human male.
2. 'ate' activates the concept INGESTION, which requires an animate agent and an edible object.
3. 'Fred' is bound to the agent position of INGESTION.
4. 'an' activates the concept of indefiniteness.
5. The system predicts the presence of a concept for some type of edible object.
6. 'apple' activates the concept of an edible fruit.
7. 'an apple' is bound to the object position of INGESTION

The final product can be represented thus: [INGEST ACTOR (FRED) OBJECT (APPLE REF (INDEF))].

On account of their almost exclusive focus on semantic and pragmatic processing, the central feature of such models is how they encode high-level information in their knowledge base. It is here that the technical devices known as "scripts," "cases," and "frames" (among others) come into play.

> One of the first … memory structures devised to capture this high-level semantic knowledge … and pragmatic knowledge necessary for language understanding … was the script.

> A script is a temporally and casually linked set of low-level concepts describing a time-ordered stereotypical sequence … paradigmatic[ally], going to a restaurant" (213).

For instance, when the system scans the input and recognizes the word 'restaurant' (or words that express semantically related concepts) it immediately activates the concepts that are typically relevant for understanding a restaurant visit—the concept of a host, a waiter, ordering, eating, china, utensils, the bill, the tip, and so forth. Each of these concepts encodes relatively sophisticated world knowledge, regarding which items are animate, which are edible, and so on.

Schank and Birnbaum describe techniques for scanning input text to find the "most important" words, and using them to activate the relevant concepts. Other input items are either discarded as "uninteresting" or kept in a buffer, awaiting expansion, which occurs if and when the information that they contain becomes needed. This strategy saves on processing time and computational resources. A piece of information comes to be needed when an activated concept is expanded and requests its "slot fillers"—i.e., concepts that would provide the information that is typically associated with the activated concept. For instance, given that a restaurant visit typically involves human participants and the food they eat, the concepts PARTICIPANT and FOOD would be slot fillers for the higher-level concept RESTAURANT. This guides the system in scanning the rest of the sentence for specific kinds of slot fillers, as dictated by the "important" concepts that have already been activated and expanded. In this way, the system can prioritize its activities and save time and resources by ignoring those words that provide no immediate slot fillers for the most important currently active concepts.

The notion of "most important concept/word" is bound to seem obscure at first. It is not clear what principled grounds there can be for determining whether a word is likely to provide important information. Offhand, it seems like *any* word can do that, given the right context. Schank et al. do provide a formal characterization, in terms what they call memory organization packets (MOPs), but the details of these data structures are beyond the scope of the present discussion. (See Schank 1979b.) However, an example will serve to convey the impressive results that such models can achieve. Schank et al. programmed their SAM and FRUMP systems to use MOPs for the task of summarizing stories and newspaper articles. Consider the following input text and the system's final output, generated in the space of several milliseconds.

Input
An Arabic-speaking gunman shot his way into the Iraqi embassy here this morning, held hostages throughout most of the day before surrendering to French policemen and then was shot by Iraqi security officials as he was led away by French officers. (From the New York Times).

Output

** MAIN EVENT **

EVENT 1 =

SCRIPT:	*TERRORISM*
ACTOR:	*ARAB GUNMAN*
PLACE:	*IRAQI EMBASSY*
INTEREST:	*9 OUT OF 10*
CITY:	*PARIS*
TIME:	*MORNING*

SCENES

EVENT 2 =

SCENE:	*HOSTAGES*
PLACE:	*IRAQI EMBASSY*
ACTOR:	*ARAB GUNMAN*
INTEREST:	*7 OUT OF 10*
TIME:	*DAY*

EVENT 3 =

SCENE:	*CAPTURE*
PLACE:	*IRAQI EMBASSY*
OBJECT:	*ARAB GUNMAN*
ACTOR:	*POLICEMEN*
INTEREST:	*6 OUT OF 10*
AFTER:	*EVENT 2*

** UNEXPECTED EVENTS **

EVENT 4 =

ACTION:	*PROPEL*
ACTOR:	*IRAQI OFFICIALS*
OBJECT:	*ARAB GUNMAN*
ITEM:	*BULLETS*
DIR. FROM:	*GUN*
INTEREST:	*5 OUT OF 10*
AFTER:	EVENT 3

RESULT

EVENT 5 =

STATE:	*DEAD*
ACTOR:	*ARAB GUNMAN*
INTEREST:	*4 OUT OF 10*

Source: Schank and Birnbaum (1984: pp. 234–5).

As the output of the system demonstrates, we are dealing with a very sophisticated language processing device. Unfortunately, at the time that Schank and Birnbaum reported these results, the example given above was not indicative of the system's performance in the general case. In actual fact, the successful outputs constituted only 10% of the total (Kinsch 1984). Schank et al. attributed this low success rate to the impoverished theoretical and encyclopedic knowledge encoded in the system's conceptual framework. Accordingly, they undertook the project of building a sophisticated knowledge base—an effort that continues to drive impressive advances in what, today, we call "expert systems."

We have seen that the integrated processing models operate largely on the basis of semantic and pragmatic knowledge. What about their syntactic resources? As noted above, Schank et al. are skeptical that the concepts from contemporary syntactic theories will play a crucial role in a comprehension system.

> Traditional notions of syntax use categories like 'part of speech' and 'phrase structure' in discussing the structure of a sentence. … [W]e would like to claim that these notions of syntax are inappropriate to attempts to describe and use syntactic knowledge in a language understanding process. (p. 228)

They maintain that syntax is only of use when local semantic information and background pragmatic knowledge are not sufficient to establish the relations between concepts. The hope is that such cases will be relatively rare and that, when they arise, information about the linear order of words in the input string will be sufficient to pick up the slack. For instance, the rule that Schank and Birnbaum propose for handling relative clauses—a rule triggered by the occurrence of a complementizer phrase ('which', 'who', 'that', etc.)—states that

> [t]o the right [of the complementizer phrase] will be found a concept with some unfilled slots. Use the concept to the left [of the complementizer phrase] to fill one of the slots, in accordance with semantic requirements. Then take the resulting conceptualization and subordinate it to the concept on the left [of the complementizer phrase]. (p. 230)

Schank and Birnbaum (1984) point out that while this rule is stated in terms of precedence relations—e.g., 'to the left of'—it ultimately has to appeal to semantically defined positions. Their point is that "syntactic and semantic knowledge cannot be distinguished by use in this model of language analysis" (231).

With this sketch in mind, we are now ready to examine the limitations of integrated processing models. The discussion in the following section draws heavily on Marcus (1984). Following Marcus, I will argue that such models, while impressive and potentially illuminating, are not adequate *even in principle* to handle any but the simplest linguistic inputs.

6.2.1 Why the "Mostly-Semantic" Models Don't Work

Marcus (1984) launched an influential criticism of the models that Schank et al. developed. He agrees with his opponents that there are reasons to *hope* that a comprehension model can make do without fancy syntactic resources. Such models would be theoretically parsimonious, making use of no information beyond what is inherent in problem domain—i.e., *linguistic meaning*. And he concedes that contemporary syntactic theories fail to either explain or to even address many interesting natural-language constructions. Nevertheless, he argues that a model that doesn't take into account *hierarchical* syntactic relationships will encounter pervasive failures, even on typical inputs.

> Many workers in both psychology and artificial intelligence hold some form of [what] I call the *no explicit syntax* hypothesis. They therefore dismiss the claims of linguists that a language has specific additional structure, which each speaker of that language knows tacitly but explicitly. … Often these workers … insist … that language can be understood by applying a system armed with only the minimum of information explicitly about the syntax of the language. … I believe that these researchers are wrong. Furthermore, I believe that the models presented to date in support of such claims, to the extent that they remain true to

the claims, are and will remain fundamentally inadequate to handle the range of grammatical phenomena well known and understood within the linguistics community for the last ten years. Perhaps surprisingly, these phenomena are part of the core of the syntax of English… In short, I argue that any model based on the hypothesis that language understanding involves no explicit syntactic structure is fundamentally inadequate to process the full range of natural language. If this is true, such models are simply not candidates for models of the process of understanding natural human language. … [And] if these models fail the prior test of being computationally sufficient, questions of psychological reality simply don't apply." (Marcus 1984: pp. 253–255, underlining added; italics in the original)

Following the presentation of Marcus (1984) we will first look briefly at models that use *only* semantic information. We then turn to those that also use linear order, and examine some simple extensions of such models, which are designed to handle problem cases. As will we will see,

these extensions, which are essentially those incorporated in the published models, are based on limited subcases of [the problematic] phenomena, and fail to handle any of their more complex manifestations. … [T]hese methods suffice to capture only the subcases of these phenomena that involve local syntactic structure (that is, in which all relevant aspects of structure are contained in a single clause), but that these phenomena can affect global structure." (p. 255)

As noted above, systems that make use of only semantic information are able to rule out some incorrect analyses by using semantic and pragmatic information—the meanings of the words in the input, combined with background encyclopedic knowledge. For instance, in *John bit the apple*, the system can tell that John did the biting—not the other way around—because it knows that inanimate things don't bite. But these resources will not always be adequate. Indeed, in the worst cases, reliance on them can lead a system to make serious errors. For example, confronted with the input *The postman bit the dog*, such a system would confidently issue the analysis, THE DOG BIT THE POSTMAN. While this analysis matches the system's pragmatic knowledge to the effect that dogs bite more often than postmen do, the analysis is obviously wrong. There will also be cases where pragmatic knowledge is either unavailable or irrelevant. For instance, a system that doesn't know much about John or Bill, will incorrectly take the sentence *John insulted Bill* to be ambiguous. Marcus draws two morals from this. First, *linear order matters*. Second, and more important:

Examples like [*The postman bit the dog*] rule out as cognitive models *semantics-mainly* approaches that use syntax only *where necessary*—that is, where a semantics-only analyzer runs into trouble. The problem, of course, is that semantics and pragmatic expectations can sometimes give a straightforward, and wrong, analysis. Wrong but highly plausible analyses will always cause such a semantics-mainly approach to overlook a right but implausible analysis. This seems to directly rule out positions like the following, which is "Principle III" in Riesbeck and Schank: "A parser must take care of syntactic considerations only when required to do so by semantic considerations… We simply believe that syntactic considerations should be done only when they are needed, that is, after other more highly ranked considerations are used. We have turned the syntactic approach of 'semantics only when needed' around… (p. 9)." Examples like [*The postman bit the dog*] show that one cannot always tell whether syntax was needed or not." (p. 261)

These considerations give rise to models that are enriched with representations of the linear order of the input string, as well as some rudimentary syntactic notions. Such models use a technique that Marcus calls "Positional Template Filling" (henceforth, PTF). To illustrate how PTF works, imagine a user of a computer system querying the system in ordinary English. Suppose the user is asking about the history a specific file, viz., file 123. The input to the comprehension system is as follows: *Has anyone accessed file 123 in the last week?*

The PTF method draws on semantic entries for predicates, which contain some basic syntactic information. This includes the typical positions of the predicate's arguments—e.g., subject and object positions. For the input sentence we are currently considering, the most important semantic entry is for the predicate 'ACCESS':

(1) ACCESS (subject: USER, object: FILE, {"on," "before," "in," …}: TIME)

A program employing the PTF method begins by scanning the input to find a verb. Having found 'access', the semantic entry in (1) guides the system to scan rightward for the object and the prepositional phrase, and leftward for the subject, resolving any ambiguities it encounters by drawing on the semantic and pragmatic resources at its disposal.

Passive sentences, like *Has file 123 been accessed by anyone in the last week?*, are handled by drawing on *separate* semantic entries for the passive form of each verb. These differ from the active-form semantic entries in that the subject position occupies a different place in the linear order and the object position is optional. The passive counterpart of (1) would thus be:

(2) ACCESS ("by": USER, subject: FILE, {"on," "before," "in," …}: TIME)

The PTF algorithm begins by finding the verb and scanning rightward, to see if the verb carries any passive morphology (e.g., 'ed'). If so, then it scans leftward, to see if there is a passive marker (e.g., 'been'). If these are both found, the passive semantic entry for the predicate 'ACCESS' is activated. This, in turn, instructs the system to scan forward from the verb to find 'by', and to scan backward from the verb to find the FILE in subject position. This strategy requires explicitly representing a bit of syntax—the order of the words and the passive morphology of the verb—but not very much.

Marcus discusses three problems with the PTF method. The first has to with passive constructions. The second problem concerns constructions that have gaps. Finally, he considers the phenomenon of pronominal binding. We will take these in turn.

The problem with the PTF strategy for handling passive sentences is that it incorrectly treats the passive form as a simple clause-bound construction. Consequently, the only cases it can handle are ones in which there is only a single clause and only one verb to be passivized. More complex cases immediately lead the strategy to fail. Consider, for instance:

(3) Is the disk crash suspected to have damaged my files?

When the system encounters (3), it finds the verb 'suspected' and activates the passive semantic entry for this predicate:

(4) SUSPECT ("by": ANIMATE-ENTITY, subject: MENTAL-OBJECT)

Searching for the subject of the suspicion, the system scans leftward and finds 'the disk crash'. One can't very well suspect a disk crash, so this is not a plausible candidate for the category MENTAL-OBJECT.[5] No other candidates are available, so the system fails. Multiple passivization, as in (5), gives rise to still further difficulties.

(5) Were the files that were suspected to have been deleted important?

First, the system will fail to fill in the passive semantic entries for *both* verbs, for the very same reasons that it fails in cases like (3). Second, even if we imagine this problem to have been solved, the system would *still* fail, on account of the presence of a "gap" after 'deleted'. Let us look more closely at the problems associated with sentences that contain gaps.

Consider the following input: *Are any files I accessed last week over 2 months old?* From the perspective of formal syntax, this sentence has a gap, as represented in (6).

(6) Are any files that I accessed ___ last week over 2 weeks old?

The gap is a position in which the object of 'accessed'—viz., 'any files'—would normally be found. However, 'any files' has been "displaced" to the front of the sentence, in accordance with the syntax of yes/no questions in English. Despite its displacement, 'any files' nevertheless serves as the thematic object of 'access'. Presented with (6), the PTF algorithm would first find the verb 'access' and activate its semantic entry. It would then scan leftwards for a subject, whereupon it will correctly locate "I". But the rightward scan for the object will not find 'any files', because 'any files' has been displaced to the left of the verb. Having found no item to fill the object slot of 'ACCESS', the algorithm would simply fail.

A proponent of the PTF strategy might suggest that this can be fixed by searching leftward for a complementizer phrase ("which," "who," or "that") and treating the material to the left as the object of 'access'. There are two problems with this suggestion. First, as (6*) illustrates there won't always be a complementizer phrase for the system to use in this fashion.

---

[5] One can, of course, suspect a disk crash *of* something—e.g., of causing some further calamity—but that's different from simply suspecting a disk crash. The event *disk crash causes calamity* would be a perfectly fine MENTAL-OBJECT; the entity *disk crash* is not.

(6*) Are any files I accessed ___ last week over 2 weeks old?

Second, and more important, the object of 'access' can be arbitrarily far back. Take, for instance, the following input:

(7a) Is the file, which the directory that I created contains, more than 2 weeks old?
(7b) Is the file, which the directory that I created ___ contains ___, more than 2 weeks old?

In this case, a leftward scan from 'that' will yield 'the directory' and a leftward scan from 'which' will yield 'the file'. The PTF system, even if it could be enriched with resources for locating gaps—especially *multiple* gaps in a single sentence—still has no way of knowing which phrases to bind to which gaps.[6] Largely the same point holds with respect to (8):

(8a) Have the announcements I queued to have the mail system send out gone out yet?
(8b) Have the announcements I queued ___ to have the mail system send out ___ gone out yet?

Consider next the fact that a gap in a *wh*-question cannot occur just *anywhere*. As we observed in the previous chapter, there are complex constraints on noun movement. The following case illustrates that movement of a noun out of a clause, as in (10), is acceptable, whereas movement out of a noun phrase, as in (12), is not.

(9) John claimed that Bill visited Sue.
(10) Who did John claim that Bill visited?
(11) John made the claim that Bill visited Sue.
(12) *who did John make the claim that Bill visited?

Now, in order to correctly process wh-questions, a system must know where gaps can appear. And in order to know *this*, it must *at the very least* mark the difference between noun phrases and clauses.[7] When combined with our earlier observations, we see that the system must also explicitly mark the orthogonal distinction between unembedded, singly-embedded and doubly-embedded phrases. The idea that semantic and pragmatic considerations alone will be sufficient to mark these distinctions, or to render them otiose, is very implausible. Marcus concludes that

---

[6] Notice that pragmatic information won't help in this case; files and directories are both the types of things that can be created and that can contain other things.

[7] In actuality, of course, the distinctions are much subtler. But even the crude distinction between noun phrases and clauses is sufficient to demonstrate the point.

> [w]hen we turn to the implemented systems [of the Yale school], we see that without exception they handle the clause-bound special case of passive. All [these] systems, which fail to handle relative clauses and "wh-" questions, also handle only the special case of clause-bound passives. Thus, no one has yet managed to make the 'semantics-mainly' idea actually work." (p. 272)

These points lead one to suspect that *any* syntactic phenomenon whose explanation requires appealing to hierarchical structure, as against mere precedence relations, will cause problems for the PTF model. This includes quantifiers, negation, adverbs, and a great deal more. Capitalizing on this insight, Marcus focuses on binding, which has been known since the late 1970s to be sensitive to the hierarchical dependence relations between the anaphor (e.g., a pronoun) and its antecedent. Consider the following set of data.

(13) File 123 disappeared when it was copied.
(14) When file 123 was copied, it disappeared.
(15) When it was copied, file 123 disappeared.
(16) It disappeared when file 123 was copied.

Note that that 'file 123' can be co-referential with 'it' in all of these cases, except the last. The formal explanation, briefly, is that in (13)–(14), the antecedent does not c-command the pronoun in its governing domain, whereas in (16) it does. It has been demonstrated experimentally that the HSPM is sensitive to the hierarchical structures that underwrite c-command relations. We considered some of the evidence in the previous chapter, particularly the study by Nicol and Swinney (1989). Focusing specifically on binding and anaphora, Frank et al. (2008) provide references to recent studies that confirm the same point.

> [W]ith respect to the anaphoric phenomena under discussion the empirical patterns of coreference possibilities that are found in studies of sentence processing are precisely aligned with those that are predicted by the abstract conditions on co-reference [in (9) and (14) below]. Experimental results from Nicol and Swinney (1989), Gordon and Hendrick (1997), Asudeh and Keller (2001), Badecker and Straub (2002) and Sturt (2003) uniformly confirm people's sensitivity to a c-command-based locality condition in pronoun and reflexive interpretation, along the lines of those in (9) [A reflexive must have as its antecedent a c-commanding noun phase in the same clause] and (14) [A pronoun may not take as its antecedent a c-commanding noun phrase within its own clause]. (p. 11)[8]

What this shows is that, in order to correctly handle sentences like (13)–(16), an effective comprehension system must know when various terms can co-refer. And this requires, at a minimum, keeping track of c-command relations.[9] Lacking such

---

[8] Frank, Mathis, and Badecker are speaking somewhat informally when they state the conditions on co-reference in terms of clauses. A more precise statement would appeal to the notion of "governing domain." See Haegeman (1994) for extended discussion.

[9] Needless to say, in the general case, syntax alone does not *determine* which phrases are co-referential. Judgments of co-reference are typically made in light of semantic or pragmatic knowledge. As noted above, syntactic processes merely *prune* the binding options, in advance of semantic processing.

distinctly syntactic information, the system will err frequently. And, again, there is just no reason to think that semantic or pragmatic knowledge can make up the difference in all, or even a significant number, of cases. The *only* way of *systematically* ruling out the incorrect interpretations is by reference to the hierarchical relations like c-command, which are defined in terms of both precedence *and* dominance.[10]

6.3 Brute-Causal Processing

The defining claim of Devitt's "brute-causal" view of linguistic comprehension is that it is a reflex-like, associative process that does not involve the construction of MPMs. He suggests that we view comprehension as a matter of mapping acoustic representations of the incoming linguistic input *directly* into thoughts. This view relies on the Language of Thought Hypothesis (J. A. Fodor, 1975), according to which propositional attitudes are couched in a linguistic medium, which (following Wilfrid Sellars) many call "Mentalese." The hypothesis is that the neurochemical vehicles of thought are *syntactically structured*. Devitt, who tentatively endorses this claim, argues for a stronger version of it, on which the resemblance between a person's thoughts—her Mentalese sentences, as it were—and her public-language expressions is relatively strong. Bearing this commitment in mind, let's look at Devitt's most detailed statement of the brute-causal model.

> Let me fill out the brute–causal view a bit more, starting with words. The view is that there is a simple association stored in memory between a linguistic word—more accurately, a representation of its phonetic [i.e., acoustic] properties—and the corresponding mental word with the result that the one leads straight to the other without any analysis. This association is established by the regular use in the community of the linguistic word to express the mental word… The same story also applies to familiar sentences—for example, 'What's the time?' and 'How are you?'—and familiar expressions, including idioms—for example, 'kick the bucket', 'butterflies in the stomach'. The view is not, of course, that these simple associations are all there is to the use of these expressions. Thus, pragmatic abilities will be called on in language comprehension to determine the reference of 'you', to remove ambiguities, and so on…But the view is that simple associations are at the core of the processes. This story for words and familiar expressions seems very plausible. And there seems to be no reason why it cannot be extended to syntax. The basic idea again is of simple associations stored in memory, in this case of syntactic structures in thought with similar (implicit and explicit) syntactic structures in language so the presence of the one prompts the other. Each of the structural features of any linguistic sentence—the features captured by structural descriptions … or by phrase-structure trees—is associated with a similar structural feature of mental sentences. This association is established by the regular use in the community of that linguistic structure to express the mental structure, a regularity that also establishes the conventional syntax of that linguistic structure. In comprehension, a person must identify the structure from clues provided by the syntactic category of words (which comes with word recognition), word order, and the like. (Devitt, 2006a, b: pp. 225–6)

[10] Carnie (2010) provides a clear and detailed discussion of various formal definitions of command in terms of precedence and dominance (see esp. Chaps. 3 and 4).

In what follows, I focus on Devitt's claims concerning associationist transitions (Sect. 6.3.1) and reflex-like simplicity and rigidity (Sect. 6.3.2). I then turn to what I take to be the crux of Devitt's case against positing MPMs—his distinction between *responding* and *representing* (6.3.3). I argue there that Devitt makes a fundamental *conceptual* error, having to do with the very notion of mental representation. The idea that mapping a sound to a thought is a kind of nonrepresentational *response* on the part of the hearer cannot be sustained. A mere response cannot do the work of a mental representation.

The confidence that mental representations can be replaced by nonrepresentational constructs is persistent within the philosophical community, particularly in the embedded/embodied camp.[11] I suspect that this confidence underlies much of the skepticism with which many theorists continue to regard claims about the role of mental representations in psychological models (Hutto and Myin 2012). Thus, I intend my assessment of Devitt's proposal to carry wider implications for the debates concerning the nature of mental representations and their indispensability in our theorizing about psychological mechanisms. In my concluding remarks, I articulate some constraints on when it is appropriate to posit mental representations, thus giving more substance to the notion of "representation" that is at play in this discussion, and setting the stage for fuller treatment of the topic in the following chapter.

6.3.1 Is Language Comprehension an Associationist Process?

Devitt's characterization of brute-causal processing makes heavy use of notions from the associationist tradition. There is, however, an issue about what exactly counts as an associationist rule of processing. Fodor et al. (1974) lament the "adherence to a terminological convention whereby *any* [process] is *ipso facto* called 'associative', whether or not there is reason to believe that it has the logical properties characteristic of associations, and whether or not there is reason to believe it satisfies associative laws" (p. 76). Though their worry was much more pertinent in the mid-1970s than it is today, their injunction to specify exactly what is meant by 'association' still holds.

Unfortunately, Devitt does not say much about this. Mainly, the clues he provides are contained in his remarks about the link between associationism and connectionism. These remarks, however, are difficult to square with the practices of present-day connectionists, especially those working on sentence processing. For instance, Devitt quotes Smolensky (1991) as saying that "simple associationism is a particularly impoverished and impotent corner of the connectionist universe." But Smolensky, like many connectionists, *resists* the common characterization of connectionist networks as implementations of old-school associationism. The quoted

[11] I do not mean to suggest that Devitt subscribes to an embedded/embodied view. My claim is merely that his proposal shares some of the difficulties facing such views.

passage continues by pointing out that it is "*a serious mistake* to presume connectionism to be committed to simple associationist principles" (Smolensky 1991: p. 165, emphasis mine). Rohde (2002) elaborates:

> One complaint often leveled against connectionist networks is that they are only able to form stimulus-response associations. Therefore, they should have been discarded as a serious account of human language processing along with behaviorism. While this may be true of the simplest one-layer networks, it is not true of multi-layer backpropagation networks composed of units with non-linear response properties, which is usually what is meant by connectionist networks these days. Multi-layer networks, during training, have the ability to develop higher-level internal representations that represent abstract properties of the environment to which the network is exposed. … The ability to develop abstract representations is … one of the principal factors that sets modern connectionist networks apart from the earlier program of associationist behaviorism, which assumed that only observable stimuli, responses, and reinforcements play a causal role in behavior and learning. (pp. 10–11).

This is particularly true of the psychologically plausible connectionist models of sentence processing that are available today—including a constraint-satisfaction model proposed by Stevenson and Smolensky (2006).

One way to clarify the notion of association is to examine the commitments of classical associationists—Aristotle, Hume, Locke, Mill, and others[12]—or, more pertinently, in the behaviorist psycholinguistics of the early twentieth century. Fodor et al. (1974): Chap. 2 do precisely this. Their conclusions do not bode at all well for the claim that human sentence processing is governed by associationist principles. Let us briefly review two of their arguments.

The classical associationists had it that one mental representation (R_1) is disposed to give rise to another (R_2) to the extent that the things that R_1 and R_2 represent are contiguous in either space or time. The principle of association-by-temporal-contiguity shows up again in the behaviorist tradition, though stripped of commitment to mental representations. It serves as the foundation of conditioning theory, according to which the presentation of two stimuli that are contiguous in time (e.g., bell sounds and food) gives rise to a conditioned response (e.g., salivation triggered by bell sounds alone).

Fodor, Bever, and Garrett (henceforth, FBG) note that if language processing were governed by the principle of association-by-contiguity, then the recognition of "grammatical dependency between two items other than adjacent ones should occur very rarely in sentences" (p. 75). But, as they go on point out, this occurs quite frequently. The HSPM is highly sensitive to dependency relations between syntactic constituents that are arbitrarily far apart in a sentence. Indeed, it *must* be sensitive to such long-distance relations, since "the interpolation of material between syntacti-

[12] See Harnish (2002: Chap. 1) for an illuminating discussion of the historical development of associationist psychology. For original sources, see Hume (1740/2000, 1748/1999). J. A. Fodor (2003) discusses Humean associationism in the context of contemporary philosophy of mind and cognitive science.

cally interdependent parts of sentences is a fundamental principle of every language" (*ibid*).

FBG continue their assault on associationist psycholinguistics by examining the *optional* character of the syntactic relationships that the HSPM must handle. They consider the following pattern of data:

(17) John phoned Mary up.
(18) *John divorced Mary up.
(19) John phoned up Mary.
(20) John phoned her up.
(21) *John phoned up her.

As the contrast between (17) and (18) illustrates, the particle 'up' is dependent on the verb 'phoned'. But the dependence is of a very specific sort. In particular, if the object of the verb is a pronoun, then the particle *must* appear *after* that pronoun (20)–(21), though it *may* appear elsewhere in other cases (9). FBG argue at length that associative principles, such as association-by-contiguity, do not "provide mechanisms for reconstructing the intricate interplay of optional and mandatory decisions involved in specifying the order relations among elements of a sentence" (*ibid*).[13] Having launched these criticisms—alongside several others, which I will pass over—FBG draw the following moral:

> [V]erbalization, like other forms of intelligent behavior, resists treatment of as a special case of a habit. Roughly speaking, the relation of association is locally transitive, probably irreflexive, and probably asymmetrical. In the case of any given organism, it holds between at most a finite number of pairs of behavioral elements, and its strength is highly responsive to the frequency with which the associated items co-occur in the organism's experience. Grammatical relations, on the other hand, are of a variety of different logical types; they characteristically hold of infinite sets of elements, and their mastery appears to be largely independent of the experience of the speaker. In short, grammatical relations differ in practically every conceivable way from associative relations; *it is difficult to imagine a less likely candidate than association for reconstructing the psychological processes of language use*. The interaction between taxonomic grammar and Hullean psychology failed to produce a viable account of the either the structure of natural languages or the psychology of the speaker hearer. (*The Psychology of Language*, pp. 76–77)

I have not attempted to cast doubt on the existence of purely associationist processes; indeed, it may well be that such processes are ubiquitous in the psychology of many creatures. The arguments I have surveyed establish only that *human language processing* does not have this character.

[13] This is best seen as a problem for associationist theories of language *production*, though there are analogues difficulties in the case of comprehension. See FBG, Chap. 2 for detailed discussion.

6.3.2 *Is Language Comprehension Simple, Rigid, or Reflex-Like?*

The brute-causal model has it that comprehension is like a reflex, in respect of its simplicity and rigidity. Devitt explicates this claim by rehearsing the infamous example of the Sphex wasp:

> Daniel Dennett finds what seems to be a nice example in the "simple, rigid and mechanical" behavior of the wasp Sphex as it drags a previously paralyzed cricket to its burrow (1978, pp. 65–66). The wasp leaves the cricket on the threshold of the burrow whilst "checking" inside. If the cricket is moved back from the threshold whilst the wasp is inside then, on emerging, the wasp will drag it back to the threshold and repeat the "checking" procedure. If the cricket is moved again, then the wasp will check again. And so on indefinitely. (Devitt 2006a: p. 55)

The Sphex responds to a specific stimulus in the same exact way, over and over, *every single time*. The stimulus-response connection is both rigid and unique: same stimulus—same response.[14] In such a case, extending our conception of the stimulus—for example, by taking it to be *multiple* consecutive encounters with the same cricket-threshold configuration—would not give us any additional predictive or explanatory leverage. For, even after having encountered that configuration multiple times, the wasp continues to engage in the very same behavior. Hence, taking the stimulus to be a single encounter with such a configuration is sufficient to predict what the wasp will do. The prior events and surrounding context make no difference. Let us now see whether this characterization is applicable to the operations of the HSPM.

We have observed several times now that the HSPM encounters stimuli that are almost always ambiguous, typically at every level of analysis. A key question for parsing theorists is how the HSPM manages to cope with such rampant ambiguity. As we saw in Chap. 5, the HSPM engages in processes of rather impressive complexity. Its processing routines are highly sensitive not only to the syntactic and semantic properties of each incoming morpheme, but also to the syntactic and semantic *context* within which that morpheme appears.

One immediate consequence of the fact that morphemes must be integrated into larger syntactic and semantic units is that *there is no unique response to a specific word or syntactic category*. In one context, a morpheme or a part of speech in the stimulus will be integrated in one way; in a different context, a different way. This feature of human parsing routines stands in stark contrast to the rigid behavior of the Sphex or, for that matter, to *any* reflex. Being struck in a particular spot on the knee *always* gives rise to the same knee-jerk response; and, as Fodor (1983) points out, if your best friend endeavors to poke you in the eye, you will *invariably* flinch. As with the Sphex, the context—prior stimuli or background knowledge—doesn't make a

[14] Dennett (2014: Sect. 72) traces the history of this example and notes that the description of the Sphex that he helped popularize may not be entirely accurate. For present purposes, this doesn't much matter. Even a fictional example can serve to fix the content of terms like 'rigid' and 'reflex-like'.

difference. This, again, is the essence of the S-R connections that Devitt sees as characteristic of brute-causal processes. But parsing is *not like that*. For the HSPM, *context always matters.*

Faced with the flexibility and variability of a system's responses to particular stimuli, one might adopt a strategy that was popular in the heyday of behaviorism—indeed, one that still is popular in the study of animal behavior: Concede that there is no unique cognitive (let alone behavioral) response to a *single morpheme*, but nevertheless insist that the response to *larger* linguistic units—sequences of words, phrases, sentences, etc.—*is*, in fact, reflex-like. But this strategy won't work in the present context; we simply know too much about language processing to take such claims seriously.

To begin with, we know that the HSPM processes linguistic stimuli *incrementally*. Rather than waiting for multiple words to come in before making its decisions regarding attachment and hierarchical dependency, it initiates syntactic analysis as soon as the first morpheme is received. Parsing is, to a first approximation, a morpheme-by-morpheme affair (Hahne and Friederici 2002). Second, the point about ambiguity and context-sensitivity applies *all the way up*, from phrases to sentences. Having encountered, say, a noun phrase in context C_1, the HSPM will perform one set of operations, while in C_2 it will perform another. And, as the processing of globally ambiguous sentences illustrates, the same principle is operative one level higher. Variations in discourse context give rise to different interpretations of the very same sentence-sized stimulus. To take a famous example, in one context a speaker will naturally interpret the string 'Visiting professors can be boring' as a comment about a certain kind of activity; in another context, the same string will be interpreted as a comment about a certain kind of professor.

Needless to say, in any particular fixed context, there is *in fact* one and only one thing that the HSPM *does*—indeed, quite predictably, as we saw in Chap. 5. For instance, upon encountering the string 'Janet knows the student', the HSPM will (*ceteris paribus*) interpret 'the student' as constituting the object of the matrix verb 'know'. Hence, the HSPM will not "expect" the string to continue with material that reveals 'the student' to be the subject of a subordinate clause (as in 'Janet knows the student is working'). Regularities such as this make the scientific study of comprehension routines possible. However, it is crucial to see that *it does not follow from this that parsing is a rigid reflex*. For, as noted above, the HSPM will *not* interpret 'the student' as the object of a matrix verb when the verb is something other than 'know' and will *not* interpret *every* NP as the object of 'know'. The parser's representation of prior context weighs crucially on such decisions.

6.3.3 The Distinction Between Responding and Representing

In light of the discussion in the preceding pages, let's consider a natural objection to Devitt's view, which he anticipates in the following dialogue with a hypothetical opponent.

> **Objection**: The way a word should be processed depends on its syntactic category, on whether it is a verb, noun, adjective, or whatever. What the processor must have then are rules for each syntactic category. It surely does not have a rule for each word of that category. So, it has rules for processing all adjectives, not a rule for each adjective. And such a general rule must operate on a representation of a word as a member of that category, on a representation like <That is an adjective>.
>
> **Response**: Why must the general rule operate on a representation? To represent a word as an adjective, the processor must first recognize that it is an adjective. Whatever clue enables this recognition could simply trigger rules appropriate for adjectives: the processor responds to it as an adjective without representing it as one. Being an adjective is not a local physical property of a word, as Rey likes to emphasize (1997: 128; 2003b: 178–9), and so recognizing a word as having that property may not always be an easy matter (although it mostly is; see 10.6). But it is not made any easier by supposing that it involves representing the word as an adjective rather than simply responding to it as an adjective. (Devitt 2006a: p. 225)[15]

Throughout his response, Devitt explicitly draws a distinction between *representing* something as having a certain property and "*responding* to it *as*" having that property.[16] The notion of responding appears again in his claim that human language comprehension mechanisms "do not operate on metalinguistic *representations* of the syntactic and semantic properties of linguistic items but are *directly responsive*, in a fairly brute-causal associationist way, to these properties" (275, emphases added). In assessing the brute-causal proposal, it is important to get clear on what notion of response Devitt is working with. There are a couple of interpretive options here, depending on what one makes of the 'as' in 'responding to it *as*'. Let's consider each option in turn.

Suppose that we ignore the 'as' for the moment, and focus on the distinction between representing and responding. This distinction is quite real—indeed, trivi-

---

[15] Following up Devitt's reference to Sect. 10.6 of his book, we find the following: "Sometimes, it is easy to tell that an object has a certain relational property because that property is well correlated with superficial properties. This makes it quite easy to tell … many English adverbs, the ones that end in 'ly'. It can also be easy to tell that an object has a certain relational property if learning to identify the object involves learning to identify it as having that property. This makes it quite easy to identify the other English adverbs; identification comes with word recognition. One way or another, it is quite easy to tell the explicit structural properties of utterances although sometimes hard to tell the implicit ones. (pp. 185–6)." This, I think, underestimates the complexity of the parsing task. J. D. Fodor (1989) provides examples of some quite common cases in which it is hard to tell the "implicit" structural properties of a stimulus expression: "[T]he antecedents of NP-trace and PRO are typically normal argument NPs, which cannot, when they are encountered, be recognized *as* antecedents. When the processor encounters the predicate … it can tell that an empty category *may* be coming, but even then it is not certain. Nicol (1988, p. 108ff.) notes that a passive verb is often confusable with an adjective, and a passive *by*-phrase may be confusable with a temporal or locative phrase. Similarly, a raising verb may be homonymous with a non-raising verb (as in *John appeared NP-trace to dislike the cake / John appeared to Susan in a dream*); and a sentence containing PRO may be temporarily indistinguishable from a non-PRO construction (e.g., *John intended PRO to leave / John intended no harm*)." (197).

[16] At other times, Devitt speaks of hearers (or their behavior) as being "sensitive to" various syntactic properties. See, e.g., Devitt (2006a: pp. 32, 73). To avoid confusion, I will continue to work exclusively with the "responding" terminology.

ally so, since not everything is a representational system. When a billiard ball strikes the bumper of a billiard table, the bumper responds to the ball as a massive object. But it would be pointless to say that the bumper *represents* the ball as a massive object. No appeal to representation is necessary in explaining the outcome of the collision. Still, while the distinction is real, there is a difficulty with Devitt's invocation of it in the present context. The fact is that there is simply no *unique* way of "responding to [a word] as an adjective" (*ibid.*) This is the now-familiar context-sensitivity point. The presence of an adjective in the stimulus does not, *by itself*, necessitate or invariably give rise to the same cognitive processes on every occasion. In the context of the string 'The wall is', the adjective 'red' will be attached below the inflectional phrase node (IP) or the tense phrase node (TP). However, if 'red' appears at the outset of a matrix clause—as in, e.g., 'Red, she painted it!'—then quite different processing operations must apply, so that that HSPM can ascertain the relevant attachment options and determine that 'Red' was, in this case, topicalized. And there are many other possible ways of "responding to" an adjective like 'red'. Because parsing is not a reflex, there is no unique context-independent response.

Now, Devitt claims that "[w]hatever clue enables this recognition [of, e.g., 'red' as an adjective] could simply trigger rules appropriate for adjectives: *the processor responds to it as an adjective without representing it as one*" (*ibid.*, emphases added). This makes crucial use of the 'as' locution, which points to a different and, I believe, deeper problem for Devitt's view. We will take a couple of passes at it, focusing first on a purely conceptual point, and then illustrating with a psycholinguistic example.

To see the problem, note first that the "clues" which enable the HSPM to "respond to" an acoustic blast as an adjective are not in the *immediate* stimulus. Such clues are typically "implicit" in the structure of the incoming phrase, or, when they are "explicitly" marked in the surface form of the sentence, they can (and often do) appear arbitrarily far back in the sentence, or even in the broader discourse context. Consider some examples of the syntactic relations that a stimulus expression might bear to the other items in the stimulus stream, awareness of which would enable the HSPM to correctly "respond" to it:

- the stimulus is modifying a verb (for, then it cannot itself be verb)
- the stimulus is governed by a referring expression (for, then it cannot itself be verb)
- the stimulus is bound by a quantifier (for, then it cannot itself be quantifier)

These clues all pertain to relations that are relatively abstract and, crucially, *non-local*. To know whether, e.g., the current stimulus is bound by a quantifier, one must have already "noted" the presence of a quantifier in prior stimuli and "kept track" of its hierarchical relation to the current stimulus.[17]

[17] Note, by the way, that not all quantifiers are explicit in the stimulus. To take just one example, plural nouns sometimes introduce implicit quantification. The universal claim 'Lions are carnivorous' differs from the existential claim 'Lions are in the building' precisely in respect of implicit quantification. (The example is due to Ernie Lepore.)

Now, as the billiard ball example that I gave above illustrates, cases of mere response—as against cases of representation—involve the sufficient conditions of the response being *wholly present* in the moment at which the response is caused. The impact of the ball's force, upon contact with the bumper, is not distributed over various discontinuous regions of time. Nor does the response of the bumper—i.e., to resist the ball and bounce it away—rely crucially on the relations between distinct events arbitrarily far back in time. The immediate impact of the ball is sufficient for that response; prior context is irrelevant. When, by contrast, the presence of a stimulus alone is *not* sufficient to explain the variety and flexibility of a system's behaviors, it is often because the system's sensitivity to prior context is making an important difference to its responses. In issuing its responses, the system is "keeping track," as we say, of various aspects of the prior context—i.e., storing information about prior events in memory. The system brings that information, or those memories—in short, those *representations*—to bear on its own subsequent processing and behavior.

Return now to our observations regarding the HSPM. The "clues" that a system needs in order to interpret an incoming stimulus as, e.g., an adjective, are *not* "wholly present" in the stimulus. For, the stimulus is only a morpheme, but the relevant clues are either arbitrarily far back in time, or not overtly present in the stimulus stream at all. Thus, something in the HSPM has to be "storing" or "keeping track of" these clues, so as to enable the HSPM to *use them as clues* later in time. And because the HSPM is demonstrably sensitive to such a wide variety of interacting contextual factors, we can conclude that it must be keeping track of *a great deal* of information. Crucially, it must also make *intelligent use* of that information, accurately discerning the interactions of the contextual factors themselves. This requires the HSPM to contain many *interacting* representations. We can think of each of these as complex multi-track dispositions, which bear in intricate ways on the processing of "nearby" representations.

The point becomes even clearer when we consider a commonsense example of high-level relational property—say, the property of being an uncle. Suppose, for a moment, that a theorist were to propose the hypothesis that mature adults do not *represent* other people as uncles, but rather merely *respond* to them as such. At first, it is tempting to protest that this theorist's notion of "responding to something as having property P" is just code for "*representing* something as having property P." But, as I urged above, we must resist this temptation. The responding/representing distinction is quite real. The theorist's proposal is thus *not* incoherent or self-contradictory. It's simply *false*.

For one thing, there is no unique or general way of "responding to someone as an uncle." How one responds to an uncle (or even how one ought to) is a matter of what other attitudes one has, both about uncles and about other things. Moreover, the mere presence of an uncle in a person's environment is not sufficient for that person to respond to the uncle "as such." (We will examine the import of the 'as such' locution momentarily.) Rather, the person must first "have in mind" certain *clues* that enable her to engage in whatever responses are "appropriate to uncles." For instance, she must know that the target individual is male and that he has had a sibling who, in turn, has a living child. Plainly, these properties of a person are not wholly present in the immediate stimulus. Indeed, the clues about *them* are typically *also* presented arbitrarily far back in time, and hence must be "stored in memory"—i.e., mentally represented—in order to be useful for guiding uncle-appropriate responses. Once again, all of these points apply directly to the case of the HSPM. There are no unique responses to adjectives, and whatever responses we do observe on any given occasion *cannot* be accounted for by the mere presence of an adjective in the stimulus. A suitably large and properly organized class of background mental representations must be wheeled in to pick up the explanatory slack.[18]

The argument above rests on two observations, both of which I take to be uncontroversial. First, one and the same linguistic stimulus can bear any number of properties that might be relevant to sentence processing. Second, at any point in processing, the HSPM may register the instantiation of any one of those properties, *without registering others*. To take just one example of this ubiquitous phenomenon, consider a case discussed by J. D. Fodor (1989), in which an expression is both a noun phrase and an antecedent of an NP-trace.

(22) These are the children who$_i$ *wh-trace*$_i$ were only very rarely observed *NP-trace*$_i$ in the classroom.

In (22), the noun phrase 'the children' is an antecedent of the NP-trace that appears after the passive verb, 'observed'. Fodor notes that sentences like (22) present special difficulties for the HSPM, because "the antecedents of NP-trace and PRO are typically normal argument NPs, which cannot, when they are encountered, be recognized *as* antecedents" (p. 197, emphasis on 'as' in the original). According to Fodor, when the passive verb in (22) is integrated into the mental phrase marker, the HSPM posits an NP-trace and proceeds to search for a suitable antecedent. As discussed in Chap. 5, the search takes time and computational resources, as reflected in a distinctive ERP signature (Featherston 2001), as well as slower response times on cross-modal priming tasks (Nicol and Swinney 1989) and visual probe recognition tasks (MacDonald 1989; McElree and Bever 1989).

Reflect now on the import of Fodor's use of 'as'. In order to explain the processing lags mentioned above, she takes the HSPM to initially represent 'the children' *as* belonging to one syntactic category, viz., *NP*, but not *as* belonging to another, i.e., *antecedent of the NP-trace*. Moments later, of course, the HSPM cottons on to the fact that 'the children' belongs to latter category as well. Since the stimulus expression is the same, the only way to describe this transition is as a change in the *representational content* of the HSPM's classification of that expression. For, viewing the HSPM as merely *responding* to the stimulus in a brute-causal fashion would require sacrificing the explanatory and predictive leverage that we gain from positing representations.

[18] The analogy between representing uncles (as such) and representing noun phrases (as such) breaks down in the long run. In Chap. 7, I argue that the latter kind of representation, unlike the former, is *subpersonal* and hence *not* a constituent of any personal-level propositional attitudes.

Devitt might suggest, at this point, that we must distinguish responding to a stimulus *as* a noun phrase from responding to that very stimulus *as* an antecedent. This would not, however, rescue the brute-causal model, for, in drawing *this* distinction, we abandon the notion of "brute-causal response" that is central to that model. When the 'as' locution is found to be playing a significant role, the term 'response' is revealed to be simply code for 'represent'.[19] To see this, consider a theorist who concedes that mature adults respond to one and the same individual sometimes *as an uncle* and sometimes *as a father*, but insists that these are mere brute-causal responses to the properties of being an uncle and being a father. The proposal is either false—because such responses are plainly mediated by representations—or it is incoherent, on account of there being no intelligible distinction between representing-*as* and responding-*as*, provided that the 'as' locution is making a genuine contribution to our inferences, predictions, and explanations. I suspect that Devitt's confidence in the brute-causal model is based on a conflation of the responding/representing distinction, which is quite real, and the responding-*as*/representing-*as* distinction, which I have argued is illusory.

6.3.4 What Does All This Tell Us About the Nature of Mental Representation?

How do the conclusions that we've reached thus far bear on broader philosophical issues? Our recent discussion offers an opportunity to clarify the very notion of representation that is at play in this discussion. In the following chapter, we will address the question "What is a mental representation?" head-on. In the service of that discussion, let us be explicit now about the conditions under which mental representations are an appropriate and explanatory posit.

As many have noted, one of the main reasons for positing mental representations is that doing so allows us to explain the *flexibility* of a creature's or subpersonal mechanism's behavior (Pylyshyn 1986; Sterelny 1991). Behavior is flexible when it is contingent upon a large variety of potentially interrelated factors, not all of which are wholly present in the stimuli that give rise to its behaviors. Varying one of the factors may influence the creature's behavior in a particular way, but only on the condition that a great many other factors are kept constant (or also varied, but only in *very* specific ways). If pressed to explain why such contingencies hold in the case of many biological systems—birds, mammals, and the like—the only answer seems to be that such creatures have *internal representations* of each of the environmental factors on which their behavior is contingent. They are, in effect, systems that *store*

[19] Here, I echo Chomsky and Fodor's most important criticism of Skinner: "as Skinner uses it, (at least when he's outside the laboratory), 'response' is really a crypto-intentional term" (Fodor 1990: p. 55).

memories of those factors.[20] This, combined with our earlier remarks about the 'as' locution, suggests that it is appropriate to posit representations in those cases where we wish to explain why one and the same stimulus gives rise to different responses, depending on which of its high-level, relational properties are relevant to its behavior. For such an explanation to be at all satisfactory, the representations that we posit must meet several conditions:

(i) they must be dynamically responsive to changes in the environment
(ii) they must be occasioned by stimuli that are sometimes temporally discontinuous
(iii) they must stand in many-to-one relations to stimuli, so that one and same stimulus can be categorized now as an NP and later as a wh-antecedent
(iv) they must stand in one-to-many relations to stimuli, so that one and the same Noun Phrase representation is occasioned by 'the boy' and 'an angry ostrich'
(v) they must be capable of misrepresenting the stimulus—as in garden-path cases, where the structure of the already-encountered phrase is misrepresented, and the prediction about future input turns out to be incorrect
(vi) they must be in constant interaction with one another, giving rise to behavior *jointly*, not individually

Point (vi) is a version of content holism. I take it to entail that a state would not be the representation that it is—would not have the representational properties that it has—if it were embedded in a different "computational context" or "cognitive economy". While most philosophers of mind would agree that representational content is not an intrinsic property (Dretske 1981), there is much debate about whether inferential or computational role of a state is a determinant of its content. The functional-role semantics that I endorse in the next chapter says that it does. But that view faces well-known objections, which we'll take up.

One final point is in order about the nature of the representations we've been considering. Though it may not be true of all mental representations, the case can now be made that MPMs are *themselves* syntactically structured. To do so, we can make novel use of the arguments that are standardly taken to establish the language of thought hypothesis (LOTH). In his classic treatment of the issue, Fodor (1987) appeals to the notions of productivity and systematicity in arguing for LOTH. Aiming at the same conclusion, Devitt (2006a) argues that the only satisfactory explanation of the rationality of human reasoning is one that appeals to computations that are sensitive to the formal and syntactic properties of thoughts.

> "I think that we should prefer LOTH [over the view that] a mental representation has some nonlinguistic complexity—for example, the complexity of a map. … [I]t is difficult to see how [the latter] could account for the mental process of thinking …. Formal logic gives us

[20] Note that Devitt agrees about the need to posit representations even for quite simple creatures. He cites approvingly the following passage from Gallistel (1990): "bees must represent the angles and distances of food sources not only with reference to the sun but also with reference to prominent features of the terrain surrounding the hive" (p. 132). Devitt notes that "Gallistel's account of the extraordinary navigational skills of insects makes this plausible" (p. 21, fn. 5).

> a very good idea of how thinking might proceed if thoughts are represented linguistically. From its very beginning, computer science has used this idea to build machines that process linguistic representations. In recent years, computer science has developed "connectionist" machines that use representations of a very different sort, if they use representations at all. Despite the striking success of these machines with some forms of problem solving, connectionist processes seem rather far from capturing anything like human inference. We still have very little idea how thinking could proceed if thoughts were not language-like." (Devitt 2006a: pp. 146–7)

There is, I think, an open question about whether arguments from productivity, systematicity, or inferential coherence suffice to establish LOTH. But, whatever the case about that, a moment's reflection reveals that those arguments apply *directly* to the case of MPMs.[21] For, unlike in the case of thoughts, it is uncontroversial that MPMs must be exactly as productive and systematic as the public-language expressions they represent. That is, after all, their domain. Moreover, in Chaps. 8 and 9 we will examine computational models in the "parsing as deduction" tradition, which serve as vivid illustrations of how the operations of the HSPM can exhibit a subpersonal analogue of inferential coherence. Such models strongly suggest that the HSPM is, in a certain qualified sense, a rational device. These considerations make it reasonable to tentatively endorse the view that MPMs are *themselves* syntactically structured representations, and that the operations of the HSPM manipulate them in ways that are sensitive to their syntactic structures.

6.4 Conclusion

I have argued that parsing requires keeping track of prior context, and the clues that are relevant for a successful parse may be arbitrarily far back and in discontinuous regions of the stimulus stream. When a system is responsive not only to the stimulus that is *currently* impinging on it, but also to contextually relevant factors, we posit representations to explain the flexibility of its behavior. On these grounds, I conclude that we should see the internal states of the HSPM as representations. This conclusion is reinforced by the fact that clear, satisfying, and as-yet-unrivaled explanations of both first-pass parsing and reanalysis routinely posit internal states that are responsive to stimuli *as* bearing one high-level relational property rather than another. I have argued that responding-*as* is the hallmark of a representational system, not a substantive alternative.

This chapter marks the final stage of my argument for the psychological reality of MPMs. Having, in the previous chapter, laid out the case in favor of positing

[21] Indeed, the argument for LOTH in Fodor (1987) uses mental phrase markers as a crucial example (pp. 144–5). Though Fodor appears to construe the mental phrase marker as a kind of propositional attitude—a thought or judgment—I will argue in the next chapter that MPMs are *subpersonal* states, hence not a species of propositional attitude, the latter being a personal-level category. Fodor (1975: pp. 52–3) claims, incorrectly in my view, that the personal/subpersonal distinction is irrelevant for the purposes of cognitive science.

them, the discussion in this chapter was devoted to examining models that eschew reference to them. We have now catalogued the empirical failures of such models and isolated the faults in their conceptual underpinnings. In the remaining chapters, my goal will be to ascertain whether a model that posits MPMs is *ipso facto* committed to the psychological reality of syntactic rules or principles—and, if so, in precisely what sense. In the next chapter, I take a closer look at the notion of representation that is relevant in this context, and distinguish representation from embodiment.

Chapter 7
Representation, Embodiment, and Subpersonal States

Abstract In the literature on psychosemantics, fully externalist views hold that nomological brain-environment relations exhaustively determine the representational properties of internal states; computational role has no bearing. Fully internalist views reverse both claims. I argue that there is no overwhelming reason to adopt either view, and that the most promising alternative is functional-role semantics (FRS). Next, I show that the main arguments against FRS fail at the subpersonal level of description—a fact obscured by the "psychofunctionalist" tradition's persistent conflation of personal with subpersonal. I survey six interrelated differences between them. Unlike personal-level states, subpersonal states are not inferentially integrated, never conscious, never expressed in speech, lack mental attitude, are not composed of concepts, and do not resist computational treatment. I then go on to distinguish between representing a rule, embodying it, and conforming to it, providing rough-and-ready characterizations of each phenomenon. Crucially, embodiment is an intermediate notion, distinguishable from the others by conceivable empirical tests. Finally, I discuss the distinction between occurrent and dispositional personal-level states—propositional attitudes such as datable judgments and standing beliefs. I argue that an analogous distinction operates at the subpersonal level, between mental phrase markers and mental syntactic principles, respectively.

Keywords Externalism • Internalism • Two-factor semantics • Inferentialism • Functional-role semantics • Wilfrid Sellars • Jerry Fodor • Robert Brandom • Martin Davies • Edward Stabler • David Lewis • Daniel Dennett • Semantic holism • Normativity • Psychofunctionalism vs. Lewis-functionalism • Personal vs. subpersonal • Inferential integration • Expressing • Mental attitude • The frame problem • Concepts • Consciousness • Representation • Embodiment • Conforming to a rule vs. obeying a rule • Hardwired rules • Connectionism • Procedural vs. declarative knowledge • Implicit vs. explicit knowledge • Common causal factor • Occurrent vs. dispositional • Mental phrase markers (MPMs) • Mental syntactic principles (MSPs) • Kingfisher • Fielders • Implicit learning • Implicit memory

D. Pereplyotchik, *Psychosyntax*, Philosophical Studies Series 129,
DOI 10.1007/978-3-319-60066-6_7

7.1 Introduction

Our goal has been to explore the psychological import of syntactic theory. We have thus far found that the syntactic structure of linguistic expressions plays an important role in language processing. The preceding chapters constitute a sustained defense of the psychological reality of mental phrase markers (MPMs). We saw that ERP studies, structural priming experiments, and research into garden-path and filler-gap processing all point to the same conclusion: A psychologically plausible comprehension model must, at some stage in its processing, construct explicit representations of the syntactic structure of linguistic input. This conclusion was further supported by the failure of the "mostly-semantics" models, developed by Schank and his associates at Yale (Schank and Birnbaum 1984), as well as the "brute-causal" model tentatively advanced by Devitt (2006a). I argued that both of these proposals face insurmountable difficulties, which can be traced to their eschewing MPMs in their models of comprehension.

Establishing the psychological reality of MPMs is important, but it does not get at what many see as the core of the debate over the psychological reality of syntax. For, what is ultimately at issue is whether, and in what sense, *the rules or principles of a grammar* are psychologically real. This is the issue I propose to explore here. Our broadly naturalistic methodology counsels us once again to look at the available empirical results and the viable models of language processing. We turn to these in the next chapter. In this chapter, we lay the groundwork by distinguishing several positions on the psychological reality of grammars. Below are the four positions that I take to be the main options, though we'll consider various refinements as we go along.[1]

REP-GRAM-PROC The structure rules of a language—i.e., the syntactic rules or principles that comprise its grammar—are (i) identical with the processing rules that govern the HSPM, and (ii) are *explicitly represented* in the competent speaker's mind/brain.

REP-GRAM-DATA The structure rules of a language (the grammar) are (i) distinct from the processing rules of the HSPM, but (ii) are *explicitly represented* in the competent speaker's mind/brain and used *as data* by the processing rules.

EMB-GRAM-PROC The structure rules of a language (the grammar) are (i) also the processing rules of the HSPM, but (ii) they are *embodied,* not explicitly represented, in the competent speaker's mind/brain.

---

[1] The first three of these positions correspond to what Devitt (2006a) calls positions **(i)**, **(ii)**, and **(iii)**, respectively. The last corresponds to what he calls position **M**, for 'minimal'. My labels are intended to serve as mnemonics: The distinction between representing and embodying is encoded in the labels '**REP**' and '**EMB**', respectively. The distinction between treating syntactic principles as a processing algorithm, on the one hand, or as data, on the other, is encoded in '**GRAM-PROC**' and '**GRAM-DATA**', respectively. The label '**CONFORM**' is self-explanatory, and equivalent to Devitt's term 'respect'.

GRAM-CONFORM Human language processing *conforms* to a grammar, in the sense that the competent speaker's mind/brain reliably takes sentences that are licensed by the grammar as inputs and produces such sentences as outputs.

A central feature of this taxonomy is the tripartite distinction between *representing* a rule, *embodying* it, and *conforming* to it. Thus, to make any progress on this issue, we must first get clear on these notions. We've noted some clues in the preceding chapters, but it's now time to put them together into an explicit framework for thinking about the theoretical posits of psycholinguistics. In Sect. 7.2.1, I locate my notion of representation within the space of available options. In Sect. 7.2.3 and 7.2.4, I contrast this notion with that of embodiment. The intervening Sect. 7.2.2 makes explicit the grounds for distinguishing between personal and subpersonal levels of description. I will suggest that the failure to draw this distinction, and to keep clearly in mind its relation to various orthogonal distinctions (conscious/nonconscious, implicit/explicit, procedural/declarative, etc.), has been the source of much needless controversy.

7.2 Theories of Mental Representation

7.2.1 Why Take a Stand on Mental Representation?

The present discussion has been concerned with the central issues raised by Devitt (2006a). Devitt's main goal was to argue that we have, at present, no grounds for what he dubs the Representational Thesis:

(RT) A speaker of a language stands in an unconscious or tacit propositional attitude to the rules or principles of the language which are represented in her language faculty.

Evaluating RT requires first getting clear on what notion of representation is at issue. Here is Devitt's characterization:

> [T]alk of representing rules raises a question: What sense of 'represent' do I have in mind in RT? The sense is a very familiar one illustrated in the following claims: a portrait of Winston Churchill represents him; a sound /the President of the United States/represents George W. Bush; an inscription, 'rabbit', represents rabbits; a certain road sign represents that the speed limit is 30 mph; the map on my desk represents the New York subway system; the number 11 is represented by '11' in the Arabic system, by '1011' in the binary system, and by 'XI' in the Roman system; and, most aptly, a (general-purpose) computer that has been loaded up with a program represents the rules of that program. Something that represents in this sense has a *semantic content*, a *meaning*. When all goes well, there will exist something that a representation refers to. But a representation can fail to refer; thus, nothing exists that 'James Bond' or 'phlogiston' refer to. Finally, representation in this sense is what various theories of reference—description, historical-causal, indicator, and teleological—are attempting to partly explain. (Devitt 2006a: p. 5)

Devitt holds that, for the purposes of assessing RT, nothing more substantive needs to be said. In particular, we can stay neutral on which *theory* of representation is correct.

> I am attempting to clarify my sense of 'represent' (and its cognates) by distinguishing that sense from other ordinary ones. I am not, of course, attempting a theory of representation, a theory of what 'represent', in that sense, refers to. The concerns of this book do not require such a theory (which is just as well given how difficult it is proving to come up with one!). (Devitt 2006a: p. 6, fn. 7)

This approach faces a difficulty. The less committal one's account of representation is, the more difficult it becomes to evaluate one's claims regarding RT, and about the psychological reality issue more broadly. In declining to set substantive constraints on the nature of representation, one risks either begging important questions or advancing merely verbal points. To avoid these problems, I propose to sketch the logical space of theories of representation, and to lay out my own commitments on that topic.

7.2.2 Internalism, Externalism, and the Functional-Role Via Media

Theories of mental representation are often categorized as either internalist or externalist. The labels can useful, but only as marking the poles of a dense spectrum. Let's start, then, by examining the extremes, with a view toward finding a reasonable middle ground.

Jerry Fodor's well-known "asymmetric dependence theory" (Fodor 1990) can be characterized as *fully externalist*, for it depicts representation as an *externalistic relation* between a creature and its environment. For present purposes, the details of this theory are less relevant than its contrast with a fully internalist position. The clearest example is Chomsky (2000), who argues forcefully that perceptual psychologists are engaged in "internalist" or "individualist" inquiries, which deal solely with the intrinsic properties of the mind/brain, irrespective of the relations that it bears to anything in the environment. To get a feel for the issue, it is instructive to quote Chomsky at length.

> Is there a problem with internalist (or individualist) approaches to other domains of psychology? So it is widely claimed, but on dubious grounds, I think. … [T]he study of visual perception along lines pioneered by David Marr (1982), … has been much discussed in this connection. This work is mostly concerned with operations carried out by the retina or, loosely put, the mapping of retinal images to the visual cortex. Marr's famous three levels of analysis—computational, algorithmic, and implementation—have to do with ways of construing such mappings. Again, the theory applies to a brain in a vat exactly as it does to a person seeing an object in motion. The latter case has indeed been studied, in work of Marr's collaborator Shimon Ullman (1979). His studies of the determination of structure from motion used tachistoscopic presentations that caused the subject to see a rotating cube, though there was no such thing in the environment; "see," here, is used in its normal sense, not as an achievement verb. If Ullman could have stimulated the retina directly, he would

> have done that; or the optic nerve. The investigation, Ullman writes, "concerns the nature of the internal representations used by the visual system and the processes by which they are derived." The account is completely internalist. There is no meaningful question about the "content" of the internal representations of a person seeing a cube under the conditions of the experiments, or if the retina is stimulated by a rotating cube, or by a video of a rotating cube; or about the content of a frog's "representation of" a fly or of a moving dot in the standard experimental studies of frog vision. No notion like "content," or "representation of" figures within the theory, so there are no answers to be given as to their nature. The same is true when Marr writes that he is studying vision as "a mapping from one representation to another, and in the case of human vision, the initial representation is in no doubt—it consists of arrays of image intensity values as detected by the photoreceptors in the retina" (Marr 1982: 31)—where "representation" is not to be understood relationally, as "representation of." … The internalist study of language also speaks of "representations" of various kinds, including phonetic and semantic representations at the "interface" with other systems. But here too we need not ponder what is represented, seeking some objective construction from sounds or things. The representations are postulated mental entities, to be understood in the manner of a mental image of a rotating cube, whether it is the consequence of tachistoscopic presentations or a real rotating cube, or stimulation of the retina in some other way; or imagined, for that matter. Accessed by performance systems, the internal representations of language enter into interpretation, thought, and action, but there is no reason to seek any other relation to the world, as might be suggested by a well-known philosophical tradition and inappropriate analogies from informal usage. (Chomsky 2000: pp. 158–160)

In saying that "[n]o notion like 'content,' or 'representation of' figures within the theory [of visual perception]," Chomsky appears to be claiming that there are two different notions of representation—the traditional philosophical notion (which Devitt and Fodor seem to have in mind) and a distinct technical notion that plays a role in cognitive science. The disagreement that Chomsky is registering can equally be characterized as one about what theory best characterizes the properties of the internal states that cognitive scientists are concerned with. Chomsky's view is that the representational properties of the states involved in vision and language processing are determined *not* by the causal or nomological relations that these states bear to the environment, but *solely* by the computational role that such states play in internal cognitive processes. This squares with Chomsky's claim that "internal representations of language enter into interpretation, thought, and action, but there is no reason to seek any other relation to the world…" (*ibid*).

Between these two extremes, we can locate a number of views that attempt to accommodate the considerations that militate in favor of both internalism and externalism. One example of such "two-factor" theories can be found in Block (1986), who distinguishes between two kinds of representational property—"wide content," which consists of objects and properties in the environment, and "narrow content," conceived as an inferential or computational role that determines a function from a state's environmental context to its wide content (cf. Fodor 1987).

Another popular "two-factor" approach takes the representational properties of only *some* states—e.g., perceptual judgments and motor commands—to be determined solely by their causal or nomological relations to the environment. Once the representational properties of these semantically basic states are fixed by (or "grounded in") the environment, the representational properties of *non*-basic states—e.g., standing beliefs, desires, suppositions, fears, etc.—can be accounted for by their causal or inferential relations to the basic states, which make more direct contact with the environment. One might also hold that particular *concepts*, whether used in perception, inference, or action, have their representational properties determined by causal relations to the world, while other concepts have their contents determined by inferential relations. Devitt (2006a) expresses such a view in this following passage:

> I think, though Fodor (1987) does not, that some [mental] words will surely not be basic but rather covered by a description theory; <bachelor> is a likely example. Meanings for these words will come from inferential associations with others; for example, <bachelor>'s association with <adult>, <unmarried>, and <male>. (p. 156)

A third type of two-factor view, which stems from the work of Wilfrid Sellars, has it that the representational properties of *every* internal state (and, by extension, every concept) are exhaustively determined by three factors: (i) the environmental conditions under which it is typically elicited, (ii) the causal relations it typically bears to other internal states, and (iii) the behaviors that it disposes a creature to perform. Contrary to the standard versions of behaviorist, covariationist, and information-theoretic accounts of representation, Sellars (1963b, 1974) held that none of the three factors is individually sufficient; only a combination of all three can be both necessary and sufficient for something to have the representational properties that it does (see also Brandom 1994: ch. 2; 2009: ch. 7). Following familiar nomenclature, let us refer to this sort of view as *functional-role semantics*. Since this is the view to which I myself subscribe, I will say a bit more about it, locating it in logical space relative to the other theories mentioned above.

Unlike some of its competitors, functional-role semantics is not committed to the claim that there are two kinds of content, narrow and wide. Rather, the idea is that a state's causal relations to the environment—both on the input and the output side—together constitute one *determinant* of its representational properties, while the internal computational or inferential role of the state constitutes the other determinant.[2] Thus, although both the external and the internal factors are relevant to the issue of what representational properties a given state has—like the components of a single vector—those representational properties are themselves univocal; there is no distinction between narrow and wide content.[3]

---

[2] Neale (forthcoming) points out that there is a crucial ambiguity in the word 'determine'. Sometimes, the term has a metaphysical reading, as in "the molecular structure of this glass determines the degree of its fragility." But, at other times, it can be used epistemologically, as in "The veterinarian determined that the goat's ailment is curable." The two senses of the word are close to 'constitute' and 'ascertain', respectively. In discussing what determines the representational properties of an internal state, I have in mind the metaphysical reading. However, I also believe that in ascertaining what representational properties a state has, we typically take into account external conditions, the internal role of that state vis-à-vis other states, and the behavioral consequences of being in that state.

[3] This is not to say that there is no distinction between *de dicto* and *de re* ascriptions of propositional attitude. The distinction is quite real, but it has to do with the differences between the communicative functions of *de dicto* and *de re* ascriptions. The discursive commitments involved in a

Similarly, the version of functional-role semantics that I endorse does not draw a distinction between basic and non-basic states or concepts. The theory entails that no state or concept has its representational properties fixed *solely* by its causal or nomological relations to the environment. While it is true that perceptual judgments and motor commands enjoy more direct causal relations to the environment and to behavior than do other states, the functional-role theorist claims that such states would not have any representational properties if they entered into *only these* causal relations. In order for an internal state to have representational properties at all, it must, in addition to bearing causal relations to the environment, also enter into computational or inferential relations with other internal states. Thus, even a perceptual judgment has its representational properties partly in virtue of the fact that it serves as a premise in inference and has a causal bearing (however indirect) on action. Functional-role semantics thus rejects what one might label "semantic foundationalism".

An initially counterintuitive aspect of the functional-role view has been explored in detail by Sellars (1974) and, more recently, by Brandom (1994, 2009) and, in another guise, Matthews (1994, 2007). On this view, to say of some state, S, that it represents X is not to claim that S bears a particular relation to X. Rather, the claim that S represents X is used to mark that state as *belonging to a particular type*—i.e., as playing a distinctive role in the creature's cognitive and behavioral economy.[4] There are, of course, many important causal and nomological relations between a creature's psychological states and its environment, both on the input side (perception) and the outputs side (action). And on the functional-role view, these relations constitute two thirds of an account of those states' representational properties. Nevertheless, no direct mind-world relation is exhaustive of a state's having the representational properties that it does. Brandom (1994) makes this point in his discussion of the closely related notion of reference.

> No doubt, as with any other two items in the causal order, there are many relations that can correctly be said to obtain between a term tokening and what it refers to. But the present considerations show that talk about referring and referents provides no reason whatever to conclude that some one of these could be singled out as *the* reference *relation*—that unique semantically significant word-world relation in virtue of which the nonexpression is the referent of the expression. Various word-world relations play important explanatory roles in theoretical semantic projects, but to think of any one of these as what is referred to as "the reference relation" is to be bewitched by surface syntactic form. (Brandom 1994: p. 325)

---

de dicto ascription are entirely those of the person to whom the mental state or speech act is being ascribed; with *de re* ascriptions, by contrast, the ascriber herself takes on some of those commitments. For instance, in saying 'Larisa thought the ghost was gone', one imputes ontological commitment to ghosts to Larisa, whereas in saying 'Of the ghosts, Larisa thought that *they* were gone', one takes on the ontological commitment oneself. The content clauses differ not in what type of mental content is being attributed to the speaker–wide or narrow–but, rather, in what social function they play. See Brandom (Making it explicit. Harvard University Press, Cambridge, MA, 1994: ch. 8) for a detailed discussion.

[4] This is what Sellars (1974) means to capture with the slogan that "meaning is functional classification." Matthews (2007) works out the idea in impressive detail, under the rubric "measurement theory of the attitudes," which has its roots in the work of Churchland (1988) and Dennett (1997), and even as far back as Goodman (1949).

Brandom goes on to argue that referring is best seen as something that people (and perhaps other creatures) *do*—a kind of communicative act, not a relation between linguistic expressions and objects, properties, or events.[5] But the functional-role view that we have been exploring extends such remarks to the notion of representation, where the contrast with Fodor's externalist position becomes clear.[6] Indeed, as I have characterized it here, functional-role semantics is a middle-ground position that contrasts most starkly with the two poles of the abovementioned spectrum—Fodor's full externalism and Chomsky's full internalism. From the perspective of the functional-role theorist, each of these views leaves out a crucial component of what makes an internal state a representation. Let me say a word about both.

The functional-role theorist holds that full externalism does not sufficiently acknowledge the importance of the internal structures and processes that, on his view, partly determine the representational properties of, e.g., perceptual judgments. Of course, to show this conclusively requires a great deal of argument, which I cannot pursue here. But a forceful case has been made by Dennett (1969: pp. 76–8) and Churchland (1979: chs. 2–3) that a neural event's being reliably elicited by something is not sufficient for that neural event's being a representation of that thing (cf. Dennett 1978, 1987). In order to be a representation of food, for instance, a state must in addition to being caused by the presence of food, cause the right sorts of representations downstream and, eventually, behaviors that are, in some sense, appropriate specifically to food (for that creature).[7] Brandom (2009: ch. 7) makes a similar point, in his discussion of the important difference between reliable *detection* and genuine *representation*.[8] Finally, it is worth noting that the project of natu-

[5] To be clear, no "idealism" or "anti-naturalism" follows from this position alone. Contra Sterelny (1991), I believe that naturalism should not be held hostage to the view that any of the numerous relations that we bear to the extra-mental world in perception, thought, and behavior is usefully labeled "the reference relation" (Price, 2011). Nor need there be such a relation for us to sustain the eminently reasonable doctrine of scientific realism. See Devitt (1996) and Chomsky (2000) for two very different defenses of these claims.

[6] The discussion in the main text presumes that representation is a more inclusive notion than reference. In defense of that assumption, the following can be said: Whereas talk of reference has its home in personal-level descriptions of speech acts and perhaps propositional attitudes, extending the notion to other cases—e.g., tree rings, computer programs, maps, pictures, and subpersonal psychological states—yields awkward consequences. For instance, the following claims would not fall comfortably on untutored ears: "This tree ring refers to the age of the tree." "This portrait refers to Plato." "This computer program refers to the restaurants in the neighborhood." "This (state of your) HSPM refers to a noun phrase." By contrast, the notion of representation extends comfortably to all such cases. Ordinary-language intuitions aren't worth much, of course, but they can, on occasion, be indicative of theoretically important distinctions.

[7] Dennett is, no doubt, following Wilfrid Sellars on this point, for whom so-called "language-language" and "language-exit" transitions (i.e., inference and action) were as important as "language-entries" (i.e., perception) in determining the intentional content of a mental state.

[8] Brandom's position requires that the state in question be usable as a premise and conclusion in full-blooded inference. Dennett, by contrast, leaves open the possibility that the downstream consequences of the state are not always personal-level processes such as inference, but are sometimes subpersonal analogs of inference. In the next section, I will argue that the states and processes involved in parsing and comprehension are subpersonal.

ralizing a fully externalist notion of representation has given rise to a number of theories in recent decades,[9] many of which are known to face potentially insuperable difficulties.[10,11]

What about full internalism—the position that Chomsky articulates in passages like the one quoted above? Chomsky focuses solely on psychological processing, ignoring the role that the environment and behavior play in determining the representational properties of internal states.[12] His reason for doing so seems to be that

---

[9]The "asymmetric dependence" approach is an alternative developed by J. A. Fodor (1990). Fodor's theory builds on the "informational semantics" introduced by Dretske (1981). Teleosemantic theories owe much of their popularity to Millikan (1984). An interestingly different version of teleosemantics is advanced by Cummins (1996a).

[10]A blog post by David J. Chalmers (June, 2011) describes the current state of play as follows: "On the sociology, for what it is worth: my sense is that the view attributed to the three 'Rutgers people' in the initial post [i.e., that "it is accepted wisdom that psychosemantics was a failure; no one ever properly solved the disjunction problem let alone naturalize semantic content"] captures the received wisdom in many parts of the profession, at least concerning causal and teleological approaches to psychosemantics. Of course received wisdom is often wrong; and even the received wisdom here allows that some future approach to psychosemantics might succeed. I don't think there's any received wisdom as to what is missing, but certainly one strand of thinking (with which I'm sympathetic) is that any successful psychosemantics, even externalist psychosemantics, must put more weight than these programs on internal factors such as inferential role…" http://philosophyofbrains.com/2011/06/20/was-psychosemantics-a-failure.aspx

[11]Nearly every publication on the topic of intentional content begins with a series of objections to each of the other available theories. Cummins (1991) and Loewer (2002) provide useful summaries of several key objections to most extant views; a more up-to-date catalogue of objections can be found in Neander (2006). Godfrey-Smith (2004) captures what seems to be a widespread sentiment in the following passage: "In the 1980s the problem of giving a naturalistic theory of mental content beckoned young philosophers like myself; this looked like a philosophical problem that was both fundamental and solvable. ... Roughly 20 years on, how has the project fared? With some sadness and much caution, I suggest that things have not gone well for the Dretske-Fodor program. I doubt that we will ever see a satisfactory version of the kind of theory that Fodor's *Psychosemantics* and Dretske's *Explaining Behavior* tried to develop. Despite this, I do think we have learned a lot from the development of this literature. Some good partial answers may have been given to important questions—but not the exact questions that Dretske and Fodor were trying to answer. So I think it is time to start looking at different approaches to the network of questions surrounding belief and representation. This rethinking will involve looking again at some of the ideas of the nay-sayers of the 1980s, like Dennett and Stich, but looking further afield as well" (pp. 147–148).

[12]I take Chomsky's view to be a version of what Harman (1987) calls "solipsistic conceptual-role semantics." Brandom (1994) calls the view "hyperinferentialism"—a position that he contrasts with his own view in the following passage: "There are three different ways in which one might take inference to be of particular significance for understanding conceptual content. The weak inferentialist thesis is that inferential articulation is *necessary* for specifically *conceptual* contentfulness. The strong inferentialist thesis is that broadly inferential articulation is *sufficient* for specifically conceptual contentfulness—that is, that there is nothing more to conceptual content than its broadly inferential articulation. … [S]trong inferentialism as it is worked out in the rest of [*Making It Explicit*] is *not* committed to the hyperinferentialist thesis, which maintains that *narrowly* inferential articulation is sufficient for conceptual contentfulness of all sorts. [T]he broad conception includes the possibility of noninferential circumstances and consequences of application. In this way, the specifically *empirical* conceptual content that concepts exhibit in virtue of

the range of stimuli that gives rise to particular representation is highly variable and unsystematic, and therefore irrelevant to the scientific classification of mental phenomena. But this line of reasoning can be resisted.

First, one might argue that the range of stimuli is not as variable as Chomsky's remarks suggest. In discussing visual processing, he mentions tachistoscopic displays, direct manipulations of the retina or the optic nerve, and the endogenously-caused phenomena that we sometimes refer to as "visualizing" or "imagining." The last of these seems importantly different from cases of perceptual judgment, hence not obviously relevant to the present discussion.[13] The other two cases involve *highly specific* manipulations of the mechanisms of the visual system. Not just *any* tachistoscopic presentation will elicit the same perceptual judgment, and neither will just any manipulation of the neurophysiology. The commonalities between the stimuli and physiological interventions that elicit the same perceptual judgment are very much of interest to the vision scientist. Nothing in the passages that Chomsky quotes from Ullman or Marr seems to contravene this.

Second, suppose that we concede that the range of stimuli that elicit any perceptual judgment is simply too heterogeneous to, *by itself*, underwrite a scientifically useful account of that judgment's representational properties. Would it follow that the stimuli are strictly *irrelevant* to determining the content of the representation? I do not see that it would. The strongest conclusion that would be warranted is that our theory of perceptual representation must take into account *both* stimuli *and* factors internal to the processing mechanism (and perhaps also the downstream behavioral consequences), the latter serving to explain why such otherwise varied stimuli give rise to such uniform behavioral and psychological effects. This conclusion is, of course, precisely the core of functional-role semantics.

Much more needs to be said—and, indeed, has been said—to flesh out and motivate the functional-role approach.[14] Again, I will not undertake that project here. I do, however, wish to forestall a common complaint against theories of representation that take inferential or computational role as a determinant of a state's representational properties. The worry is that each cognitive economy is so different from every other that substantive generalizations cannot be defined over representational states. Addressing this objection, and obviating any number of others, requires

---

their connection to language entries in *perception* and the specifically *practical* conceptual content that concepts exhibit in virtue of their connection to language exits in *action* are incorporated into the inferentialist picture. … The hyperinferentialist about conceptual content (adopting a position not endorsed here) would allow only *inferential* circumstances and consequences of application" (pp. 131–132).

[13] It is, in any event, arguable that the representational properties of endogenously-caused visual imagery can only be fully cashed out in terms of their relations to sensations, perceptual judgments, thoughts, and other phenomena, including outward behavior and mind-independent states of affairs.

[14] Different versions of functional-role semantics are discussed in Sellars (1953, 1963a, 1963b, 1974), Dennett (1969, 1978, 1987), Field (1978), Stich (1985), Braddon-Mitchell and Jackson (1996), Horwich (1998, 2005), Harman (1975, 1982, 1987, 1999), Brandom (1994, 2009), and Greenberg and Harman (2006).

drawing a clear distinction between the personal and subpersonal psychological levels of description.

7.3 The Personal/Subpersonal Distinction

7.3.1 Eight Marks of a Subpersonal State

Following Sellars (1963a), I hold that the functional-role theory of *intentional content*—one type of representational property, but certainly not the *only* type[15]—applies, in the first instance, to speech acts. But the need to account for speech and rational action gives rise to the theoretical posit of internal states—the propositional attitudes. These are personal-level states, whose role in a creature's cognitive economy is captured by a suitable formulation of folk psychology (Lewis 1972). The folk-psychological posit of propositional attitudes suffices only for a relatively coarse-grained way of describing a creature. Digging deeper, one wants to know how a creature can so much as *have* such states. Here, the strategy of attributing *sub*personal mechanisms becomes useful.[16] Focusing specifically on the case of language comprehension, we can say the following: Whereas folk psychology lets us talk about the sensation of a sound giving rise to linguistic comprehension—a judgment to the effect that a speaker said that *p*—cognitive psychology tells us what happens *between* the sensation and the judgment.

Dennett (1987) draws a useful distinction between taking "the intentional stance"—i.e., using folk psychology to predict, explain, and describe a creature's behavior—and adopting "the design stance," which involves thinking of a creature as an aggregate of purposeful mechanisms, each of which has the function of performing a specialized task. It's from the design stance that we isolate the HSPM and attribute subpersonal states to it. These states have *some* of the features that we take to be characteristic of personal-level propositional attitudes. In particular, they bear systematic relations to the environment, to behavior, and to one another. This is what makes it both reasonable and useful to think of them as representations. Doing so allows us to abstract away from the largely unknown neural mechanisms that underpin language processing and to see the causal relations between these states as resembling the inferential relations that hold between propositional attitudes. This, in turn, allows us to *rationalize* subpersonal mechanisms—i.e., to understand them as being engaged in purposeful activities and as taking reasonable cognitive steps

[15] There are other types of mental representation—e.g., sensory representation. Rosenthal (2005: Chap. 7) argues, compellingly to my mind, that our account of the representational properties of purely qualitative sensations must differ from our account of the representational properties of purely intentional states.

[16] Theorists who maintain a firm distinction between the personal and subpersonal levels of description (though not always in those terms) include: Dennett (1969), Stich (1978b), Egan (1991), Collins (2008b), and Hornsby (2000, 2001).

toward accomplishing their goals. These similarities between personal and subpersonal states make it tempting to think of the latter as *propositional attitudes* with intentional content. But that would be a mistake, for there are at least eight salient differences between the two kinds of state. In elucidating these differences, I will use mental phrase markers (MPMs) as my stock example of a subpersonal state.

First, subpersonal states are not "inferentially integrated" with personal-level states (Stich 1978b). We cannot draw personal-level inferences whose premises are MPMs. There may well be evidential relations between an MPM and the price of oil futures, but no one can ever make that connection. It's hard to credit the idea that anyone—even an exceptional psycholinguist—reasons *with* MPMs at the personal level, regardless of how much practice they have reasoning *about* them.

Second, subpersonal states are *always* nonconscious, whereas personal-level states are sometimes conscious and sometimes not. For this reason, the conscious/nonconscious distinction should not be confused with the personal/subpersonal distinction. This is a pernicious error that pervades many introductory chapters in linguistics textbooks (e.g., Larson 2010, p. 12). An explanation of why subpersonal states are never conscious is beyond the scope of this discussion. But, for present purposes, suffice it to say that the reasons for this are *very* different from those that are operative in cases of Freudian repression, which involves emotions and psychological defense mechanisms, and in cases of subliminal perception (e.g., in masked priming experiments), which invite still another type of explanation, having mostly to do with weakness of activation in the primary sensory cortex.

Third, subpersonal states seem to be susceptible to a computational description. By contrast, there are well-known and potentially insurmountable problems—loosely captured under the label "the frame problem"—for the enterprise of giving a computational description of personal-level states and processes. Jerry Fodor puts the point in a characteristically comical way: "I should like to propose a generalization; one which I fondly hope will someday come to be known as 'Fodor's First Law of the Nonexistence of Cognitive Science'. It goes like this: the more global (e.g. the more isotropic) a cognitive process is, the less anybody understands it." (Fodor 1983: p. 107).[17] Although much fruitful work on reasoning has been done since Fodor made this pronouncement, it is not yet possible to say with confidence that rational thought and planning have succumbed to a plausible computational model. By contrast, impressive computational models of language processing have been developed, as we will see in detail in the coming chapters.

Fourth, whereas personal-level states are *paradigmatically* conceptual (or "discursive"), there is an open question as to whether the same is true of subpersonal states. True, MPMs represent linguistic stimuli *as* having one property or another (Chap. 6, Sect. 2.3). In that sense, they *describe* rather than merely *label*, to use the language of Sellars (1957) and Brandom (2009).[18] But Sellars' and Brandom's con-

[17] For extended arguments against the viability of a computational treatment of belief revision, see Putnam (1988) and Brandom (2008: ch. 3).

[18] "Classification as the exercise of reliable differential responsive dispositions (however acquired) is not by itself yet a good candidate for *conceptual* classification, in the basic sense in which apply-

ditions on a state's being conceptual or discursive would certainly rule out subpersonal states. (See the brief discussion of normativity below.)[19] Of course, weaker notions of "concept" are available. Whether *any* such notion is theoretically useful is currently an open question (Machery 2009; Weiskopf 2009). Even if it turns out (as I suspect) that the notion of a concept can be made to do some real explanatory work, it remains unclear that this would yield the result that *subpersonal* representations (e.g., MPMs) as composed of concepts.

Fifth, subpersonal states have no mental attitude. Personal level states have assertoric and desiderative attitudes, which divide in subtle ways: suspicion, prediction, expectation, anticipation, and discovery are all assertoric, whereas wishing, longing, desire, need, and preference are all desiderative. Subpersonal states have nothing like this richness. Psycholinguists' occasional talk of the HSPM having, e.g., frustrations, is amusing but obviously fanciful. Of course, one can dig in one's heels and insist that the mental attitude of subpersonal states is always assertoric. But it's not clear what the point of this maneuver might be. Folk psychology derives a massive payoff from drawing the assertoric/desiderative distinction. What comparable payoff is there in the subpersonal case?

Sixth, subpersonal states like those involved in the early stages of language comprehension are not expressible in speech. They are, of course, *causally involved* in speech production, but that is not sufficient for being *expressed*, let alone conveyed or communicated. For a speech act to *express* an internal state, the two must have similar intentional contents, and the mental attitude of the internal state must correspond to the illocutionary force of the speech act (Rosenthal 2005). But subpersonal states do not seem to have mental attitudes, and their representational properties are arguably not usefully assimilated to intentional content—at least as that technical term is employed in traditional philosophy of mind. And even if subpersonal states had mental attitudes and intentional contents, it's plain that their

ing a concept to something is *describing* it. Why not? Suppose one were given a wand, and told that the light on the handle would go on if and only if what the wand was pointed at had the property of being *grivey.* One might then determine empirically that speakers are grivey but microphones are not, doorknobs are but windowshades are not, cats are and dogs are not, and so on. One is then in a position reliably, perhaps even infallibly, to apply the *label* 'grivey'. Is one also in a position to *describe* things *as* grivey? Ought what one is doing to qualify as applying the *concept* grivey to things? Intuitively, the trouble is that one does not know what one has found out when one has found out that something is grivey, does not know what one is taking it to be when one takes it to be grivey, does not know what one is describing it *as.* The label is, we want to say, uninformative. … What more is required? Wilfrid Sellars gives this succinct, and I believe correct, answer: 'It is only because the expressions in terms of which we describe objects, even such basic expressions as words for the perceptible characteristics of molar objects, locate these objects in a space of implications, that they describe at all, rather than merely label' ["Counterfactuals, Dispositions, and the Causal Modalities," in *Minnesota Studies in the Philosophy of Science,* vol. 2, ed. H. Feigl, M. Scriven, and G. Maxwell (Minneapolis: University of Minnesota Press, 1957), sec. 107.]" (Brandom 2009: pp. 202–3)

[19] This is not to say that subpersonal states have "nonconceptual content," in the sense that is operative in the recent philosophical literature on sensation and perception (e.g. Kelly 2001). For, as many have noted, discussions of *that* topic leave it strikingly unclear what the notion of nonconceptual content amounts to (Bermúdez and Cahen 2010).

attitudes and contents would come apart, often dramatically, from the forces and contents of the speech acts that they facilitate. An MPM that facilitates a production of "Is it raining?" is not itself a question, and it represents 'rain', not rain.

The final two points concern a controversial view according to which representational properties are, in some sense, *inherently normative* (Dennett 1987; Brandom 1994). I am not, myself, confident that we can draw a fact/norm distinction in a way that animates such claims. But suppose that one is fully committed to that sort of view. One must then say something about the *kind* of normativity that's in play, and about the naturalistic *ground* of that normativity. In the case of speech acts and personal-level, conceptual states—the propositional attitudes—the strategy of choice (Kripke 1982; Brandom 1994) seems to be an appeal to community practices of sanctioning various kinds of behavior. But should we credit the idea that community sanctions *also* ground the representational properties of subpersonal states? I think we should not. It is implausible that human communities reward or punish their members for implementing one parsing algorithm rather than another—especially as long as both algorithms preserve, to a similar degree, the socially relevant behavioral effects. If there is normativity to the representational properties of subpersonal states, then it must, I think, be a *biological* kind of normativity, not a social one. The ground of this kind of normativity would have to be something about our evolutionary history (Millikan 1984), or perhaps some notion of fitness. It is much less plausible that it will have anything to do with our "reciprocal relations of authority" to other members of the community (Brandom 2009; Pereplyotchik 2017).

This connects to a final point: The ascription of subpersonal states presupposes only that they have biological purposes, for which they were selected, over the course of evolution or ontogeny. As Dennett makes clear, adopting "the design stance" involves assuming only that a system has a purpose, and that it is not currently malfunctioning. By contrast, the ascription of personal-level states (e.g., thoughts about quarks or chess positions) requires taking "the intentional stance," which rests on weightier assumptions—e.g., that the system is a rational agent, whose terms mostly refer, whose judgments are mostly true, and whose inferences are mostly good. That is why erroneous judgments and inferences leave a person open to rational criticism, whereas misrepresentations in language processing are best seen as malfunctions in a dedicated "subpersonal" mechanism within a person's mind/brain.

We have now reviewed eight reasons for distinguishing personal-level states—e.g., propositional attitudes—from the subpersonal states that include MPMs. I think we have strong grounds for rejecting any view that fails to honor this distinction, or to recognize its importance for cognitive science. One very influential view that arguably falls into this category is due to J. A. Fodor (1975, 1983).

> It is, perhaps, not very important to this Neocartesian story that what is innate should be, strictly speaking, *knowledge*. After all, knowledge is—or so many philosophers tell us—*inter alia* a normative notion, having much to do with the satisfaction of standards of justi-

fication. Chomsky is himself quite prepared to give up the claim that the universal linguistic principles are innately *known* in favor of the explicitly neologistic (hence sanitized) claim that they are innately "cognized." … It is, however, important to the Neocartesian story that what is innately represented should constitute a bona fide object of propositional attitudes; what's innate must be the sort of thing that can be the value of a propositional variable in such schemas as 'x knows (/believes, /cognizes) that P'. … Here is why this is important. As previously remarked, it is the fate of the (presumed) innate information to interact with the child's primary linguistic data, and this interaction is assumed to be *computational*. Now, the notion of computation is intrinsically connected to such semantical concepts as implication, confirmation, and logical consequence. Specifically, a computation is a transformation of representations which respects these sorts of semantic relations. (See Fodor 1975; Haugeland, 1981.) It is, however, a point of definition that such semantic relations hold only among the sorts of things to which propositional content can be ascribed; the sorts of things which can be said to *mean that P*. The idea that what is innate has propositional content is thus part and parcel of a certain view of the ontogeny of mental capacities—viz., that in cognitive development, what is endogenously given is computationally deployed" (J. A. Fodor 1983: pp. 5–6).

In this passage, Fodor moves directly from the claim that the states of the language faculty are *computational* to the very different claim that they are on a par with belief and knowledge, solely on the grounds that both kinds of state have representational properties. This is of a piece with the remarks in Fodor (1975), concerning the irrelevance of the personal/subpersonal distinction for cognitive psychology.

> [W]hile I have argued for a language of thought, what I have really shown is at best that there is a language of computation; for thinking is something that *organisms* do. But the sorts of data processes I have been discussing, though they may well go on in the nervous systems of organisms, are presumably not, in the most direct sense, attributable to the organisms themselves. … But whatever relevance the distinction between states of the organism and states of its nervous system may have for some purposes, there is no particular reason to suppose that it is relevant to the purposes of cognitive psychology. … What cognitive psychologists typically try to do is to characterize the etiology of behavior in terms of a series of transformations of information. … If one has these ends in view, it turns out (again on empirical rather than conceptual grounds) that the ordinary distinction between what the organism does, knows, thinks, and dreams, and what happens to and in its nervous system, does not seem to be frightfully important. … [T]he states of the organism postulated in theories of cognition would not count as states of the organism for purposes of, say, a theory of legal or moral responsibility. But so what? What matters is that they should count as states of the organism for … purposes of constructing psychological theories that are true. (J. A. Fodor 1975: pp. 52–3)

Chomsky (2000), Matthews (2007), and Collins (2008a: ch. 5) expose the unpalatable consequences of this view. I will not pause to review their conclusions here. Instead, I will note that one good reason for honoring the personal/subpersonal distinction in cognitive science is the need to avoid errors concerning the format of various kinds of representation. One should not prejudge the open empirical question of whether MPMs are couched in the same representational format as thoughts about quarks.

7.3.2 Functionalism and Functional-Role Semantics

Functionalism in the philosophy mind is arguably the reigning orthodoxy. But there are, broadly speaking, two versions of that view, and it matters for present purposes which of the two we adopt.[20] One version stems from the work of Sellars and Lewis, who emphasize personal-level states and folk-psychological descriptions, pitched from the intentional stance. The other, which stems from the early work of Putnam and J. A. Fodor, deals with computational states, of the sort that cognitive psychologists posit. The latter brand of functionalism has, I think, led philosophers astray, by blurring the distinction between personal and subpersonal levels of description and making it appear as though the states involved in language processing are propositional attitudes.[21]

One upshot of getting clear on the personal/subpersonal distinction, and the corresponding distinction between the two types of functionalism, is that doing so paves the way for a reconsideration of a common complaint against functional-role theories of representation. Functional-role semantics is sometimes rejected on the grounds that it underwrites an argument for semantic holism, which many see as having implausible consequences with regard to communication, theory change, and other phenomena (Fodor and Lepore 1992). There is, however, a question about whether these arguments against holism about personal-level states carry over to the *sub*personal case. In some cases, it seems that they do not. For instance, issues pertaining to the success of communication among ordinary speakers cannot arise in the subpersonal case, for the simple reason that subpersonal states are never directly expressed, communicated, or conveyed in speech. (If they were, psycholinguistics would be much easier.) Nor do any issues arise about the possibility of interpretation, agreement, or disagreement across divergent or allegedly incommensurable theories, paradigms, or conceptual schemes.

One problematic issue does, however, arise in connection with the holistic characterization of the representational properties of subpersonal states. One might reasonably wonder whether such a characterization would make it impossible for the psychologist to formulate lawlike generalizations about these properties. I do not think that the functional-role view has this consequence. First, subpersonal systems are, in all likelihood, *designed* to be as they are—i.e., they are adaptations, in the sense of Pinker (1994, 1997). It follows that they are species-wide properties, largely similar across speakers, both neurophysiologically and behaviorally. As with any biological system, there will almost certainly be exceptional cases. But the psycholinguist can freely idealize away from these, treating them as cases where *ceteris* is not *paribus*. Given that the laws of psychology are not *strict* laws, this type of idealization is perfectly appropriate. Alongside the point about adaptation, an

[20] The distinction I have in mind is a descendent of the one that Block (1978) drew, between "psychofunctionalism" and "analytic functionalism." I would not, however, accept several of the claims that Block makes in the course of drawing that distinction.

[21] A paradigmatic case of this phenomenon is the introductory chapter in Fodor et al. (1974).

appeal to idealization is, I believe, sufficient to quell concerns about the consequences of adopting holism in the subpersonal case. The predicates of subpersonal psychology would remain projectable and psychological laws would remain true (*ceteris paribus*).

7.4 Computational Psychology, Hardwired Rules, and Embodied Procedural Knowledge

Taking on board the main claims of the foregoing discussion, let us restrict our focus solely to the subpersonal level of description, and treat talk of representation as a mode of functional classification. We are now in a position to be more precise about the kinds of commitments that a psychologically plausible computational model might make about the representations involved in language processing. In particular, we can draw a distinction between two species of psychologically reality—declarative representation and procedural embodiment.

Stabler (1983) defines a variety of claims that can be made within the framework of computational psychology. On his view, a computational system is one that goes into a physical state that represents the output of some function whenever the system is in a state that represents the corresponding input to that function.[22] We say that a system computes a function when this causal pattern is "regular and predictable" enough for it to be "convenient" for us to so describe it; the description is used because it is "clear and useful" (Stabler 1983: fn. 1) Departing slightly from Stabler's terminology, let us call such a function the system's "target function."

In Stabler's terms, giving a *first-level theory* of a computational system involves saying what target function it computes. Such a theory does not, however, tell us *how* the system performs its computations. Seeking a more detailed description, we may, in some cases, formulate a *second-level theory*—a specification of the *program* that the system uses in computing that function. A program consists of more basic functions, whose sequential computation produces the output of the target function.

Some systems—e.g., the modern-day personal computer—compute their target functions by explicitly representing, as a separate data structure in their memory banks, the very programs that they are running. Such systems have "control states" that explicitly represent the *instructions* of a program and are causally involved in the inner workings of the machine. Each such instruction encodes one of the basic functions that, together with other basic functions, conspire to produce the output of

[22] Cf. Chalmers, 1995: "A physical system implements a given computation when there exists a grouping of physical states of the system into state-types and a one-to-one mapping from formal states of the computation to physical state-types, such that formal states related by an abstract state-transition relation are mapped onto physical state-types related by a corresponding causal state-transition function" (p. 392).

the target function. Saying which programs are explicitly *represented* in the system (rather than merely computed by it) is giving a *third-level theory*.

Importantly, not all computers are "stored-program" systems of this kind. Some are hardwired circuits, for which we can give only a second-level, not a third-level theory. Such circuits, Stabler says, "embody" a program, but do not explicitly represent it. The notion of a hardwired or embodied rule is a difficult one, so it will be helpful to consider a number of examples of systems that embody rules without representing them. We can then get a better sense of how to apply Stabler's theoretical framework to the case of language processing.

Horgan and Tienson (1999) recommend that we think of some types of connectionist networks as hardwired circuits, susceptible only to a second-level theory. They write:

> It is sometimes assumed that connectionist representations cannot exhibit syntactic structure, and that lack of syntax therefore constitutes an essential difference between connectionism and classicism (e.g., Churchland 1995). On this view, connectionist models depart from the classical "rules and representations" conception of cognition because they eschew traditional symbolic, syntactically structured representations. But in fact some connectionist models do employ syntactically structured representations and exhibit structure-sensitive processing— although syntactic constituency in these models is not a simple part-whole relation. Examples of such models include Pollack (1990), Smolensky (1990), and Berg (1992). So connectionism need not eschew syntax—and arguably should not (Horgan and Tienson 1999). In connectionist models, rules for processing representations are not explicitly represented. It is sometimes assumed that classicism is committed to explicitly represented rules, and that lack of such rules therefore constitutes an essential difference between classicism and connectionism (e.g., Hatfield 1991). But although programs are explicitly represented as stored "data structures" in the ubiquitous general-purpose computer, stored programs are not an essential feature of the classical point of view. In some computational devices—including, for example, many hand-held calculators—the rules are all hardwired into the system and are not explicitly represented" (p. 725)

Devitt (2006a) picks up on this discussion and claims that the rule for, e.g., addition is not explicitly represented in a simple mechanical calculator.

> Think of a really simple calculator, a mechanical one. When the calculator adds it goes through a mechanical process that is governed by the rules of an algorithm for addition. We have already noted that we cannot assume that the calculator embodies a particular rule simply because the calculator behaves as if it does. The present point concerns the rules that do, as a matter of fact, govern the operations of the calculator. Perhaps these rules operate on representations of numbers like 28 and functions like addition, but the rules themselves are not represented in the calculator. The rules are hardwired but not encoded in the calculator. And, of course, the calculator does not know that the governing rules are rules for addition, subtraction, etc. We can be quite confident about this because the calculator is not the sort of thing that can know about anything. Finally, in virtue of the calculator being governed by those rules we can say, if we like, that the arithmetical "information" that those rules reflect is embodied in the calculator even though the calculator does not represent the information. [fn. 8: Dwyer and Pietroski wonder, surprisingly, how having a mechanism in a speaker that is "correctly described by a linguistic theory is not tantamount to [the speaker's] having the relevant beliefs" (1996: 341). It is not tantamount because the theory could correctly describe the algorithm that is embodied in the mechanism in the way that an arithmetical algorithm is embodied in the calculator.] (2006a: 48)

Finally, consider the account of "implicit knowledge" that one finds in Davies (1987, 1989, 1995). Davies elaborates this key notion as follows: A device that effects transitions between a set of inputs and their respective outputs can be said to have implicit knowledge of a rule just in case the device contains a state or a mechanism that serves as a *common causal factor* in all of those transitions. In a series of clever examples, Davies demonstrates that there are ways of satisfying this definition which are logically weaker than explicit representation (Stabler's "third-level theory") but, at the same time, *stronger* than mere "conformity" with a rule (Stabler's "first-level" theory).

The internal wiring of a simple pocket calculator can provide an intuitive example of such common causal factors. The transitions between '7 + 1' and '8', '8 + 1' and '9', and so on, might all have a common causal explanation—i.e., a single feature of the calculator's internal wiring that mediates each one of these transitions. There may be a mechanism in the circuit devoted solely to computing the function of addition, $f(x, y) = (x + y)$, which is activated when the '+' button is pressed. A component of that mechanism has the task of incrementing by one whatever number serves as the argument to the function, $f(x) = (x + 1)$. Another component increments by 2 whatever number serves as the argument to function, $f(x) = (x + 2)$. And so on. Applying Davies' idea that such common causal factors are cases of implicit knowledge, the whole addition mechanism is revealed as having implicit knowledge of a rule for addition. What Davies calls "implicit knowledge" is a case of what I, following Devitt (2006a), have been calling "embodiment."

Turn now to the psychologist's notion of *procedural* knowledge. Devitt (2006a: Sects. 3.1 and 11.5) asks us to consider a language processing systems like ACT-R, which is described here by Vasishth and Lewis (2006)

> All procedural knowledge is represented as production rules—asymmetric associations specifying conditions and actions. Conditions are patterns to match against buffer contents, and actions are taken on buffer contents. All behavior arises from production rule firing; the order of behavior is not fixed in advance but emerges in response to the dynamically changing contents of the buffers. … In summary, this sentence processing model (a) provides an integrated, quantitative account of both length and structural complexity effects in both ambiguous and unambiguous constructions; (b) predicts reading times in both ambiguous and disambiguating regions; (c) probabilistically predicts both parsing failures and reading times; and (d) provides single-parameter quantitative predictions across multiple experiments and paradigms. (pp. 414–415)

In considering this sort of model, Devitt argues that procedural knowledge does not consist in representational states.

> ACT-R theory, the most recent ACT theory, "accounts for a wide range of cognitive phenomena, including perceptual-motor tasks, memory, and problem-solving" (Johnson et al. 2003: 32). The key question for us is whether this achievement requires that the production rules that are learnt be represented. Descriptions of ACT are often vague on this matter but the received view is that the rules are represented: "in ACT both declarative and procedural knowledge are represented in an explicit, symbolic form (i.e. semantic networks plus productions)" (Sun et al. 2001: 235; see also Masson 1990: 223). Yet … since the production rules constitute procedural not declarative knowledge there seems to be no immediate and pressing need to take them as represented. Because ACT theories are based on general-

> purpose computer models (Anderson 1983: 2) it is perhaps not surprising that the cognitive architecture they propose involves the representation of production rules. Still, we wonder whether we should take this aspect of the model seriously if we are looking for a simulation of skills that exist in real organisms. Is there any reason to think that the IF–THEN rules that become embodied in an organism as a result of practice are represented rather than merely embodied? Perhaps we can suppose that the organism has simply learnt to respond to the working memory representation of a certain condition with the appropriate action. Is there any explanatory gain in supposing further that it does this by representing the rule governing this response and applying it? Anderson himself remarks that "the production is very much like the stimulus-response bond" (Anderson 1983: 6). And Pinker calls it "a knee-jerk reflex" which is "triggered" (1997: 69). (*Ignorance*, p. 215)

It is important to contrast Devitt's notion of procedural embodiment with a related but crucially different notion of procedural representation. Consider, for instance, the following passage from J. A. Fodor.

> Though it gets thrown around a lot in cognitive science, the notion of a procedural representation isn't itself transparent. At its least tendentious, however, a procedural representation is just a representation of a procedure. This is the construal that's suggested by examples like sentence parsing. It may be that what underlies the child's ability to assign syntactic forms to utterances is something like an algorithm for mapping sentence tokens under acoustic representation onto their structural descriptions. A grammar of the child's language may be "explicit" in the parser, or it may be merely "implicit" in the algorithm in the sense that, whereas the latter contains "declarative" representations (like "'the' is a word"), the former contains "imperative" representations (like "if you find a phonological sequence #/t/ /h/ /e/#, label it a token of the word-type 'the'."). Notoriously, parsers and grammars needn't be trivially interconvertible. Going in one direction, there are grammars for which parsers aren't constructible; and, going in the other direction, there needn't be any fact of the matter about which of an infinity of extensionally equivalent grammars a given parser realizes. So, on the construal of "procedural representation" as "representation of a procedure," there really can be something of substance to the distinction between procedural representations and others. (J. A. Fodor 1998: p. 136)

Fodor's distinction is between representations that have, as it were, different mental attitudes.[23] The declarative representation has an assertoric attitude—it purports to give a description of how things are; the procedural representation has an imperative attitude—it tells the system what to do in a particular circumstance. But regardless of these differences, Fodor seems to take both kinds of representation to be stored in a data structure and accessed (i.e., read) by the system—a separate processing step. Devitt's contrast between represented and embodied rules is different. He would consider states with both assertoric and imperative mental attitudes to be representations, on account of their being stored and accessed. What is distinctive about embodied rules, on his view, is precisely that they are *not* stored in a separate data structure and *not* read by the system. Rather, they are *directly implemented*—i.e., hardwired into the architecture of the system, and carried out without further ado.

[23] I urged above that subpersonal states do not have mental attitudes. Thus, respecting the personal/subpersonal distinction requires viewing talk of, e.g., imperative subpersonal representations as mere metaphor. As I noted earlier, Fodor deliberately elides this distinction.

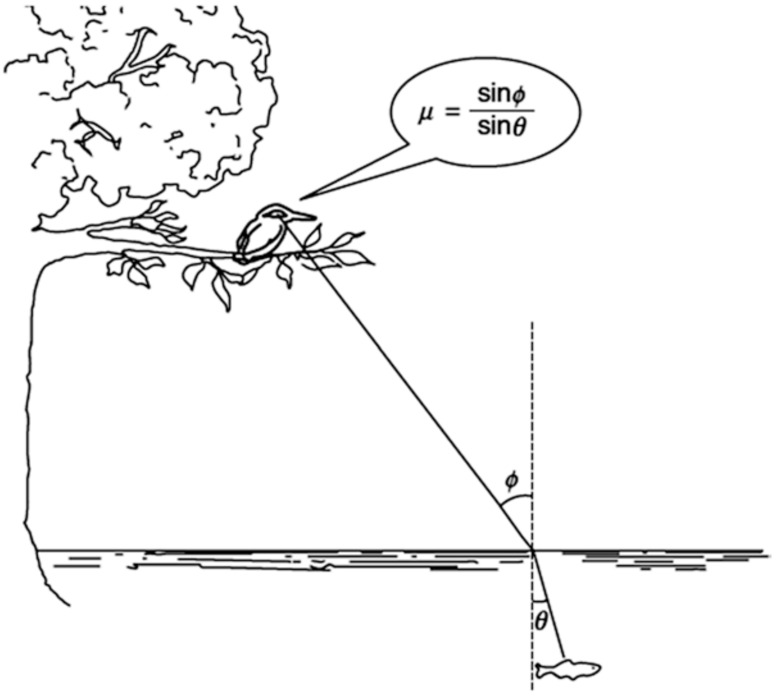


Fig. 7.1 The kingfisher (*Source*: Devitt 2006a), p. 48. (Devitt gets the figure from Boden 1984: p. 153.)

To clarify this contrast further, let us consider Devitt's example of the kingfisher—a bird that feeds on fish by diving at an angle and reliably catching its prey without pursuing it underwater. In order to do this by sight alone, the kingfisher has to compensate for the way that light refracts as it exits the water and enters air. Otherwise, the image of the fish would be systematically misleading as to its actual location (Fig. 7.1).

The question arises: How does the kingfisher accomplish this feat? Devitt writes:

> [I]t would be absurd to suggest that the kingfisher goes through the following process: it represents (tacitly knows) that μ for water-to-air refraction is such and such; it represents that the angle ϕ is so and so; it represents that $\mu = \sin \phi / \sin \theta$; it uses this information to calculate the angle θ. … Those mechanisms, reflecting information about water-to-air refraction, are simply built into the kingfisher "by nature" just as the rules for the calculator are simply built into it by an engineer" (2006a: p. 49)

Now, if we hold fast to the distinction between personal and subpersonal levels of description, or (what comes to the same thing) if we make it clear that we are describing the kingfisher from the design stance, not from the intentional stance, then it is perhaps not absurd to suggest that some subsystem within the kingfisher constructs perceptual representations of angles. But there is no obvious reason for holding that this subsystem also represents, in a declarative fashion, the value of μ or the equation describing water-to-air refraction ($\mu = \sin \phi / \sin \theta$). That is, there are no grounds for claiming that it draws on such a representation as a premise in its computations.

A more parsimonious way to characterize the kingfisher's internal processes is to say that a subsystem within the kingfisher embodies a certain procedure. The procedure is to token a representation of the appropriate value for the unperceived angle θ, as a direct consequence of tokening a perceptual representation of the value of ϕ. Such a procedure would be, as Devitt says, built into the very structure of kingfisher's visuomotor system, rather than being represented in a declarative fashion in a separate data structure.

Of course, we can be more parsimonious still. Instead of positing representations of perceived and unperceived angles within the kingfisher, we could posit a direct reflex-like sensory-motor link. We would thus claim the system transforms activations in the visual cortex, occasioned by a refracted angle ϕ (within some range of magnitudes), into motor-neuron activations that directly control the diving behavior that is appropriate to angle θ. If the only behavior that we were concerned to explain were the kingfisher's diving for fish, then this sort of account—which posits no representations whatsoever—would be suitable. We would thus say that the kingfisher behaves *as if* it stores representations, though in actuality it does not. (One is reminded here of the hydraulically-driven mechanical statues that once stood in the royal French gardens—the early automata that so fascinated Descartes.)

Note that this would constitute a major disanalogy between the kingfisher's diving capacity and the case of human sentence processing. I have argued that the HSPM constructs mental phrase markers in order to keep track of the hierarchical syntactic relations that hold between stimuli dispersed over discontinuous regions of time. If it were not for this aspect of human sentence processing, we would be able to view the states of the HSPM as mere responses to linguistic stimuli, not as genuine representations. In the case of the kingfisher, however, it is possible that this crucial element of the story is missing. It may be that the kingfisher's motor outputs are directly responsive, in a reflex-like manner, to its immediate sensory inputs, rather than being sensitive to a rich set of contextual factors that extend arbitrarily far back in time. In this case, we can view its internal states as mere responses, rather than genuine representations. Whether this is correct is, of course, an empirical issue that we need not settle here.

Devitt discusses a slightly different case, involving the visuomotor skills of humans. Experiments show that, when running to catch a fly ball,

> skilled fielders "ran at a speed that kept the acceleration of the tangent of the angle of elevation of gaze to the ball at 0" (McLeod and Dienes 1996: 531). It is not known how they manage this (p. 542) but they surely don't manage it by representing the algorithm for being in the right place and applying it to a series of representations of the acceleration of the tangent of the angle. (2006: p. 50)

Surely, it is correct to say that the *fielders* don't represent the algorithm. Should we say that a subpersonal system inside of them does? Devitt suggests that the algorithm for computing the target position is embodied in them, not represented. This seems to be the most parsimonious explanation of the data. Still, a question remains, parallel to the two raised above regarding mental phrase markers in parsing and representations of angles in the kingfisher: Does the fielder—or any subsystem of her mind/brain—represent “the acceleration of the tangent of the angle of elevation of gaze to the ball,” or any of the mathematical determinants of that complex value? Again, our verdict will depend partly on the anatomical facts regarding whether there is a direct neural connection between the dorsal horn of the visual stream and the motor areas that are responsible for coordinating the fielders’ direction of motion and gaze. It also partly depends on whether the mechanisms involved in this task are routinely recruited by a variety of other seemingly disparate tasks, and whether these mechanisms are relatively flexible in their operation. These are empirical issues that are beyond the scope of our present concerns.

To round off this discussion, let us consider the phenomena that psychologists call implicit learning and implicit memory. The issues surrounding these notions are, unfortunately, somewhat murky, in part because psychologists are not in agreement about how to draw the relevant distinctions. As Devitt points out:

> our view of the nature of skills depends on our view of the nature of procedural knowledge and implicit memory. Here there are lots of interesting ideas but no consensus. Indeed, our knowledge of this matter is at an early stage: “a great deal remains to be learned about the cognitive and neural mechanisms of implicit memory” (Schacter 1999: 395); … “there is no consensus regarding the details of the dichotomies” (Sun 2003: 698). … Axel Cleeremans (2003) refers to “as many as eleven different definitions… [of] implicit learning” (p. 491).

Davies (1995) examines the same issue and argues that the psychologist’s implicit/explicit distinction may not be helpful in addressing the general topic of what he calls implicit knowledge (which, again, seems to map neatly onto what we have called embodied rules).

> [T]he distinction in experimental psychology between explicit and implicit memory tests does not help us. In an implicit memory test “memory for a recent experience is inferred from facilitations of performance, generally known as repetition or direct priming effects, that need not and frequently do not involve any conscious recollection of the prior experience” (Schacter 1989, p. 695). In contrast, explicit memory tests “make explicit reference to and demand conscious recollection of a specific previous experience.” (ibid.) Given that we are considering human information processing much of which is unconscious, and that we also want our notions of knowledge of rules to be applicable to systems—such as small connectionist networks—for which the question of conscious recollection does not even arise, this … usage of “explicit” is of no help to us…(pp. 158–9)

Davies’ remarks in this passage point to a problem in the literature on implicit and explicit memory and, incidentally, also in discussions of the declarative/procedural distinction. Theorists sometimes conflate both of those distinctions with the conscious/nonconscious distinction and fail to take note of the personal/subpersonal distinction.[24] The moral that I draw from such considerations is that making prog-

[24] Focusing for the moment on the declarative/procedural distinction, rather than the (possibly orthogonal) implicit/explicit distinction, we can compare two examples of declarative memory representations: (1) the factual knowledge that quarks are particles—clearly a personal-level state, which is sometimes conscious (especially when expressed in speech), and (2) lexical memory—e.g., information about the subcategorization frames of verbs—a *subpersonal* state and hence *never* conscious. It appears that some declarative knowledge is accessible to consciousness, but some isn’t.

ress on the psychological reality issue requires being as clear as possible about the relationships between five different distinctions: conscious/nonconscious, personal/subpersonal, implicit/explicit, declarative/procedural, and occurrent/dispositional. I have attempted to put this moral into practice in this chapter. Putting aside the difficult issues pertaining to implicit and explicit memory, my hope is that the examples discussed above give us a tolerably firm grip on the notion of an *embodied rule*. In the next section, I offer an analogy that will shed further light on this notion by bringing into play the last of the five distinctions mentioned above.

7.5 Grammar as a Set of Subpersonal Embodied Procedural Dispositions

MPMs are the subpersonal analogous of perceptual judgments, in that their function is to represent the syntactic properties of incoming linguistic stimuli. These are relatively transient states. Though priming data suggest that they remain active for a short time after coming into existence, their "lifespan" is typically measured in milliseconds (Pickering and Ferreira 2008). Consider now a *different* theoretical construct—what we might call *mental syntactic principles* (MSPs). Suppose that these are standing structures—architectural features of the human parsing mechanism—that *embody* the rules or principles of a person's language. When the language faculty is inactive—i.e., when no comprehension or production processes are taking place—these structures are, so to speak, *dormant*. In the jargon of contemporary metaphysics, we can say that they are *dispositional*, rather than occurrent.

One way of getting a grip on this distinction between MPMs and MSPs is by focusing on an analogous distinction between perceptual judgments and beliefs. Philosophers sometimes apply the term 'belief' to *any* assertoric propositional attitude, regardless of what other properties it has. This usage elides an important distinction between *occurrent* assertoric attitudes, on the one hand, and *dispositions to have* occurrent assertoric attitudes, on the other. The dispositional/occurrent distinction is not only significant for the purposes of theory construction, but is also clearly marked in folk psychology, and frequently makes a difference to everyday practices of prediction and explanation.[25] Thus, on the more careful usage that I recommend here, beliefs are dispositions to have occurrent thoughts or judgments when the occasion arises, or to process various occurrent thoughts and judgments in specific ways.[26] By contrast, thoughts and judgments are states that we come to be in

[25] Q: "Why did the student get this question wrong on the math exam?" A: "He must have thought that the square root of 4 can only be a positive integer." Q: "But if you had asked him straight out, he would have told you that it can be negative too." A: "Yes, I'm sure he *believes* that it can, but he must have not *thought* of that at the time."

[26] Notice that I am *not* proposing that belief is a disposition to express oneself in speech or in nonverbal behavior. There is a difference between being disposed to think something and being disposed to say it or to act on it. Beliefs are dispositions to have thoughts—i.e., occurrent psychological

occurently, either by way of perceptual processing or in the manifestation of dispositional beliefs in inference. Similar remarks go for non-assertoric states. There is a distinction between standing preferences, on the one hand, and occurrent *wants* or *urges*, on the other.[27]

Some philosophers assume that the occurrent/dispositional distinction lines up with either the conscious/nonconscious distinction or with the personal/subpersonal distinction. Neither of these assumptions has any merit. The folk-psychological practices of predicting and explaining behavior rest on the ascription of both thoughts and beliefs to *whole persons*, not to subpersonal mechanisms.[28] Moreover, intentional states can occur both consciously and nonconsciously. Intentions, thoughts, expectations, hopes, worries, fears, desires, volitions, and decisions, as well as many other more "exotic" propositional attitudes have all been shown in experimental settings to occur without a person's awareness *of* their occurrence (Rosenthal 2005). Some of these are occurrent states and others are dispositional. We are not aware, for instance, of all of our beliefs and memories at any given time, though we are able to answer questions like "Have you ever ridden an antelope?" without skipping a beat (Dennett 1978). When the question is presented, a standing belief gives rise to an occurrent thought, which is then expressed in speech. If the question is presented in a clever masked-priming study or in a variety of other (ordinary and not-so-ordinary) conditions, the occurrent thought may still be formed, though nonconsciously. Being an occurrent state, then, is neither a necessary nor a sufficient condition for being a conscious state. And being a dispositional state is neither necessary nor sufficient for being a nonconscious state.[29]

Returning to the case of language processing, we can now put our analogy to use. If mental syntactic principles (MSPs) are embodied, rather than explicitly represented, they can be thought of as a subpersonal analogue of standing beliefs. And, continuing the analogy, mental phrase markers (MPMs) can be seen as a subpersonal analogue of perceptual thoughts or judgments. To illustrate the use and plau-

states. I do, however, hold the view that when one has an occurrent thought, one is thereby disposed—to whatever extent—to express that thought in speech and nonverbal behavior. This disposition is typically overridden by a number of factors, including a rudimentary sense of tact, etiquette, or moral decency.

[27] Q: "Do you prefer chocolate ice-cream or vanilla?" A: "I prefer chocolate, but I don't want any right now because I just ate two portions of cake and I have no urge to eat any more dessert."

[28] And even if choose to extend our use of the term 'belief' to cases of subpersonal information processing, we can just as easily extend our use of 'thought' and 'judgment' in that way.

[29] Searle (1990) argues that only occurrent conscious states can have what he calls "aspectual shape"—a property whose ascription gives rise to an intensional-with-an-'s' context. Searle thinks of aspectual shape as that property of mental states in virtue of which intensional differences make a difference *to the person who is in the state*. But, as Rosenthal (1990) points out, there is no reason at all to believe that only conscious states can make a difference to a person. A belief to the effect that water is water is a disposition to have a rather boring thought; a belief to the effect that water is H_2O is a disposition to have a marginally interesting one. The latter belief can give rise to a nonconscious thought whose function in a person's inferential economy is different from that of thought that water is water. This is a difference that makes a difference *to the person*.

sibility of this analogy, let us run through two psychological explanations—one pitched at the personal level and another at the subpersonal level.

At the personal level—from the "intentional stance", so to speak—we tell stories such as the following: "Ever since he was a kid, Jeff has believed that spiders are dangerous. So when he just now saw several spiders crawl out of a box in his attic, he reasoned that there are probably more dangerous bugs in the box." In this story, the perceptual judgment regarding the presence of spiders interacts with a standing belief to yield an occurrent thought as a conclusion.[30]

The process of syntactic parsing, though it takes place subpersonally, may have roughly the same structure. To illustrate, consider what happens when the HSPM encounters (2).

(2) Have the sick soldiers….

The HSPM has two standing belief-like states that embody two principles: (i) a sentence-initial auxiliary verb serves to introduce a question; (ii) a subsequent noun phrase will be the subject of the clause. As a result of encountering the first word of (2), it constructs a mental phrase marker that represents 'Have' as an inverted auxiliary verb. From this MPM and the two standing MSPs, the HSPM "concludes" that that the subsequent noun phrase will be the subject of the clause.

Given this analogy between the belief/judgment contrast at the personal level and the MSP/MPM contrast at the subpersonal level, we can address our concerns about the psychological reality of MSPs by asking the following question: Are beliefs, construed in the manner above, psychologically real? That is, can we attribute a psychological reality to a "mere procedural disposition"? Can we do so simply on the strength of the fact that it's a disposition to give rise to an occurrent representation?

Let us imagine a scenario where we learn that there is nothing more to having the belief that red strawberries are ripe than having the disposition to infer 'This is ripe' *directly* from 'This is a red strawberry', 'It's red' *directly* from 'It's a ripe strawberry', and so forth. I think we would still grant that the standing state underlying this cluster of dispositions is psychologically real. The "procedural" nature of this state does not hamper our ability to explain and predict a person's behavior by positing it. And given that we have, as yet, no way of identifying the state in neurophysiological terms, it seems that the predictive and explanatory leverage that we derive from classifying it in intentional or semantic terms is enormous. Moreover, when we do eventually look at the neurophysiological details, we expect—in accordance with the view of embodiment that I endorsed above—to find that there is a common causal factor in the brain that accounts for the aforementioned inferences. For these

[30] We can imagine two accounts of how a perceptual judgment interacts with a belief. The first account has it that the interaction is direct and unmediated. The second has it that the belief gives rise to an occurrent thought with the same intentional content, which, in turn, interacts directly with the perceptual judgment. Which of these accounts one prefers will depend on the details of one's theory of cognitive architecture.

reasons, I see no compelling grounds for resisting the ascription of psychological reality to the dispositional states of belief. Similar considerations can be adduced for the case of standing preferences.

These remarks carry over, I suggest, to the subpersonal case. Mental syntactic principles may well turn out to be nothing more than embodied procedural knowledge—a systematic disposition to move directly from one type of mental phrase marker to another. This would not, however, impugn their explanatory value as "psychologically real" theoretical posits.

7.6 Summary and Conclusion

In Sect. 7.2, I discussed several competing views of representation. The major lines of division, I suggested, pertain to the role of internal processing factors (computational and inferential role) and external environmental factors (causal and nomological mind-world relations). On a fully externalist view, like the one championed by Fodor (1990), internal factors do not determine the representational properties of psychological states. At the opposite end of the spectrum, we find fully internalist views like those of Chomsky (2000), which deny that causal and nomological relations to environment have any bearing on the representational properties of internal states. Between these two poles, a variety of views have been developed and defended. Of these mixed views, the one that appears most promising is functional-role semantics, as developed by Sellars, Harman, Brandom, and others. I went on to explore the core tenets of this view and to contrast it with other positions in logical space. Finally, I suggested that there is no overwhelming reason to adopt either a full internalism or a full externalism.

Section 7.3 surveyed a number of important differences between personal and subpersonal levels of description. Unlike personal-level states, subpersonal states are not inferentially integrated, never conscious, never expressed in speech, lack mental attitude, are not composed of concepts, and do not resist a computational treatment. I went on to relate the personal/subpersonal distinction to two others—viz., the distinction between Dennett's intentional and design stances, and the distinction between the sort of functionalism propounded by David Lewis and the sort that has been popularized by Jerry Fodor. I suggested that the latter sort of functionalism has given rise to a conflation between subpersonal and personal-level states. One upshot of this conflation is that a set of arguments commonly invoked against functional-role semantics is often not evaluated in light of the personal/subpersonal distinction. At the close of Sect. 7.3, I provided reasons for thinking that these arguments, whatever their force in the personal-level case, do not militate against a functional-role treatment of the representational properties of subpersonal states.

For the purposes of the present project, a crucial distinction must be drawn between two ways in which a set of rules or principles might be psychologically real. In Sect. 7.4, I sketched the theoretical framework of Stabler (1983), who distinguishes between computational systems in which rules are declaratively represented and ones in which those very rules are "hardwired" or embodied. I surveyed several examples of the latter type of system, in an effort to clarify the represented/embodied distinction, and to differentiate it from the distinction that Fodor (1998) draws between declarative and procedural representations. In the course of that discussion, I also appealed to a view developed by Davies (1987, 1989, 1995), according to which a rule is embodied in a system just in case there is a common causal factor within the system that mediates between the rule's domain and its range.

With this distinction between embodied and represented rules in place, I went on in Sect. 7.5 to discuss the distinction between occurrent and dispositional propositional attitudes—e.g., datable judgments and urges vs. standing beliefs and preferences. I noted that we might find an analogous distinction, at the subpersonal level, between mental phrase markers and mental syntactic principles. I went on to ask what would follow from a discovery that beliefs are nothing more than the reliable inferential dispositions to move procedurally from one occurrent thought or judgment to another (e.g., from "Lightning is seen now" to, e.g., "Thunder will be heard soon,"). I suggested that it would not follow from this that beliefs lack psychologically reality. We would continue to think of beliefs as useful, explanatory posits that illuminate the contents of a creature's mind and, consequently, figure in our accounts of its outward behavior. By analogy, I think that we should be willing to ascribe psychological reality to those states of the HSPM that embody syntactic rules or principles, rather than explicitly representing them. Doing so underwrites the legitimacy of psycholinguistic explanations, of the kind that we surveyed in Chap. 5.

What remains to be seen is which of the positions mentioned at the outset of this chapter derive support from the computational parsing models that have some claim to psychological plausibility. In the following chapters, we will conduct a historical survey of such models, in the hope of determining whether the rules or principles of syntax are represented or embodied in the HSPM.

Chapter 8
Computational Models and Psychological Reality

Abstract The main claim of this chapter and the next is that all psychologically plausible parsing models either represent or embody a grammar. I substantiate this claim by surveying top-down, bottom-up, and left-corner parsing algorithms, illustrating the ways in which they can draw on explicit representations of grammatical principles. I then discuss the Parsing as Deduction approach, wherein a proof procedure takes the rules of a grammar as axioms and derives MPMs as theorems, using a subpersonal analogue of natural deduction. This constitutes the most concrete implementation of the idea that the HSPM draws on syntactic principles as data. Finally, I turn to three strategies for dealing with the massive structural ambiguity that any parser will encounter in the input stream. Resource-based approaches emphasize parsing heuristics that minimize the use of computational resources, like short-term memory. Frequency-based approaches use statistical analyses of corpuses and treebanks to guide parsing decisions. Grammar-based approaches appeal directly to Minimalist syntactic principles in accounting for the HSPM's behavior in the face of ambiguity. The latter possibility is particularly exciting, as it would show that a Minimalist grammar is not only suitable for describing abstract formal relations, but also the real-time operation of psychological mechanisms.

Keywords Top-down parsing • Bottom-up parsing • Left-corner parsing • Parsing as deduction • Structural ambiguity • Ambiguity resolution • Computational resources • Computational efficiency • Statistical analysis • Statistical parsing (see also: frequency-based parsing) • Corpus • Treebank • Structure rules vs. processing rules • Strong competence hypothesis • Oracle • The Earley algorithm • The CYK algorithm • Minimal attachment • Late closure • The Minimal Chain principle • Context-free grammar (CFG) • Probabilistic context-free grammar (PCFG) • Lexical vs. grammatical rules • Chart parsing • Mitchell P. Marcus • Shift-reduce parsing • Right- vs. left-branching • Center embedding • Definite clause grammar (DCG) • *n*-gram models • Beam search • A*-search • Serial vs. parallel parsers

D. Pereplyotchik, *Psychosyntax*, Philosophical Studies Series 129,
DOI 10.1007/978-3-319-60066-6_8

8.1 Introduction

In this chapter, we begin our survey of computational parsing models. Laying out some of the details of these models puts us in a position to determine to what extent they support one or another construal of the "psychological reality" of grammar. Here, again, are the available positions on this issue:

REP-GRAM-PROC The structure rules of a language—i.e., the syntactic rules or principles that comprise its grammar—are (i) identical with the processing rules that govern the HSPM, and (ii) are *explicitly represented* in the competent speaker's mind/brain.

REP-GRAM-DATA The structure rules of a language (the grammar) are (i) distinct from the processing rules of the HSPM, but (ii) are *explicitly represented* in the competent speaker's mind/brain and used *as data* by the processing rules.

EMB-GRAM-PROC The structure rules of a language (the grammar) are (i) also the processing rules of the HSPM, but (ii) they are *embodied,* not explicitly represented, in the competent speaker's mind/brain.

GRAM-CONFORM Human language processing *conforms* to a grammar, in the sense that the competent speaker's mind/brain reliably takes sentences that are licensed by the grammar as inputs and produces such sentences as outputs.

The logic of the situation is this: **GRAM-CONFORM** is entirely uncontroversial, saying nothing more than that our minds are capable of processing sentences of some natural language. The other three positions make stronger commitments to the role of the grammar in processing. **REP-GRAM-PROC** and **REP-GRAM-DATA** are both committed to RT, the idea that the rules or principles of a grammar are mentally represented, whereas **EMB-GRAM-PROC** is not.

In this chapter and the next, I argue that comprehension models that support **REP-GRAM-PROC** and **REP-GRAM-DATA** always face more parsimonious rivals, in which the grammar is not represented but embodied. Suppose that's right. This would leave us with **EMB-GRAM-PROC** and **GRAM-CONFORM**. The latter is mandatory, but makes no commitment concerning the psychological reality of any grammar. It is compatible with *any* account of language processing, including one that makes no use of the grammar at all. A question arises, then, about what grounds we have for adopting any more committal position. How do we know that a grammar is even *embodied* in the mind/brain? The short answer, to be developed in the remainder of this book, is this: The HSPM *must* construct mental phrase markers, and *all psychologically plausible models of* that *process either represent or embody a grammar*. The only alternative is sheer magic. In what follows, I substantiate this claim by surveying models of comprehension that employ a range of grammars and have some claim to psychological plausibility.

It will be helpful to begin with a general "anatomy" of language processors. Steedman (2000) lays out the basic components of any such system.

> All language-processors can be viewed as made up of three elements. The first is a grammar, which defines how constituents combine to yield other constituents. The second is an algorithm for applying the rules of the grammar to a string. The third is an oracle, or mechanism for resolving nondeterminism. The oracle decides which rule of grammar to apply at points in the analysis where the nondeterministic algorithm allows more than one rule to apply. [fn. 1: The division of processing labor between a nondeterministic algorithm and an oracle is not always made explicit, particularly in implementations. However, all processors can be viewed in this way.] (pp. 226, 280)

What Steedman calls 'the grammar' is a set of *structure rules*, which govern the language that constitutes a parser's domain.[1] Talk of "applying rules" reflects Steedman's commitment to what Bresnan and Kaplan (1982) called the "Strong Competence Hypothesis," which Steedman formulates as follows: "the grammar that is used by *or implicit* in the human sentence processor is the competence grammar itself" (226, emphasis added).[2] Plainly, a great deal hangs on what is meant by a grammar's being "used by" and "implicit in" the HSPM.

In this chapter, we follow Steedman's partitioning of the elements of the parser and provide examples of grammars, algorithms, and oracles. (In a later section, we will explore a fourth element of some contemporary parsing systems—what theorists in the statistical parsing tradition call "the decoder".) We begin by discussing bare-bones context-free phrase structure grammars (Sect. 8.2). Though these are now known to be too weak to capture important fragments of natural language, we will use them here to introduce several foundational parsing algorithms, which can be used with more powerful formalisms, including those discussed in the next chapter: transformational grammars, ATN grammars, the Government and Binding (GB) Theory, and Minimalist grammars.[3] The top-down, bottom-up, and left-corner parsing algorithms discussed in this chapter form the foundation of the majority of parsing models.

---

[1] Devitt (2006a) characterizes the notion of a structure rule as follows: "The outputs of a linguistic competence … are governed by a system of rules, just like the outputs of the chess player, the logic machine, and the bee. Something counts as a sentence only if it has a place in the linguistic structure defined by these structure rules. Something counts as a particular sentence, has its particular syntactic structure, in virtue of the particular structure rules that govern it, in virtue of its particular place in the linguistic structure. Like the theory of the idealized outputs of the chess player, logic machine, and bee, our theory can be used to make distinctions among the nonideal. Strings that are not sentences can differ in their degree of failure. For they can differ in the sort and number of linguistic structure rules that they fail to satisfy" (p. 24).

[2] Steedman (2000) actually makes a stronger commitment: "It is important to note that the strong competence hypothesis as stated by Bresnan and Kaplan imposes no further constraint on the processor. In particular, it does not limit the structures built by the processor to fully instantiated constituents. However, the Strict Competence Hypothesis proposed in this book imposes this stronger condition" (228).

[3] Constraints of space preclude a fuller discussion of a number of formalisms that are currently popular in parsing theory. These include the lexicalist grammars that rely on feature unification—e.g., Lexical Functional Grammar (Bresnan 2001) and Head-driven Phrase Structure Grammar (Pollard and Sag 1994)—as well as Tree-Adjoining Grammar (Schabes et al. 1988) and Combinatory Categorial Grammars (Steedman 2000). Throughout the discussion, I will occasionally mention these, but I reserve a detailed treatment for future work. The philosophical conclusions pertaining to the psychological reality issue are not affected by this omission.

Section 8.3 provides an overview of the Earley and CYK algorithms, and illustrates the ways in which they can draw on explicit representations of grammatical rules.[4]

Section 8.3 also includes a discussion of an exciting development in parsing theory, known as the Parsing as Deduction (PAD) approach. Here, syntactic analysis is seen, quite literally, as a species of *natural deduction*, in a variant of first-order logic. Parsing models that implement this approach run through an explicit proof procedure that takes the rules of a grammar as axioms and derives theorems concerning input strings. The PAD framework has been shown to subsume the Earley and CYK algorithms and to be applicable to a wide array of syntactic formalisms. It also constitutes the most concrete implementation of the core idea of **REP-GRAM-DATA** —viz., that structure-rules are drawn on *as data* in the course of comprehension. Models that adopt the PAD approach treat rules as *truth-evaluable claims*, which the parser uses as *premises* in the course of its deductive procedures. The success of such models substantiates my claim in Chap. 6 (Sect. 6.3.4) that the HSPM can exhibit a subpersonal analogue of inferential coherence.

In Sect. 8.4, we discuss the third element of Steedman's anatomy—the oracle. Here, we take up pressing questions about the computational efficiency of the Earley and CYK algorithms, and explore various efforts to tackle the daunting problem of structural ambiguity in the input. These fall into three broad categories. The resource-based approach emphasizes parsing heuristics such as Minimal Attachment, Late Closure, and the Minimal Chain Principle, on the grounds that these minimize the parser's use of limited computational resources, such as short-term memory. The frequency-based approach makes use of statistical analyses of a corpus or treebank to guide parsing decisions. Statistical parsing is arguably the reigning paradigm in computational linguistics, where probabilistic extensions of context-free grammars (and other formalisms) are being developed and refined at a rapid pace, with impressive gains in both efficiency and psychological plausibility. Finally, there is the grammar-based approach, which attempts to leverage the resources of an independently-motivated Minimalist grammar in accounting for the HSPMs behavior in the face of ambiguity. The success of this approach would constitute a powerful argument for **REP-GRAM-PROC**.

8.2 The Grammar, First Pass: Context-Free Phrase Structure Grammars

Context-free phrase structure rules, sometimes called "production" or "rewrite" rules, constitute the starting point of many contemporary syntactic theories. Figure 8.1 contains an impoverished context-free grammar (CFG) for a fragment of English. Like all CFGs, this grammar divides its rules into two types: lexical and grammatical. The lexical rules make claims about the syntactic categories to which the words of a language belong. They do so by mapping *nonterminal* symbols like

[4]This claim pertains to the implementations of such algorithms in conventional computers. How such algorithms are implemented in the human brain, if indeed they are, is a separate question. As noted above, it may well be that the brain *embodies* the rules, without explicitly representing them.

Lexical rules

Noun → *flight* | *breeze* | *trip* | *morning*
Verb → *is* | *prefer* | *like* | *need* | *want* | *fly* | *leave*
Adjective → *cheapest* | *non-stop* | *first* | *latest* | *other* | *direct*
Pronoun → *me* | *I* | *you* | *it*
Proper-Noun → *Alaska* | *Baltimore* | *Los Angeles* | *Chicago* | *United* | *American*
Determiner → *the* | *a* | *an* | *this* | *these* | *that*
Preposition → *from* | *to* | *on* | *near*
Conjunction → *and* | *or* | *but*
Complementizer → *that*

Grammatical rules	**Quasi-English Examples**
S → *NP VP*	I + want a morning flight
S → *VP*	Leave me in Chicago.
NP → *Pronoun*	I
NP → *Proper-Noun*	Los Angeles
NP → *Det Nominal*	a + flight
Nominal → *Nominal Noun*	morning + flight
Nominal → *Noun*	flight
VP → *Verb*	do
VP → *Verb NP*	want + a flight
VP → *Verb NP PP*	leave + Boston + in the morning
VP → *Verb PP*	leave + on Thursday
VP → *Verb Complementizer S*	prefer + that + the flight leave this morning

Fig. 8.1 A toy context-free grammar. In the lexical rules, the symbol '|' expresses exclusive disjunction (Adapted from Jurafsky and Martin 2008: p. 394)

'*Noun*' and '*Verb*', which denote those categories (sometimes called "parts of speech") into *terminal* nodes like 'morning' and '-ed'—i.e., actual words or morphemes. The lexical rules, taken together, can be regarded as the lexicon of the language that the grammar is used to describe.

The lexicon presented in Fig. 8.1 is rather poor, describing only the basic syntactic category of any given word. No description of the words' phonological, morphological, or semantic properties is provided. Similarly, this lexicon does not specify the more fine-grained syntactic properties of the words—e.g., the verbs' subcategorization frames or the nouns' agreement features. Nor does it provide any statistics about the frequency with which the words are used, either alone or in combination. In later sections, we will explore the prospect of adding these features to the grammar—particularly frequency information, which yields a *Probabilistic* Context-Free Grammar (Sect. 8.4.2).

Unlike the lexical rules, the grammatical rules of a CFG are used to make lawlike claims about the relationships *between* syntactic categories. For instance, the rule '*VP* → *Verb NP'* amounts to the modal claim that if something *were* to consist of a standalone verb followed by a noun phrase, then it *would* thereby be a verb phrase.[5]

[5] We can trade in the awkward subjunctive locutions for ordinary material conditionals, so long as we keep firmly in mind that the latter are *lawlike*, in the sense of Goodman (1954/1983). See Chaps. 1 and 2 for a discussion of why this matters for the ontology of linguistic theory.

The grammar is recursive, hence capable of describing an infinite number of sentences (if there happen to be that many; see Chap. 1). The recursive nature of the grammar is due to the following two rules:

S → *NP VP*
VP → *Verb Complementizer S*

Jointly, these two rules entail that the language contains structures in which a sentence is embedded within another sentence, as in the following examples:

(1) I prefer that the flight leave this morning.
(2) I prefer that I prefer that the flight leave this morning.
(3) I prefer that I prefer that I prefer that the flight leave this morning. etc.

The '→' symbol must be treated with care. A common practice is to read a rule like '*S* → *NP VP*' as "S *goes to* NP VP." But the import of 'goes to' is rather opaque. And the same is true of the popular "rewrite" and "produce" locutions. The claims that "S *can be rewritten as* NP VP" or that "S *produces* NP VP" are equally obscure. (What modality attaches to 'can'? Who is doing the rewriting or the "producing"? What does it even mean to "produce" here?) For some purposes, of course, this nonspecificity is desirable; and for many purposes it's either harmless or irrelevant. But for *our* purposes such formulations are dangerously sloppy, for they slur over the distinction between structure rules and processing rules.[6] Let us, then, be pedantically explicit: If we wish to treat the above statements as structure rules of English, then we should read '*S* → *NP VP*' as claiming that something *would be* a sentence *if it were* to consist of a concatenation between a noun phrase and a verb phrase. The same goes, *mutatis mutandis*, for all of the grammatical rules listed above.

Context-free rewrite rules impose a binary distinction between the strings that are and the strings that are not sentences of some language.[7] The question arises: Which language? If we take a language to be a set of sentences, there are several options. Consider first the language that consists of precisely those strings that are sentences according to the claims of the grammar. Trivially, the grammar is true of *this* language. But now contrast this with your own idiolect—i.e., the variant of English that you currently speak (assuming, for the moment, that there really is just one; see Chap. 3). The grammar listed above will not be true of *it*, so long as you would, in some relatively ordinary circumstance, express yourself by uttering the string in (4) and so long as you are strongly disposed to *not* express yourself by uttering the string in (5).

(4) Linguistics is difficult.
(5) Fly morning on breeze.

[6] See Kartunnen and Zwicky (1985: pp. 3–5) for a discussion of closely related issues.

[7] Strictly speaking, in order to secure this result, we would have to include an explicit rule to the effect that *nothing else* is a sentence of the language. Following convention, I have omitted this in the main text.

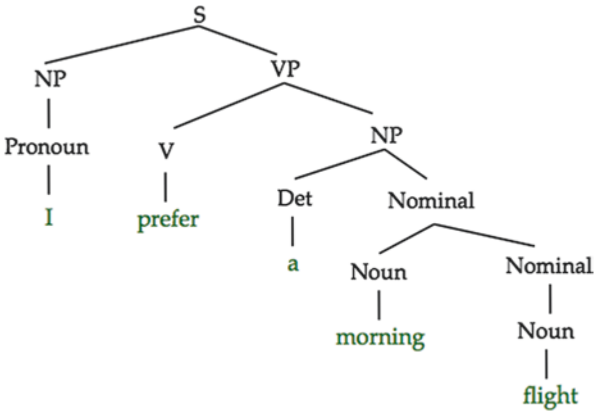
The reason that the grammar presented above is not true of *your* idiolect is that it both undergenerates and overgenerates. (4) uses lexical items that do not appear in that grammar, while the aberrant (5) *can* be generated by the rules of that grammar. Similar remarks can be made for a public *dialect* of English (Chap. 3).

A context-free grammar can be used to do more than just demarcate sentences from nonsentences. It can also be used to describe the *hierarchical structure* of the sentences in a language. For instance, the grammar presented above describes (6) as having a particular structure, which can be represented with brackets (7), trees (8), or in ordinary English (9).

(6) I prefer a morning flight.

(7) [$_{S}$ [$_{NP}$ [$_{Pronoun}$ I]] [$_{VP}$ [$_{Verb}$ prefer] [$_{NP}$ [$_{Determiner}$ a] [$_{Nominal}$ [$_{N}$ morning] [$_{Nominal}$ [$_{N}$ flight]]]]]]

(8)



(9) The sentence 'I prefer a morning flight' consists of the concatenation of a noun phrase and a verb phrase.The noun phrase consists of the pronoun 'I'.The verb phrase consists of a concatenation of the verb 'prefer' and another noun phrase.This (second) noun phrase itself consists of a concatenation of the determiner 'the' and a nominal. The nominal consists of a concatenation of the noun 'morning' and another nominal.This remaining nominal consists of the noun 'flight'.

As (7)–(9) illustrate, it is trivial to express in ordinary English any claims that one can make in either of the other notations. But the bracket and tree notations can be very convenient for a number of purposes—brackets render any such claim shorter; trees render it easier to read. The important point to keep in mind, though, is that these are merely notational variants. Whatever the notation, (7)–(9) all serve to make claims about the structure of a sentence. Without further argument, nothing at all follows about the psychology of a language user, nor about the internal operations of a computational system designed to parse sentences of a language.

Let us now explore the role that CFGs play in existing parsing algorithms, including the top-down Earley algorithm (Sect. 8.3.1) and the bottom-up CYK algorithm (Sect. 8.3.2), both of which draw on explicitly represented context-free rules, which are stored in a separate data structure. We will also consider left-corner parsers (Sect. 8.3.3), and discuss how the Parsing as Deduction framework uses first-order logic and other familiar formalisms to render CFGs suitable for use in deductive proof procedures (Sect. 8.3.4).

8.3 The Algorithm

Computational linguists sometimes complain that psycholinguists' proposals concerning the operations of the HSPM are imprecise, in that they don't make clear the underlying computational architecture. Consider, for instance, the following remarks from Harkema (2001).

> The psycholinguist selects a set of sentences with grammatical constructions that are predicted to engage the human sentence processor in a way that will reveal its inner workings, designs an experiment with these sentences and runs it with a number of human subjects, interprets the results and proposes certain properties that the human sentence processor should have in order to account for the experimental results. The results of this approach greatly expand our knowledge about the human sentence processor, but it sometimes fails to satisfy linguists of a more formal bent, because a mere description of its properties does not constitute a complete specification of a parser. In most psycholinguistic proposals, the operations of the parser are only specified insofar as they bear on the natural language constructions being explored. Moreover, this specification is usually very informal. While the parser works for the set of examples provided, absence of a description of the overall architecture of the parser leaves open the question whether a parser with the desired properties covering the entire language actually exists. Another drawback of some of the work in psycholinguistics is that it is sometimes based on simplistic and outdated conceptions of syntactic structure. (pp. 2–3)

To avoid falling prey to such concerns, in what follows we will consider processing rules that are pitched at a fairly basic level—what one might call the "algorithmic level," following Marr (1980). In doing so, our primary goal will be to examine the ways in which psychologically plausible parsing models might make use of the structure rules of various grammars.

The discussion in this chapter draws heavily on Jurafsky and Martin (2008), a canonical text in computational linguistics. The authors open their chapter on parsing algorithms by drawing a distinction between structure rules and processing rules.

> We defined parsing … as a combination of recognizing an input string and assigning a structure to it. Syntactic parsing, then, is the task of recognizing a sentence and assigning a syntactic structure to it. This chapter focuses on the kind of structures assigned by context-free grammars …[S]ince they are based on a purely declarative formalism, context-free grammars don't specify *how* the parse tree for a given sentence should be computed. We'll therefore need to specify algorithms that employ these grammars to produce trees. (p. 431)[8]

---

[8] There is a distinction between "recognizing" a sentence and assigning a syntactic structure to it. A system's recognizing a string, or "accepting" it, amounts to no more than that system's issuing a judgment to the effect that the string in question is grammatical, relative to the grammar with

Jurafsky and Martin view the construction of syntactic representations as a necessary step toward the final goal of semantic interpretation.

> [P]arse trees serve as an important intermediate stage of representation for semantic analysis … and thus play an important role in applications like question answering and information extraction. For example, to answer the question *What books were written by British women authors before 1800?* we'll need to know that the subject of the sentence was *what books* and that the *by* adjunct was *British women authors* to help us figure out that the user wants a list of books (and not a list of authors). (p. 431)

They go on to discuss two of the most basic methods available for constructing parse trees: *top-down* and *bottom-up* parsing.[9] The distinction between these two approaches reflects a more fundamental distinction, with which philosophers are well acquainted.

> Regardless of the search algorithm we choose, there are two kinds of constraints that should help guide the search. One set of constraints comes from the data, that is, the input sentence itself. … The second kind of constraint comes from the grammar. … These two constraints … give rise to the two search strategies underlying most parsers: top-down or goal-directed search, and bottom-up or data-directed search. These constraints are more than just search strategies. They reflect two important insights in the western philosophical tradition: the rationalist tradition, which emphasizes the use of prior knowledge, and the empiricist tradition, which emphasizes the data in front of us. (p. 433)

In what follows, I present the top-down and bottom-up parsing algorithms in their most austere forms. I turn to various refinements in Sect. 8.4.

8.3.1 Top-Down Parsing and the Earley Algorithm

In a top-down parser, the process of assigning syntactic structure to a linguistic input proceeds in phases, sometimes called *plies*. At each ply, the parser uses the internally represented grammar to issue predictions, in something like the way that a scientist uses her background theory to issue predictions prior to performing an experiment. The parser begins by positing an initial tree-node—traditionally an S.[10] This node, alone, constitutes the first ply. In the second ply, the S node is expanded

---

which the system is operating. Parsing, by contrast, involves constructing one or more representations of the string's syntactic structure and, in the case of ambiguity, selecting one of these as the *privileged* representation—the one that constitutes the system's "ultimate decision" about how the input should be interpreted. For a discussion of this distinction, see Berwick and Weinberg (1984), p. 252.

[9] Fodor et al. (1974) refer to top-down and bottom-up techniques as *analysis-by-analysis* and *analysis-by-synthesis*, respectively. Computer scientists sometimes use the terms *recursive-descent* and *shift-reduce*, for reasons that will become apparent below.

[10] In contemporary syntactic theories in the P&P tradition (Chap. 9), the phrasal category S has been replaced by other phrasal types—different ones in different theories. The list includes, *inter alia*, inflectional phrase (IP), complementizer phrase (CP), and tense phrase (TP). For ease of exposition, I ignore this and related complications.

Grammatical rules	**Lexical rules**
S → NP VP	*Det → that | this | a*
S → Aux NP VP	*Noun → book | flight | meal | money*
S → VP	*Verb → book | include | prefer*
NP → Pronoun	*Pronoun → I | she | me*
NP → Proper-Noun	*Proper-Noun → Houston | NWA*
NP → Det	*Aux → does*
Nominal → Noun	*Preposition →from | to | on | near | through*
Nominal → Nominal Noun	
Nominal → Nominal PP	
VP → Verb	
VP → Verb NP	
VP → Verb NP PP	
VP → Verb PP	
VP → VP PP	
PP → Preposition NP	

Fig. 8.2 The context-free grammar 'G' (Adapted from Jurafsky and Martin 2008: p. 394)

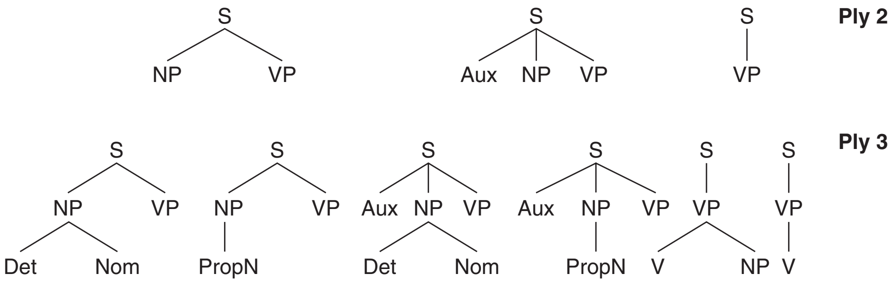


Fig. 8.3 Plies 2 and 3 in a simple top-down parse of the input *Book that flight*, assuming grammar G (Fig. 8.2)

in accordance with the internally represented grammar. For purposes of illustration, let us assume that the parser is working with a pitifully simple CFG, which we label 'G' in Fig. 8.2. Figure 8.3 depicts the parser's outputs at two successive plies.

In the second ply, the S node is expanded three separate times into (i) NP + VP, (ii) Aux + NP + VP, and (iii) a standalone VP. These are the three expansion options that G allows. They constitute the parser's disjunctive prediction to the effect that the input will consist of either (i) noun phrase and verb phrase, (ii) an auxiliary phrase, a noun phrase and a verb phrase, or (iii) just a verb phrase. In the third ply, these nodes are themselves expanded, again in accordance with G.

In theory, this process of expansion would proceed until all of the nodes are parts of speech, which cannot be expanded in accordance with G. When the parser reaches a ply in which all of the leaves (i.e., terminal nodes—usually words) of all of the parse trees are parts of speech, it finally looks at the input string to see whether it matches any of the items in the lexicon.[11] This, in turn, would allow the parser to see which of its predictions turned out to be correct. For instance if the input is (10), whose structure is given in (11), then the parser could immediately rule out any parse tree whose first leaf is *Determiner*.

(10) Book that flight.
(11) $[_S\,[_{VP}\,[_V$ Book]] $[_{NP}\,[_{Det}$ that]$[_{Nom}\,[_N$ flight]]].

It could then systematically rule out all of the other parse trees, except the correct one. The parsing process would then be complete.

A major complication is that if grammar G were recursive, the parser would never reach a ply in which all of the leaves of the parse trees are terminal nodes—i.e., parts of speech. Consider the combined effect of the following two rules.

(12) a. *S → NP VP*
b. *VP → Verb Complementizer S*

These rules will conspire to continuously create more and more candidate parse trees by expanding the S node into parse trees that contain a VP node, and then expanding that VP node into parse trees that contain an S node. The process would repeat without terminating.

This problem can be overcome by making use of the classic "generate-and-test" strategy. The internally represented phrase-structure rules can generate some candidate parse trees (say, two plies' worth) and then stop, allowing the system to access some of the input data prior to generating more parse trees. The strategy is no longer *purely* top-down, as it is allowed to "look at" the input prior to generating more predictions, but this is motivated by the need to avoid issuing infinitely many predictions.

Though it is no longer purely top-down, the parser is still quite simplistic and inefficient. The top-down process still generates an unacceptable number of predictions that ultimately fail to accord with the input. Indeed, many of its predictions might be consistent with the input *up until a point*, but then disconfirmed thereafter. And because it takes time and computational resources for a physical system to generate such predictions, each prediction that is ultimately inconsistent with the

[11] In principle, a top-down parser could make predictions even about which *specific lexical items* will appear in the input stream. But such predictions, even if made on the basis of frequency information and pragmatic/contextual clues, would still be rather risky. I assume, then, that lexical retrieval is in significant measure a data-driven process. This raises a question about whether the "matching" involved in lexical recognition is brute-causal, in the sense introduced by Devitt (2006a). The answer is that it's not. The HSPM constructs *phonological* representations prior to, and in the service of, lexical retrieval. As with syntactic processing, the activation of a phonological representation is highly context-sensitive and dependent on factors that are not present in the immediate stimulus, but scattered over discontinuous chunks of time (Fernández and Cairns 2011: ch. 6). Moreover, higher-level syntactic decisions exert a downward influence on lexical retrieval and actively guide the correction of errors in the retrieval process.

input is a waste of limited time and resources. With a realistic broad-coverage grammar, the resulting inefficiency is devastating.

Another possibility is to *pre-compute* the search space allowed by the grammar and store it in memory so that the parser does not have to generate all possible parse trees each and every time before checking the input. As Marcus (1980: p. 29) points out, the parsers known as STRING (by Naomi Sanger), LUNAR (by Bill Woods), and SHRDLU (by Terry Winograd) are all instances of the top-down approach. Focusing his attention on the LUNAR parser, Marcus notes that it is based on an ATN grammar, which is essentially a pre-defined search space that models the syntactic options of an interesting fragment of English.[12] The job of the parser, then, is to navigate the search space. Parsing thus becomes a search task. Marcus goes on to argue that this, too, is highly inefficient, because at each choice point in the search, the parser will have to make a decision that is potentially incorrect—an error that might be discovered arbitrarily late in the sentence. Consider, for instance, the following sentences.

(13) Is the woman in the office?
(14) Is the woman in the office on the third floor?
(15) Is the woman in the office on the third floor of the new building?

(16) Is the woman in the office helpful?
(17) Is the woman in the office on the third floor helpful?
(18) Is the woman in the office on the third floor of the new building helpful?

If the parser assumes that "the woman in the office…" is a reduced relative clause, then when it encounters the question mark in (13)–(15), it will have to backtrack, retracting its commitments earlier in the sentence, one by one, and trying all of the available options at each prior choice point, until it reaches the site of the initial error. If the parser instead assumes that the "the woman in the office…" is a noun phrase, it will encounter the same problem when it reaches the word 'helpful' in (16)–(18). Either assumption, when found to be incorrect, leads to an unacceptable amount of backtracking and revision of commitments. (See Marcus 1980: Chaps. 1 and 2 for more in-depth discussion.)

A solution to some these problems was developed by Earley (1970).[13] The Earley algorithm, named after its inventor, avoids generating the same partially successful predictions again and again at each successive ply, or backtracking through prior choice points one by one and duplicating work already done. This is accomplished by storing some number of successful *partial* parse trees in a data-structure known

[12] The infinitude of the space is handled by a simple formal device that we shall introduce in our discussion of ATN parsers in Chap. 9. For a visual aid, the reader can skip down to Figs. 8.11, 8.12, 8.13, which illustrate the type of search space I have in mind here.

[13] The Earley algorithm is discussed, in various levels of detail, in Hale (2001), Jurafsky and Martin (2008), Kilbury (1985), Harkema (2001), Aho and Ullman (1972), Morawietz (2000), Pereira and Warren (1983), and Shieber, Schabes, and Pereira (1993).

as a “chart” or a “well-formed substring table.” The successful partial parses are then called on when needed, instead of having to be reconstructed anew.

The Earley algorithm is widely used in top-down parsing systems and has been applied to probabilistic context-free grammars (Hale 2001, 2003), as well as to Minimalist and other mildly context-sensitive grammars (Harkema 2001). Let’s take a closer look at its operations. The algorithm (Fig. 8.5) has three main functions: *Predict*, *Scan*, and *Complete*. The *Predict* function adds to the chart all expansions of a node that are consistent with the grammar, which is internally represented in a declarative data structure. The *Scan* function looks at the input, prior to issuing further predictions. The *Complete* function stores all of the confirmed predictions, and propagates them to the next stage of the process, while removing from the chart all of the disconfirmed predictions. The rules of the internally represented grammar are annotated, by convention, with a dot (‘•’), which indicates how much of a rule has already been completed. Figure 8.4 illustrates the operations of the Earley parser on the input sentence ‘Book that flight’.

A cursory glance at the pseudocode provided by Jurafsky and Martin (2008) reveals that each of the three main functions of the Earley algorithm draw on an internally represented grammar (Fig. 8.5). This is particularly obvious in the case of the *Predictor* function, which makes explicit reference to a data structure that the authors have aptly labeled GRAMMAR-RULES.

It is a significant feature of the parser described here that the rules of the grammar are stored in a separate data structure. This makes it possible to swap out one grammar and replace it with another, leaving the rest of the system intact. Other parsers do not share this feature. For instance, the Marcus parser (Marcus 1980), and the Augmented Transition Networks models that we discuss in the next chapter, use the rules *procedurally*. That is, the rules are *embodied* in the parser’s operations, not explicitly represented and accessed as data.

A number of refinements can be introduced to make the Earley algorithm more efficient. For instance, instead of blindly making predictions on the basis of all of the rules in its internally represented grammar, the parser can be equipped with an oracle (Sect. 8.4), that decides intelligently which rules to apply at each point in the parse, and in what order to call the *Predict*, *Scan*, and *Complete* functions.[14] This allows for an implementation of the least-effort parsing principles discussed above. For instance, Late Closure dictates that the HSPM will incorporate newly encountered material into the most recent phrase or clause. This can be implemented by having the predictor insert into the chart only those rules that expand nodes in such a way that an already-recognized category appears to left of the dot. Another method that has been used to combat the problem of inefficiency makes use of probabilistic information. The Earley algorithm has been applied to Probabilistic Context-Free Grammars (PCFGs), which include information about the frequency with which

[14] For details, see Jurafsky and Martin (2008: pp. 452–454). Kaplan (1973) contains an early but prescient discussion of various chart-parsing techniques.

Chart[0]	S0	*γ → • S*	[0,0]	Dummy start state
	S1	*S → • NP VP*	[0,0]	Predictor
	S2	*S → • Aux NP VP*	[0,0]	Predictor
	S3	*S → • VP*	[0,0]	Predictor
	S4	*NP → • Pronoun*	[0,0]	Predictor
	S5	*NP → • Proper-Noun*	[0,0]	Predictor
	S6	*NP → • Det Nominal*	[0,0]	Predictor
	S7	*VP → • Verb*	[0,0]	Predictor
	S8	*VP → • Verb NP*	[0,0]	Predictor
	S9	*VP → • Verb NP PP*	[0,0]	Predictor
	S10	*VP → • Verb PP*	[0,0]	Predictor
	S11	*VP → • VP PP*	[0,0]	Predictor
Chart[1]	S12	*Verb → book •*	[0,1]	Scanner
	S13	*VP → Verb •*	[0,1]	Completer
	S14	*VP → Verb • NP*	[0,1]	Completer
	S15	*VP → Verb • NP PP*	[0,1]	Completer
	S16	*VP → Verb • PP*	[0,1]	Completer
	S17	*S → VP •*	[0,1]	Completer
	S18	*VP → VP • PP*	[0,1]	Completer
	S19	*NP → • Pronoun*	[1,1]	Predictor
	S20	*NP → • Proper-Noun*	[1,1]	Predictor
	S21	*NP → • Det Nominal*	[1,1]	Predictor
	S22	*PP → • Prep NP*	[1,1]	Predictor
Chart[2]	S23	*Det → that •*	[1,2]	Scanner
	S24	*NP → Det • Nominal*	[1,2]	Completer
	S25	*Nominal → • Noun*	[2,2]	Predictor
	S26	*Nominal → • Nominal Noun*	[2,2]	Predictor
	S27	*Nominal → • Nominal PP*	[2,2]	Predictor
Chart[3]	S28	*Noun → flight •*	[2,3]	Scanner
	S29	*Nominal → Noun •*	[2,3]	Completer
	S30	*NP → Det Nominal •*	[1,3]	Completer
	S31	*Nominal → Nominal • Noun*	[2,3]	Completer
	S32	*Nominal → Nominal • PP*	[2,3]	Completer
	S33	*VP → Verb NP •*	[0,3]	Completer
	S34	*VP → Verb NP • PP*	[0,3]	Completer
	S35	*PP → • Prep NP*	[3,3]	Predictor
	S36	*S → VP •*	[0,3]	Completer
	S37	*VP → VP • PP*	[0,3]	Completer

Fig. 8.4 The first column contains labels of the successive charts that are created in the course of parsing the sentence '$_0$ Book $_1$ that $_2$ flight $_3$' (where the subscripted numbers denote positions in the sentence). The second column contains labels of the steps that constitute the process. The third column displays the "dotted" rules of the internally represented grammar. Before the process is complete, the right-hand side of each rule has some nodes to the right of the dot ('•'), indicating that these nodes have not yet been expanded. The nodes to the left of the dot have been successfully expanded. For instance, state S14 indicates that the *Verb* node has been expanded and completed, but the *NP* node has not. The fourth column indicates how much of the material from the input has been successfully integrated. For instance, '[1,3]' in line S30 indicates that this state has managed to integrate the words 'that flight' in '$_0$ Book $_1$ that $_2$ flight $_3$'. The fifth column displays the function responsible for entering each state into the chart. A parse is complete when all of the input material has been integrated and the dot is on the rightmost edge of a rule that begins with the symbol 'S'—in this case, state S36. *Source*: Jurafsky and Martin (2008), p. 451

```
procedure PREDICTOR((A → a • B b , [i, j]))
        for each (B → g ) in GRAMMAR-RULES-FOR(B, grammar) do              ← !!!
            ENQUEUE((B → • g , [ j, j]), chart[ j])
    end

procedure SCANNER((A → a • B b , [i, j]))                                  ← !!!
        if B ⊂PARTS-OF-SPEECH(word[j]) then
            ENQUEUE((B → word[ j], [ j, j+1]), chart[j+1])
    end

procedure COMPLETER((B → g •, [ j,k]))
        for each (A → a • B b , [i, j]) in chart[ j] do                    ← !!!
            ENQUEUE((A → a B • b , [i,k]), chart[k])
    end
```

Fig. 8.5 An excerpt from the pseudocode for the Earley algorithm. I have marked explicit references to grammatical rules with the symbol '!!!' (*Source*: Jurafsky and Martin 2008, p. 448)

each rule of the grammar is applied, as well as the frequencies of particular lexical items. We discuss this strategy in Sect. 8.4.2.[15]

8.3.2 Bottom-Up Parsing and the CYK Algorithm

Bottom-up parsing was the earliest approach to have been explored, with proposals dating back to 1955. The pioneering research was conducted by computer scientists, largely in connection with artificial programming languages (Aho and Ullman 1972). Kartunnen and Zwicky (1985) discuss the relation between parsing algorithms designed for these purposes and ones that aim for psychological plausibility. As they point out, the concerns of psycholinguists and computer scientists overlap in a number of respects, though the differences between natural languages and programming languages sometimes obscure this. Whereas artificial languages tend to be simple, largely unambiguous, and have a known grammar, natural language is complex, rife with ambiguity, and not yet fully understood from a formal perspective.[16] Moreover, the notions of communication, discourse, and semantic interpreta-

[15] See also Schabes, Abeille, and Joshi (1988), who provide an instructive application of the Earley algorithm to lexicalized versions of context-free grammars, as well as to the mildly context-sensitive tree-adjoining grammar (TAG). The authors discuss the considerable gains in efficiency stemming from the lexicalization of these grammars.

[16] The terms "simple" and "complex" can be given a formal interpretation. Whereas programming languages tend to belong to the class of context-free languages, it has been known for some time that natural languages are slightly stronger than context-free. This was demonstrated by Shieber (1985), who discussed cases of cross-serial dependencies in Dutch and in Scandinavian languages. Syntacticians are in the process of constructing formalisms that fit this specification while avoiding overgeneration—i.e., without allowing the formulation of grammars that are not attested by any known natural language. The Minimalist grammars currently being explored in one branch of syntactic theory are committed to the existence of discontinuous constituents—syntactic chains in

tion are significantly different when applied to programming languages, as against natural languages. Nevertheless, with respect to syntax and parsing, there is much that the computer scientist and the psycholinguist can learn from one another.[17] The fact that bottom-up parsing is used in both computational and psycholinguistic applications is an instance of this fruitful exchange of ideas.

The simplest bottom-up parser is data-driven; it begins by immediately looking at the input, rather than making predictions about it. Having searched the lexicon for the grammatical categories to which the words in the input can belong, the parser begins to build trees that are compatible with those categories and the grammar. At each step, the parser "looks for places in the parse in progress where the right-hand side of some [context-free] rule might fit" (Jurafsky and Martin 2008: p. 434). Figure 8.10 shows how a bottom-up parser can draw on explicit representations of the rules and use them as templates to determine which parses are licensed at each ply.

Bottom-up parsers, better known to computer scientists as *shift-reduce* parsers, operate by attempting to integrate already recognized lexical items into phrasal groups.[18] The parsing process is completed when the partial structures that were built out of the items in the input are integrated into a full sentence—a structure with an S node at its root. In general, it will not be possible to incorporate *all* of the partial structures into a tree rooted in an S node.[19] Many of these will turn out to be

which antecedents do not c-command the traces or copies that they leave behind after movement. A grammar that generates discontinuous constituents is stronger than context-free. Minimalist grammars thus belong to the class of mildly context-sensitive grammars (Stabler 2001.) Shieber, Schabes, and Pereira (1993) discuss other mildly context-sensitive grammars, e.g., tree-adjoining grammars (TAGs), and provide a schema for constructing efficient parsers for these grammars.

[17] Though we'll see in Sect. 8.4.2 that the differences between their concerns can also cause confusion.

[18] Another term for these parsers that may be familiar to computer scientists is 'LR(k)'. The symbol 'LR' encodes the fact that the parser works from left to right. The symbol '(k)' is a variable that defines the size of the parser's "look-ahead window"—i.e., how much of the input it is allowed to take in before initiating some operation. An LR(2) parser, for instance, can look at two items of the input before deciding what to do next.

[19] There is a way in which this sort of remark can be misleading. Devitt (2006a: pp. 69–71) notes an important distinction between the expressions of a language, on the one hand, and structural *descriptions of* those expressions on the other. The latter can be derived from a theory of a language, but the former cannot. "[W]hat is derived from a grammar is not an expression of the language but a description of an expression, just as what is derived from an astronomical theory is not, say, a star, but a description of a star" (p. 69). If we are careful not to run afoul of this use/mention distinction, we must say that the parser uses its internally represented grammar—conceived now as (a subpersonal analogue of) a *theory* of a language—to generate *structural descriptions* of incoming linguistic stimuli. Hence, when we say that the parser produces "a structure that has an S node at its root," we do *not* mean that it produces a sentence; rather the parser produces a description of the incoming stimulus, thus characterizing the stimulus *as* a structure that has an S node at its root. The question then arises: How do we get from such a descriptive characterization to the final product of language comprehension? If the final representation *uses* words, rather than *mentioning* them, then what accounts for the transition between a description (which merely mentions the words) to the final product (which uses them)? This puzzle disappears if say that the final representation merely "mentions" the words, though in a sense that's closer to indirect discourse (e.g., "Kurt just said that it's raining.")

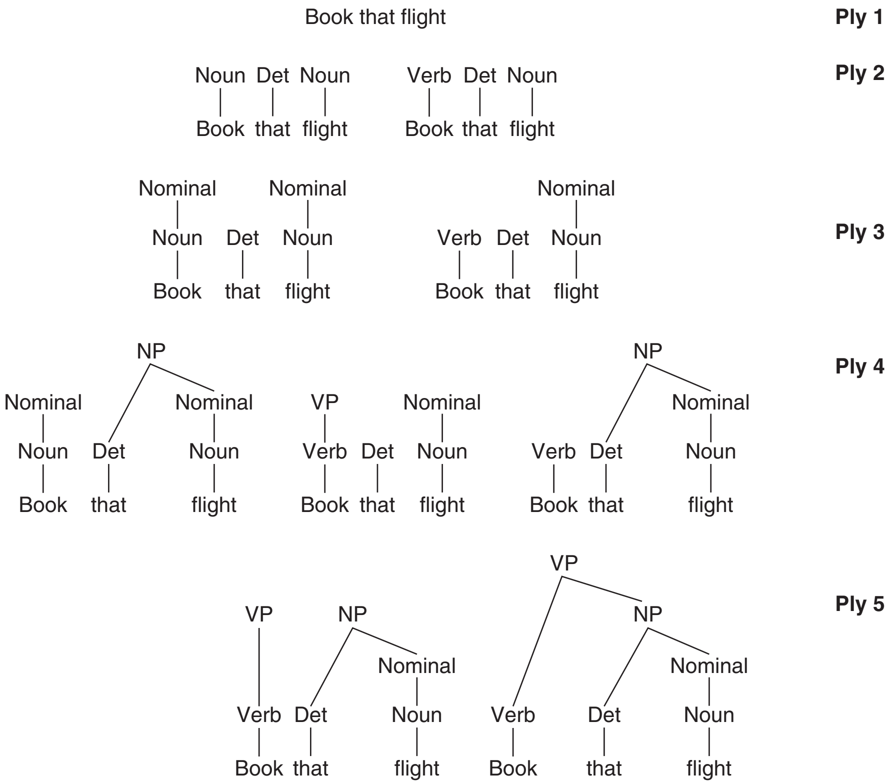


Fig. 8.6 The first five plies of a bottom-up parse of *Book that flight* (*Source*: Jurafsky and Martin 2008, p. 435)

"duds" and will have to be discarded. But if the input is a sentence of the language that is adequately captured by the parser's internally represented grammar, then at least one of the parser's initial guesses about the structure of the input will turn out to be correct.

Figure 8.6 illustrates the first five plies of a bottom-up parse of (25), by reference to the grammar G, which we introduced in our discussion of top-down parsers (Fig. 8.2).

(25) Book that flight.

Figure 8.6 seems to suggest that all of the lexical items are recognized at the same time and that all of the structures are built simultaneously only after the full input sentence has been scanned. This need not be the case.[20] To see this point, let's exam-

[20] Indeed, it had better *not* be the case, if psychological plausibility is our goal. As noted in Chaps. 5 and 6, the HSPM is an "eager" mechanism—the assignment of syntactic structure is never delayed; syntactic analysis begins at the very first hint of linguistic input and continues incrementally, morpheme by morpheme.

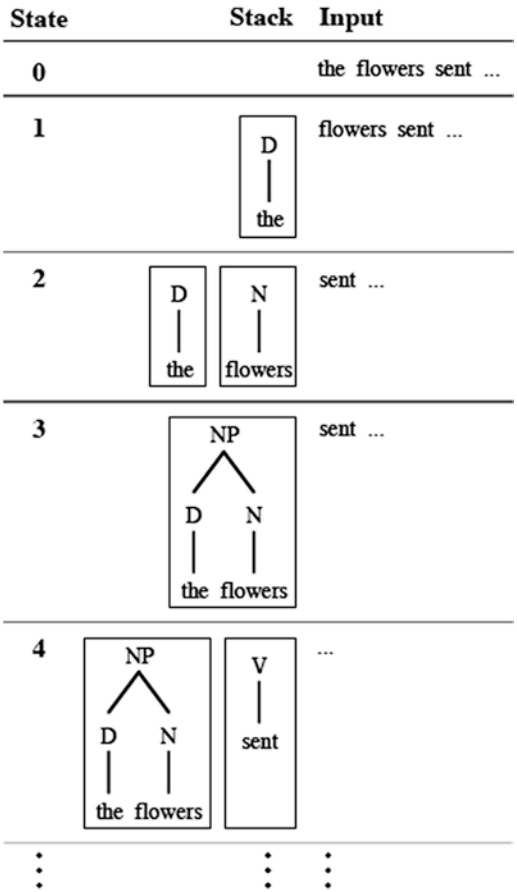


Fig. 8.7 An incremental bottom-up parse of *The flowers sent…*. (*Source*: Shieber and Johnson 1993)

ine the bottom-up parsers that computer scientists call *shift-reduce* parsers, which operate on essentially the principles described above. These have two basic operations: *Shift* and *Reduce*. The *Shift* operation puts new material from the input into a memory stack. The *Reduce* operation unifies distinct items into a single unit. It does so by taking sequences of items from the stack and, if they match the right-hand side of an internally represented grammar rule, replacing them by the left-hand side. For instance, if the stack contains an NP and a VP, and the grammar contains the rule *S* → *NP VP*, then the NP and VP will be "reduced" to an S—i.e., the system will posit an S node that immediately dominates both the NP and the VP. This process is illustrated in Fig. 8.7.[21]

The inefficiency of top-down parsers, we saw, stems from the fact that they make predictions without first looking at the input, thus wasting resources on predictions that are ultimately found to be inconsistent with the input. Although it turns out that

[21] Similar visual aids can be found in Abney and Johnson (1991) and Crocker (1999).

bottom-up parsing is, in practice, more efficient than top-down parsing (O'Donnell 1995: Chap. 9), there are nevertheless several problems with bottom-up parsers as well. Looking at Fig. 8.6, we notice that many of the structures that the parser considers at the first ply do not—indeed, cannot—eventuate in an S. Building these structures is therefore a waste of time and resources. In the case of long, highly-ambiguous inputs, one and the same mistake might be made several times over. A top-down parser would never make such mistakes, because it starts out predicting only legal trees—ones that have an S node at their root. By contrast, a bottom-up parser has no way of knowing in advance which structures will be successful as higher and higher nonterminal nodes are "reduced." A purely bottom-up parser must have some mechanism for either ignoring the mistakes it makes or fixing them when necessary. The former option is inherent in the parallel parsing approach, which simply sets aside the unsuccessful parse trees that are constructed on the way to the final, successful parse. However, parallel architectures require large amounts of memory or storage capacity. The other option—fixing mistakes—necessitates backtracking and revising earlier commitments. As noted above, backtracking requires a great deal of time and resources in realistic applications.

We saw earlier that some of the inefficiency associated with a simplistic top-down approach can be dealt with by storing successful partial parses in a data structure known as a "chart," and calling on them in order to construct a solution to a larger problem. A similar technique is available for the bottom-up approach. A particularly elegant parsing algorithm was discovered in the late 1960s by three separate researchers: John Cocke, Daniel Younger, and Tadao Kasami. The algorithm is frequently labeled 'CYK' (occasionally 'CKY'), in honor of its discoverers. Jurafsky and Martin (2008), describe the basic operations of the CYK algorithm, using as an example the input sentence (26).

(26) Book the flight through Houston

First, a chart is constructed—a matrix of cells with individual cells corresponding to the words in the input. This is shown in Fig. 8.8a. The cells marked [0,1], [1,2], [2,3], [3,4], and [4,5] correspond to the words 'Book', 'the', 'flight', 'through', and 'Houston', respectively. The algorithm begins by using its lexical rules to identify the parts of speech to which these words belong. It then runs through the remaining cells from left to right, bottom to top, as illustrated in Fig. 8.8b. In accordance with the rules of an internally represented context-free grammar,[22] cells that occupy higher positions in each column are filled in with higher-level phrasal categories, on the basis of the contents of the cells that have already been viewed. For instance, in accordance with the rule *NP→ Det Noun*, an NP is placed in cell [1,3] on the basis of there being a determiner (*Det*) in cell [1,2] and an noun in cell [2,3]. Similarly, the rule *VP→ Verb NP* is used to fill cell [0,3] with a VP, on the basis of there being

[22] Note that the CFG must first be transformed into a binary-branching format known as Chomsky Normal Form (CNF). This transformation is well-defined and computationally trivial. See Jurafsky and Martin (2008: pp. 441–2).

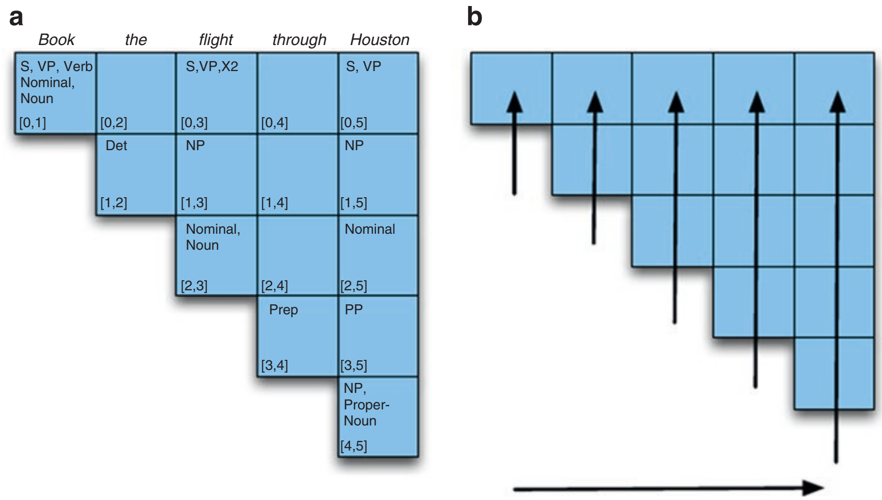


Fig. 8.8 The CYK algorithm in action. The figure in (**a**) shows a completed chart for sentence (26). The cells at the bottom of each column correspond to the words in the sentence and encode their parts of speech. The algorithm proceeds from *left to right*, *bottom to top*, as shown in figure (**b**) (*Source*: Jurafsky and Martin 2008, p. 443)

```
function CKY-PARSE(words, grammar) returns table                          ← !!!
    for j←from 1 to LENGTH(words) do
        table[j–1, j]←{A | A → words[j] ∈grammar }                         ← !!!
        for i←from j–2 downto 0 do
            for k←i+1 to j–1 do
                table[i,j]←table[i,j] ∪
                    {A | A → BC ∈ grammar,                                 ← !!!
                         B ∈ table[i,k],
                         C ∈ table[k,j] }
```

Fig. 8.9 The pseudocode for the CYK algorithm. I have marked explicit references to grammatical rules with the symbol '!!!' (*Source*: Jurafsky and Martin 2008, p. 444)

a *Verb* in cell [0,1] and an NP in cell [2,3]. This process is repeated, left-to-right, bottom-to-top until all of the cells have been filled. The success of the parse is indicated by the presence of an S node in the very last cell, marked [0,5].

Just as we did with the Earley algorithm, we can survey the pseudocode of the CYK algorithm and see that explicit reference is made to a data-structure in which the rules of the grammar are explicitly represented (Fig. 8.9).

The reader may also be interested to learn that the CYK algorithm has been successfully implemented in connectionist networks. This requires converting a context-free grammar into what Smolensky and Legendre (2006) call a *Harmonic Grammar*. Entering into the details would, unfortunately, take us too far afield. A full description of this implementation can be found in Hale (1999) and Hale and Smolensky (2006).

8.3.3 Left-Corner Parsing

Purely top-down and bottom-up parsers have problems with memory capacity, some of which we mentioned above. Harkema (2001) summarizes one of the problems as follows:

> Pure bottom-up and top-down parsers are generally rejected as adequate models for human sentence processing because they require unbounded memory for right-branching and left-branching structures respectively, while the human sentence processor, which is equipped with finite memory, seems to have no particular problems with structures of this kind (e.g., Crocker 1999).

There is, however, a third approach, which combines the insights of both top-down and bottom-up parsers. Known as *left-corner parsing*, this technique uses bottom-up methods to look at the input, thus avoiding the pitfalls associated with blind guessing, and then top-down methods to make informed predictions about further input. The strategy is to build each constituent bottom-up and then posit a plausible parent node as well as whatever further structure is required by that node. Figure 8.10 illustrates this idea.[23] Note again that which predictions are made on the basis of the input is a function of what grammar the algorithm employs.

Left-corner parsing is psychologically plausible on a number of measures.[24] First, it is incremental, which we know the HSPM to be (Chap. 5). Second, unlike the bottom-up and top-down parsers, which have problems with right- and left-branching structures, the left-corner parser requires unbounded memory only for center-embedded structures. (Figure 8.11 provides schemas for the three types of structure.) These are known to be problematic for human readers and listeners, as the example in (27) illustrates.[25]

(27) The guitar the woman the man loved played is broken.

For these reasons, many researchers now subscribe to the view that the HSPM is itself a left-corner parser.

Dynamical chart-parsing techniques like the Earley and CYK algorithms have been applied to the left-corner parsing approach. Moore (2000) presents a comparison between the performance of a left-corner chart-parsing system and that of previous parsers. His analysis concludes that left-corner chart-parsers enjoy an overall lead in efficiency. We will see below (Sect. 8.4.2) that the left-corner strategy is particularly useful for probabilistic parsers (Hale 2011).

---

[23] Similar visual aids can be found in Abney and Johnson (1991) and Crocker (1999).

[24] See Abney and Johnson (1991), Stabler (1994), Crocker (1999), Harkema (2001). Abney and Johnson (1991) distinguish between what they call arc-eager and arc-standard left-corner parsers. The latter, they argue, are less efficient than the former. I omit the details here. Note also that left-corner parsing bears a close resemblance to what Fodor and Frazier (1980) term 'information-paced parsing'. Frazier and Fodor argue that strictly top-down and bottom-up routines are too rigid to deal effectively with locally ambiguous inputs.

[25] For an in-depth discussion of center-embedding, see Thomas (1995). Stabler (1994) discusses a range of center-embedded structures and related constructions from languages other than English.

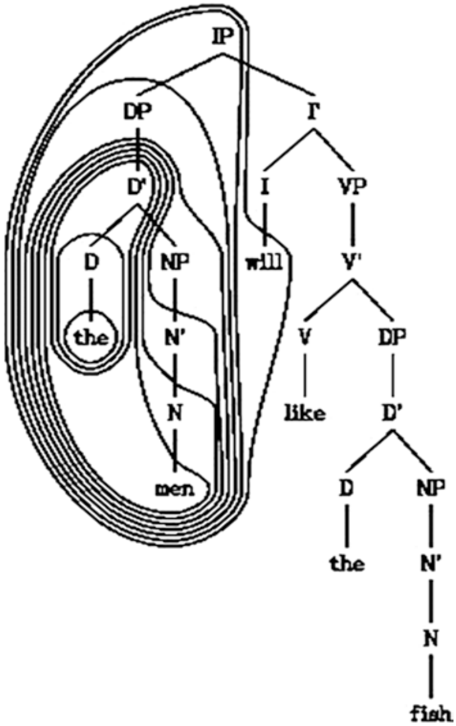


Fig. 8.10 A left-corner parse of *The men will like the fish*. The figure is from Stabler (1994), who describes the process as follows: "We can diagram an LC parse of a sentence by circling the nodes that have been completed at each step. We have done this for the first 10 steps of a left corner parse [here]. The steps indicated there are the following, beginning with the smallest circle:
(i) The word 'the' is heard.
(ii) The parent D is built, since its left corner ('the') has been completed.
(iii) The parent D is built, since its left corner D is complete, and the sibling NP is predicted.
(iv) The word 'men' is heard.
(v) The parent N is built, since its left corner is complete.
(vi) The parent N is built, since its left corner is complete.
(vii) The parent NP is built, since its left corner is complete, and this constituent is attached as the predicted NP.
(viii) The parent DP is built, since its left corner D is complete.
(ix) The parent IP is built, since its left corner DP is complete.
(x) The word 'will' is heard"

8.3.4 Parsing as Deduction

An exciting development in parsing theory, dating back to Pereira and Warren (1983), is the interpretation of the parsing process as a natural deduction procedure in first-order logic. The *Parsing as Deduction* approach (henceforth, PAD) views parsing as a species of theorem-proving, in which the grammar is formulated as a set of axioms. The deductive procedure begins with these axioms and derives a

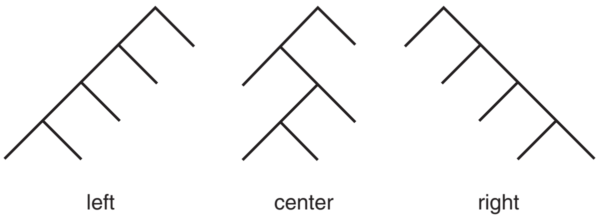


Fig. 8.11 Possible tree-geometries (*Source*: Abney and Johnson 1991)

theorem to the effect that the input string is a sentence of the language generated by the grammar. Harkema (2001) summarizes this idea as follows:

> Taking a logical perspective on parsing, the definition of the recognizer includes a specification of a grammatical deductive system and a specification of a deduction procedure. The formulae of the deductive system, which are commonly called items, express claims about grammatical properties of strings. Under a given interpretation, these claims are either true or false. For a given grammar and input string, there is a set of items that, without proof, are taken to represent true grammatical claims. These are the axioms of the deductive system. Goal items represent the claim that the input string is in the language defined by the grammar. Since our objective is to recognize a string, the truth of the goal items is of particular interest. The deductive system is completed with a set of inference rules, for deriving new items from ones derived earlier. The other component of the definition of the recognizer is the specification of a deduction procedure. This is a procedure for finding all items that are true for a given grammar and input string. (p. 96)

PAD is a vivid implementation of **REP-GRAM-DATA**, the idea that grammatical rules are explicitly represented and used "as data." Here, we introduce the approach by reference to CFGs. In the next chapter, we discuss its application to GB and Minimalist grammars.

CFG rules can be formulated in the language of first-order quantified predicate logic (with identity). For instance, the rule *S* → *NP VP* would be translated into (28).

(28) $\forall x\, [\exists yz\, [(\mathrm{NP}(y)\ \&\ \mathrm{VP}(z))\ \&\ x = yz] \Rightarrow \mathrm{S}(x)]$

In their classic discussion of PAD, Pereira and Warren (1983) opt for a slightly different notation, known as Definite Clause Grammar (DCG). In DCG, the rule *S* → *NP VP* would be translated into the formula in (29). (All variables are implicitly quantified.)

(29) $\mathrm{S}(s_0,s_2) \Leftarrow \mathrm{NP}(s_0,s_1)\ \&\ \mathrm{VP}(s_1,s_2)$

Pereira and Warren use the 's' variables to represent positions in the input string, so (29) can be paraphrased as follows: The material between positions 0 and 2 in the input string constitutes a sentence if between positions 0 and 1 there is a noun phrase and between positions 1 and 2 there is a verb phrase. Pereira and Warren go on to show how the Earley algorithm discussed above can be reinterpreted in the PAD framework. The basic functions of the parser—*Scan*, *Predict*, and

Complete—are interpreted as inference rules, on a par with, say, *modus ponens* and *existential instantiation* in first-order logic.

The example below, adapted from Pereira and Warren (1983) shows how the Earley algorithm would appear when rendered in the deductive formalism. In the justifications table, *instantiation* corresponds to the *Predict* function, while *reduction* corresponds to the *Scan* and *Complete* functions. Sentence (30) is assumed to have been divided into distinct morphemes. The outputs of a background lexical decision process are treated as axioms for the purpose of the deduction.

(30) $_0$ **Agatha** $_1$ **'s** $_2$ **husband** $_3$ **hit** $_4$ **Ulrich** $_5$

Lexical rules	**DCG notation**	**English translation**
Axiom 1	N(0,1)	The item between positions 0 and 1 is a noun
Axiom 2	Gen(1,2)	The item between positions 1 and 2 is a genitive marker
Axiom 3	N(2,3)	The item between positions 2 and 3 is a noun
Axiom 4	V(3,4)	The item between positions 3 and 4 is a verb
Axiom 5	N(4,5)	The item between positions 4 and 5 is a noun

Grammar rules	**DCG notation**	**Context-free notation**
Axiom 6	$S(s_0,s) \Leftarrow NP(s_0,s_1)$ & $VP(s_1,s)$	S → NPVP
Axiom 7	$NP(s_0,s) \Leftarrow Det(s_0,s_1)$ & $N(s_1,s)$	NP → Det N
Axiom 8	$Det(s_0,s) \Leftarrow NP(s_0,s_1)$ & $Gen(s_1,s)$	Det → NP Gen
Axiom 9	$Det(s_0,s) \Leftarrow Art(s_0,s)$	Det → Art
Axiom 10	Det(s,s)	Det → ∅
Axiom 11	$S(s_0,s) \Leftarrow NP(s_0,s_1)$ & $VP(s_1,s)$	VP → V NP

Deductive procedure		**Justifications**
Line 1	$ans \Leftarrow S(0,5)$	Goal statement (i.e., assumption)
Line 2	$S(0,5) \Leftarrow NP(0,s_1)$ & $VP(s_1,5)$	Line 1 instantiates Axiom 6
Line 3	$NP(0,s) \Leftarrow Det(0,s_1)$ & $N(s_1,s)$	Line 2 instantiates Axiom 7
Line 4	$Det(0,s) \Leftarrow NP(0,s_1)$ & $Gen(s_1,s)$	Line 3 instantiates Axiom 8
Line 5	$Det(s_0,s) \Leftarrow Art(0,s)$	Line 3 instantiates Axiom 9
Line 6	$NP(0,s) \Leftarrow N(0,s)$	Axiom 10 reduces Line 3
Line 7	NP(0,1)	Axiom 1 reduces Line 6
Line 8	$S(0,5) \Leftarrow VP(1,5)$	Line 7 reduces Line 2
Line 9	$VP(1,5) \Leftarrow V(1,s_1)$ & $NP(s_1,5)$	Line 8 instantiates Axiom 11
Line 10	$Det(0,s) \Leftarrow Gen(1,s)$	Line 7 reduces Line 4
Line 11	Det(0,2)	Axiom 2 reduces Line 10
Line 12	$NP(0,s) \Leftarrow N(2,s)$	Line 11 reduces Line 7
Line 13	NP(0,3)	Axiom 3 reduces Line 12
Line 14	$S(0,5) \Leftarrow VP(3,5)$	Line 13 reduces Line 2

Line 15	Det(0,s) $\Leftarrow$ Gen(3,s)	Line 13 reduces Line 4
Line 16	VP(3,5) $\Leftarrow$ V(3,s_1) & NP(s_1, 5)	Line 14 instantiates Axiom 11
Line 17	VP(3,5) $\Leftarrow$ NP(4,5)	Axiom 4 reduces Line 16
Line 18	NP(4,5) $\Leftarrow$ Det(4,s_1) & N(s_1,5)	Line 17 instantiates Axiom 7
Line 19	Det(4,s) $\Leftarrow$ NP(4,s_1) & Gen(s_1,s)	Line 18 instantiates Axiom 8
Line 20	Det(4,s) $\Leftarrow$ Art(4,s)	Line 18 instantiates Axiom 9
Line 21	NP(4,s) $\Leftarrow$ Det(4,s_1) & N(s_1,5)	Line 19 instantiates Axiom 7
Line 22	NP(4,5) $\Leftarrow$ N(4,5)	Axiom 10 reduces Line 18
Line 23	NP(4,s) $\Leftarrow$ N(4,s)	Axiom 10 reduces Line 21
Line 24	NP(4,5)	Axiom 5 reduces Line 22
Line 25	VP(3,5)	Line 24 reduces Line 17
Line 26	Det(4,s) $\Leftarrow$ Gen(5,s)	Line 24 reduces Line 19
Line 27	S(0,5)	Line 25 reduces Line 14
Line 28	ans	Line 27 reduces Line 1

Shieber, Schabes, and Pereira (1993) extend the work of Pereira and Warren (1983) by subjecting the CYK algorithm and a variety of other parsing algorithms to the same treatment.

> [W]e will in this paper investigate … how to synthesize parsing algorithms by combining specific logics of grammaticality claims with a fixed search procedure. In this way, deduction can provide a metaphor for parsing that encompasses a wide range of parsing algorithms for an assortment of grammatical formalisms. We flesh out this metaphor by presenting a series of parsing algorithms literally as inference rules, and by providing a uniform deduction engine, parameterized by such rules, that can be used to parse according to any of the associated algorithms. The inference rules for each logic will be represented as unit clauses and the fixed deduction procedure, which we provide a Prolog implementation of, will be a version of the usual bottom-up consequence closure operator for definite clauses. As we will show, this method directly yields dynamic-programming versions of standard top-down, bottom-up, and mixed-direction (Earley) parsing procedures. (p. 4)

They also go on to apply the PAD approach to syntactic formalisms other than CFGs, including combinatory categorial grammars (CCGs) and tree adjoining grammars (TAGs). Harkema (2001) and Hale (2003) apply the approach to Minimalist grammars. These richer formalisms (all slightly stronger than context-free) are widely seen as providing a more descriptively adequate treatment of natural language than CFGs. We shall see in the following chapter that principle-based grammars, such as Government and Binding Theory, likewise receive a natural interpretation in the PAD framework. Indeed, the PAD approach has been crucial in the development of principle-based parsers (Berwick 1991a, b; Johnson 1989, 1991).

From the point of view of a computational linguist, the PAD approach has an immediate payoff: it makes possible the application of well-known logic programming techniques to the task of parsing natural language. The programming languages *LISP* and *Prolog* have been adapted to natural language processing with great success. For *philosophical* purposes, the payoff is quite different, though no less intriguing. It is difficult to find up-and-running examples of psychological processes *convincingly* modeled as deductive operations defined over truth-evaluable statements. The PAD approach shows that such an example is available in the case of at least one aspect of language comprehension. We should not, however, overstate the claim. It is not as though neurocognitive research has revealed that the language inscribed in the brain is in fact first-order logic. Rather, what has been shown is that *if* the CYK or Earley algorithms play a role in a psychologically plausible model of the HSPM, as the results from Hale (2003) and others indicate, *then* there is a perfectly workable interpretation of the HSPM according to which the neural mechanisms that underpin it can be said to carry out deductive procedures. This substantiates my claim in Chap. 6 (Sect. 6.3.4) that the HSPM can exhibit a subpersonal analogue of inferential coherence.

The availability of an interpretation on which the HSPM is a deduction engine can be brought to bear on the recent debate in the philosophy of linguistics, concerning whether knowledge of language is, strictly-speaking, *propositional* (Knowles 2000; Rattan 2002). The fact that a rich array of successful parsing models can be represented in a familiar propositional format shows, at the very least, that it is *not incoherent* to claim that knowledge of language should be propositional. Nevertheless, I have argued (Chap. 7, Sect. 7.3) that it is better to regard the HSPM *subpersonal* system, which operates on representations that are not usefully assimilated to propositional attitudes.

8.4 The Oracle: Dealing with Inefficiency and Ambiguity

Both top-down and bottom-up parsers are, *in their purest and simplest forms*, wildly inefficient, hence unusable for realistic natural language processing, whether in machines or in humans. Devitt (2006a) draws on this point in the following argument against **REP-GRAM-DATA**—the idea that syntactic rules are explicitly represented and treated as data.

> How can the represented rules be used as data in language use? Consider language comprehension. Suppose that, somehow or other, the processing rules come up with a preliminary hypothesis about the structure of the input string. In principle, the represented rules might then play a role by determining whether this hypothesis could be correct (assuming that the input is indeed a sentence of the language). The problem in practice is that to play this role the input would have to be tested against the structural descriptions generated by the rules and *there are just too many descriptions*. The "search space" is just too vast for it to be plausible that this testing is really going on in language use. (209)

A closely related problem has to do with local structural ambiguities. Any parsing model that seeks to achieve psychological plausibility must somehow deal with the massive ambiguity of the input—a point that came up repeatedly in our discussion of the evidence for MPMs (Chap. 5). Efficiency problems and ambiguity problems can be seen as two sides of the same coin; after all, it is ambiguity—the availability of too many legal moves at any given point—that gives rise to inefficiency.

Computational linguists are all too aware of these problems and have devised an array of strategies for dealing with them. Above, we saw examples of clever computational tricks, like the storage and reuse of partial solutions.[26] But these can only take one so far. What's needed, at a minimum, is what Steedman (2000) calls "the oracle"—a mechanism that actively guides the parser's choice of rule-applications, and hence the resolution of ambiguities. Weinberg (1999) offers a helpful taxonomy of the various approaches to modeling human parsing preferences, and hence to characterizing the oracle that is inherent in the HSPM.

> Research in the theory of human sentence processing can be characterized by three styles of explanation. Researchers taking the first track have tried to motivate principles of structural preference from extralinguistic considerations like storage capacity in working memory, or bounds on complexity of incremental analysis. Frazier and Rayner's (1982) Minimal Attachment and Right Association [i.e., Late Closure] principles, and Gorrell's simplicity metric, are examples of this type of theory. The second track eschews "parsing strategies," replacing them with a fairly complex tuning by speaker/hearers to frequency in the hearer's linguistic environment. The difficulty of recovering an analysis of a construction in a particular case is a function of how often similar structures or thematic role arrays appear in the language as a whole. The work of Trueswell et al. (1994), Jurafsky (1996) and MacDonald et al. (1994) are examples of frequency or probability based constraint satisfaction theories. The third track takes a more representational view and ties processing principles to independently needed restrictions derived from competence and language learning. This approach claims that the natural language faculty is extremely well designed in the sense that the same set of principles that govern language learning also contribute to a theory of sentence processing. This track is represented by the work of Gibson (1981), Gorrell (1995) Pritchett (1992), Philips (1995, 1996) and Weinberg (1992), who argue that processing can be seen as the rapid incremental satisfaction of grammatical constraints such as the Theta Criterion, which are needed independently to explain language learning or language variation. (p. 283)

The grammar-based approach—the third on Weinberg's list—is, in my view, the most exciting of the three. It has the potential to unify our understanding of the structure-rules of natural language with both the parsing algorithm and the ambiguity-resolution principles of the HSPM. The structure-rules that comprise the grammar may, after all, be of the right sort to *also* govern language processing. Indeed, Weinberg (1999) argues that

> a particular version of the Minimalist Program (Chomsky 1993; Uriagereka 1999) provides principles needed to explain both initial human preferences for ambiguous structures and provides a theory of reanalysis, explaining when initial preferences can be revised given subsequent disconfirming data, and when they lead to unrevisable garden paths.

Now, if a deeply explanatory grammar could be made to do this double duty, accounting for *both* structural generalizations *and* human parsing preferences, this would constitute a powerful argument for **REP-GRAM-PROC** and hence for RT—a major contribution to the psychological reality debate. We must, however, postpone

[26] In principle-based parsers, to be discussed in the next chapter, "interleaving" the principles of a modular grammar such as GB leads to impressive gains in efficiency. See Berwick (1991a, b) and Merlo (1995).

discussion of this possibility to the next chapter, after we introduce the resources of more sophisticated grammars, such as GB and Minimalism.

In the remainder of this section, we explore the first two approaches in Weinberg's taxonomy: resource-based and frequency-based. Each of these enjoys certain distinctive advantages. The former can account for a wide range of human parsing preferences and other psycholinguistic data by appeal to a handful of intuitive principles (Sect. 8.4.1). But, though elegant, this approach does not have much to say about many of the inefficiency problems that arise in the course of constructing wide-coverage parsers for a realistic corpus. That task has been taken up by theorists in the statistical parsing tradition, which has come to dominate research in computational linguistics in recent years. Many in this area hold that the parsing principles MA, LC, and MCP are only coarse-grained generalizations, which can be shown to fail in some cases. Some theorists (e.g., Hale 2011) construct models that seek to *derive* the generalizations of MA, LC, and MCP from lower-level facts about how the parser uses frequency information to guide its decisions in real time.

A philosopher of science would want to know whether this is a case of two rival theories of the same subject matter, or a case of a higher-level theory being subsumed by an explanatorily deeper and ontologically more basic one. To answer this question, in Sect. 8.4.2 we examine the successes and limits of various statistical parsing models, with an eye toward determining the role of frequency information in human language processing.

8.4.1 Resource-Based Approaches

As we saw in Chap. 5, psycholinguistic research in the Garden-Path tradition (Frazier and Fodor 1978; Fodor and Frazier 1980; Fodor and Inoue 1994, 1998, 2000a, b) has shored up three general parsing principles: Minimal Attachment, Late Closure, and the Minimal Chain Principle.

(MA) The parser will attach incoming material into the existing mental phrase marker in such a way as to minimize the number of nonterminal nodes in the resulting structure.

(LC) The parser will incorporate newly encountered material into the most recent phrase or clause of the mental phrase marker that it has already constructed, thus preferring attachments lower in the tree over attachments higher in the tree.

(MCP) The parser will (i) avoid postulating unnecessary chain members (e.g., gaps/traces) and, (ii) having identified an item in the input as an antecedent, it will posit a gap/trace in the very first position at which it is licensed by the grammar.[27]

[27] This formulation is a paraphrase of the one found deVincenzi (1991). The second conjunct is equivalent to what Frazier and Clifton (1989) refer to as the "Active Filler Hypothesis." I use the locution 'gap/trace' to avoid commitment to formalisms that have traces and movement operations in their theoretical toolkit.

I will refer to MA, LC, and MCP, collectively, as the *least-effort parsing principles*, for they reflect a tendency on the part of the HSPM to adopt whichever syntactic analysis requires expending the fewest computational resources.

> In general, 'garden-path' explanations of processing difficulty account for asymmetries between sentences where pure memory load explanations do not. Our own model is of the garden path variety. The parser chooses to do whatever costs it the least effort; if this choice turns out to have been correct, the sentence will be relatively simple to parse, but if it should turn out to have been wrong, the sentence will need to be reparsed to arrive at the correct analysis (Frazier and Fodor 1978: p. 295).

Appealing to these principles has allowed psycholinguists to derive both clear predictions and satisfying explanations of a wide range of ambiguity resolution preferences—hence, of a great deal of behavioral and neurocognitive data (Chap. 5). It will come as no surprise, then, that versions of MA, LC, and MCP are thought by many psycholinguists to constitute the "oracle" of the HSPM.

It is important to be clear about the status of these principles vis-à-vis the psychological reality debate. To my knowledge, it has never been suggested that they are *represented* in the HSPM. Rather, these are intended to be *descriptive* principles; they are *true of* the HSPM in much the same way that the principles of celestial mechanics are *true of* our solar system. Moreover, while admitting the descriptive accuracy of such principles does commit one to the psychological reality of MPMs (Chap. 5), it does *not* automatically commit one to any view regarding the psychological reality of grammars—at least not without additional argument.[28]

To see more clearly the connection between these principles and the computational resources of a parser, consider the ways in which later formulations of Minimal Attachment depart from the original statements in Frazier and Fodor (1978) and Fodor and Frazier (1980). For instance, McRoy and Hirst (1990) write:

> MA has been defined as a preference for attachments that require adding the fewest nodes to attach a new word to the current structure; but … this definition is too dependent on assumptions about the underlying grammar. In addition, this definition can also make the system's account of the interactions among MA and other preferences implausibly rigid (Church 1980; Wanner 1980; Wilks et al. 1985) or force the system to delay attachment decisions (Shieber 1983), which is also implausible. … Thus, we use a more general statement of MA: The attachment that can be computed most quickly (i.e., with lowest time cost) is the one that will be favored. … [T]o find the most minimal attachment of a constituent C is to find (in decreasing order of preference): 1. A place along the right edge of the current parse tree where C is expected (or a place where C expects the current parse tree); 2. A place along the right edge of the current parse tree where C is allowed; 3. A place along the right edge of the parse tree where C could be attached by first proposing some intermediate structure such as a phrase or a clause. (323)

---

[28] J. D. Fodor and L. Frazier (1978, 1980) do supply an additional argument for RT. Fodor and Frazier claim that their parsing model provides a *principled* explanation of why MA and LC are true of the parser, precisely in virtue of the model's commitment to representing the grammar of a language in a separate data structure. They point out that the competing models, which implement the grammar "procedurally," can build in such principles only in an *ad hoc* way, if at all. From this, they conclude that it "seems unavoidable that *the well-formedness conditions on phrase markers are stored independently of the executive unit*, and are accessed by it as needed." (Frazier and Fodor 1978: 322n). We will return to this argument in Chap. 9.

This definition requires an independent specification of which computations have the "lowest time cost." McRoy and Hirst go on characterize the computational cost of an operation by listing a number of empirically confirmed generalizations:

> Lexical information is faster to identify than syntactic information, presumably because lexical retrieval is easier than tree traversal. Simpler attachments are faster to build than more complex ones, because it takes more time to build more structure. More local hypotheses are faster to recognize than more distant ones, because it takes more time to traverse more of a parse tree. Syntactic information is faster to access than semantic information, because tree traversal is easier than knowledge-base inference (cf. Ferreira 1986; Ferreira and Clifton 1986). Possibilities that have been primed, either by semantics or syntax, are faster to identify and build than unprimed ones (Frazer et al. 1984; Hirst 1987). Possibilities that are more highly expected are faster to identify than ones that are less expected. (322)

Although MA, LC, and MCP can account for a wide range of data, the principles have faced empirical challenges. Some of these have been effectively defused. For instance, the Late Closure principle encountered trouble when it was discovered that, in Spanish, Dutch, and perhaps other languages, adjuncts sometimes prefer to attach high in the tree structure, rather than low as Late Closure would predict (Cuetos and Mitchell 1988; Carreiras and Clifton 1993).

The sentence in (31) illustrates the problem.

(31) The spy shot the daughter of the colonel who was standing on the balcony.

LC predicts that readers will interpret (31) as meaning that it was the colonel who was standing on the balcony, not his daughter. But, contrary to this prediction, experiments showed that English speakers were equally happy with either interpretation, and that Spanish speakers preferred the daughter-on-balcony interpretation. This issue was effectively resolved by Fodor (1998a), who proposed that prosody—whether implicit or explicit—determines whether or not Late Closure applies. Roughly, the idea is that prosodically "heavy" material is attached high, while prosodically light material is attached low.

Still, other challenges remain, and proponents of alternative approaches to ambiguity resolution capitalize on their ability to predict data that the resource-based approach seems to miss. For instance, reading times on (32) appear to violate the prediction of Minimal Attachment (Weinberg 1999).

(32) John decided the contest was fair.

A proponent of the frequency-based view would explain this by pointing out that 'decide' occurs more frequently with sentential complements than with NP complements. Similarly, Hale (2011) presents the following pair, adapted from Gibson (1991).

(33) I gave her earrings on her birthday.
(34) I gave her earrings to another girl.

LC predicts that (34) should be fine, but that (33) should be a mild garden-path. Hale claims that readers perform equally well on both sentences. He then draws on his own probabilistic parsing model to explain the lack of preference.

Proponents of the frequency-based approach do not limit themselves to accounting for cases where the resource-based models seem to go wrong. Rather, they attempt to *derive* principles like MA, LC, and MCP from what they take to be more basic facts about the parser's sensitivity to statistical information. For instance, with regard to a particular type of garden-path effect, Hale (2011) says the following: "Whereas Frazier (1979) accounts for the effect with the Minimal Attachment (MA) heuristic, the present model recasts it as a side effect of optimistic expectations about the way sentences typically end" (p. 414). Here again, a question arises about what exactly competing models of the oracle seek to explain.

One option is this: Frequency-based and resource-based accounts agree that the MA heuristic is in fact the HSPM's usual strategy—i.e., that the HSPM actively strives to keep the number of nonterminal nodes to a minimum at each increment in the parse—and they seek to explain *this* fact, by *subsuming* MA under more basic, lower-level theories. In this case, MA is taken for granted, and two rivals accounts are proposed to explain why it is true. But another option is that frequency-based and resource-based accounts disagree about how to explain *the observed processing difficulties*—e.g., the behavioral and neurocognitive data that we sampled in Chap. 5. This formulation of the debate leaves it open that some frequency-based accounts will try to explain the data *without* committing themselves to any claim about the HSPM's minimization of nonterminal nodes, or, indeed, to any claim stated in the vocabulary of a hierarchical grammar. The get a clearer sense of this issue, let's turn our attention to frequency-based models, and see what bearing they have on the psychological reality issue.

8.4.2 Frequency-Based Approaches and Probabilistic Context-Free Grammars

Theorists who opt for the "probabilistic," "statistical," or "frequency-based" approach to parsing hold that the problems posed by ambiguity in linguistic input can be met by a system that takes into account lexical, syntactic, and semantic frequency information. Such information is typically derived from a corpus, such as the Brown Corpus, or from a treebank—most commonly, the Penn Treebank (Collins 1999; Bod et al. 2003; Jurafsky and Martin 2008: Chap. 14; Clark 2010; Hale 2011).[29] This view derives support from the compelling evidence that such frequencies are psychologically real and play an important role in lexical retrieval

[29] Some grammars are drawn from a hand-crafted corpus, which can be a fragment of natural language (Magerman 1995), or an artificial language generated by a hand-crafted grammar (Rohde 2002).

and parsing (MacDonald et al. 1994; Filip et al. 2002; Jurafsky 2003).[30] The core idea of this approach is that, when faced with an ambiguity, the HSPM will prefer the most probable analysis—i.e., the one that has occurred most frequently in prior input. In this way, processing difficulties can be explained by the system's low probability estimate of what turns out, in fact, to be the correct parse.

While it's clear that frequency information plays a significant role in comprehension, it's often hard to determine *which* frequencies matter—i.e., which factors the HSPM is keeping track of and which ones it brings to bear on a particular occasion. The space of logical possibilities is vast. On one end of the spectrum, frequencies can be defined over quite coarse-grained properties of sentences, like having an N-V-N structure. On the other end of the spectrum, we find suggestions that the HSPM keeps track of lexical frequencies defined over extremely fine-grained, recherché categories, cobbled together from semantic and pragmatic ingredients. To take a typical case, Filip et al. (2002) propose that frequency information about the distribution of "proto-Patient and proto-Agent properties" plays a role in language comprehension. They define these categories as follows:

> *Proto-Patient*: Undergo a change of state, serve as incremental themes, are causally affected by another event participant, are stationary relative to another event participant, and do not exist independently of the event described
>
> *Proto-Agent*: Volitional, sentient causers of the event described, who move and exist independently of the verb described (*sic*).

While frequencies defined over coarse-grained categories (N-V-N, etc.) are not terribly useful, those defined over fine-grained categories are known to cause "sparse data problems," as when a category is either very rare in the corpus or treebank, or missing entirely (Manning and Schütze 1999). The sheer number of fine-grained categories renders a system computationally inefficient, and gives it an *ad hoc* air.[31]

---

[30] Though see Weinberg (1999) for a discussion of some of the predictive failures of these models. I summarize Weinberg's argument in Chap. 9, Sect. 9.6.

[31] Manning and Schütze (2000) comment on this point, providing valuable methodological guidance to computational linguists: "It is not hard to induce some form of structure over a corpus of text. Any algorithm for making chunks—such as recognizing common subsequences—will produce some form of representation of sentences, which we might interpret as a phrase structure tree. However, most often the representations one finds bear little resemblance to the kind of phrase structure that is normally proposed in linguistics and NLP. Now, there is enough argument and disagreement within the field of syntax that one might find *someone* who has proposed syntactic structures similar to the ones that the grammar induction procedure which you have sweated over happens to produce. This can and has been taken as evidence for that model of syntactic structure. However, such an approach has more than a whiff of circularity to it. The structures found depend on the implicit inductive bias of the learning program. This suggests another tack. We need to get straight what structure we expect our model to find before we start building it. This suggests that we should begin by deciding what we want to do with parsed sentences. There are various possible goals: using syntactic structure as a first step towards semantic interpretation, detecting phrasal chunks for indexing in an IR system, or trying to build a probabilistic parser that outperforms n-gram models as a language model. For any of these tasks, the overall goal is to produce a system that can place a provably useful structure over arbitrary sentences, that is, to build a parser. For this

The tradeoff between informative categories and efficient computation is a primary concern for all research in probabilistic language processing.

In computational linguistics, one of the most common ways of making use of probabilistic information is to define frequencies over the tree structures of a context-free grammar. This yields an enriched grammar that contains not only various rewrite rules, but also estimates of how often the nonterminal nodes on the left-hand sides of each rule are expanded into the nodes on the right-hand side. The most naïve way of "inducing" a probabilistic context-free grammar (PCFG) from an annotated corpus is to literally count the number of times that each individual context-free rewrite rule is employed, or, equivalently, the number of times that particular syntactic structure appears in the input. (I make liberal use of this equivalence between tree-structures and derivations in what follows.) This provides a measure of how often a particular nonterminal is expanded in one way rather than another.

Far more complex ways of inducing a PCFG from a corpus have been devised, yielding correspondingly finer-grained and more useful statistical information (Charniak 1993, 1996, 1997; Bod 1998; Manning and Schütze 2000: Chap. 11; Clark 2010). Indeed, one of the major selling points of PCFGs is that they can in principle be learned from positive data alone, unlike ordinary CFGs.[32] Another desirable feature is their robustness in the face of the grammatical mistakes, disfluencies, and errors that pervade real text and speech. This problem can be avoided to some extent with a PCFG by ruling out nothing in the grammar, but by just giving implausible sentences a low probability. A toy example of a PCFG is given in Fig. 8.12. For realistic PCFGs, see Charniak (1993, 1996, 1997).

PCFGs have been employed in bottom-up, top-down, and left-corner parsers. To illustrate their use, consider how the CYK algorithm can use frequency information to compute a probability estimate of there being an NP, given the presence of a determiner and a noun, as well as the probability associated with the rule *NP* → *Det N*. Figure 8.13 provides the illustration.

Note that the probability estimates are based on purely structural factors, which don't take into account lexical co-occurrence. This renders PCFGs less good at predicting subsequent input than bigram and trigram models. But the two approaches can be combined.

> In practice, a PCFG is a worse language model for English than a n-gram model (for n > 1). An n-gram model takes some local lexical context into account, while a PCFG uses none. … PCFGs are not good models by themselves, but we could hope to combine the strengths of a PCFG and a trigram model. An early experiment that conditions the rules of a PCFG by word trigrams (and some additional context sensitive knowledge of the tree) is presented in Magerman and Marcus (1991) and Magerman and Weir (1992). (Manning and Schütze 2000: p. 387)

goal, there is no need to insist that one begins with a *tabula rasa*. If one just wants to do a good job at producing useful syntactic structure, one should use all the prior information that one has" (407–408).

[32] Here, "in principle" means "identification in the limit," where no bounds are placed on the amount of data the learning model is allowed to see.

Grammar		Lexicon
S → *NP VP*	[.80]	*Det* → *that* [.10] | *a* [.30] | *the* [.60]
S → *Aux NP VP*	[.15]	*Noun* → *book* [.10] | *flight* [.30]
S → *VP*	[.05]	| *meal* [.15] | *money* [.05]
NP → *Pronoun*	[.35]	| *flights* [.40] | *dinner* [.10]
NP → *Proper-Noun*	[.30]	*Verb* → *book* [.30] | *include* [.30]
NP → *Det Nominal*	[.20]	| *prefer*; [.40]
NP → *Nominal*	[.15]	*Pronoun* → *I* [.40] | *she* [.05]
Nominal → *Noun*	[.75]	| *me* [.15] | *you* [.40]
Nominal → *Nominal Noun*	[.20]	*Proper-Noun* → *Houston* [.60]
Nominal → *Nominal PP*	[.05]	| *NWA* [.40]
VP → *Verb*	[.35]	*Aux* → *does* [.60] | *can* [40]
VP → *Verb NP*	[.20]	*Preposition* → *from* [.30] | *to* [.30]
VP → *Verb NP PP*	[.10]	| *on* [.20] | *near* [.15]
VP → *Verb PP*	[.15]	| *through* [.05]
VP → *Verb NP NP*	[.05]	
VP → *VP PP*	[.15]	
PP → *Preposition NP*	[1.0]	

Fig. 8.12 A toy probabilistic context-free grammar (PCFG). The numbers in brackets denote the probability of applying a production rule—i.e., the frequency with which that rule has been applied in an annotated corpus, like the Penn Treebank. In a "probabilistically proper" PCFG, all probabilities for rules with the same left-hand side should sum to 1. For instance, the three rules for expanding the S node are 0.85, 0.10, and 0.05 (*Source*: Jurafsky and Martin 2008, pp. 465)

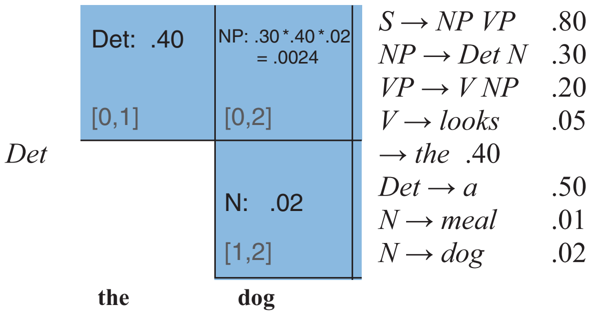


Fig. 8.13 The CYK algorithm applied to probabilistic CFGs. The input is the phrase 'the dog'. Here we see the parser computing a probability estimate of there being an NP in cell [0,2], on the basis of the presence of a determiner (Det) in cell [0,1], a noun in cell [1,2], and the rule *NP* → *Det N*. The probability of the determiner in cell [0,1] is 0.4 and the probability of the noun in cell [1,2] is 0.02. This is then multiplied by the probability of applying the *NP* → *Det N* rule, which is 0.3. The calculation 0.3 × 0.4 × 0.02 yields the value .0024 in cell [0,2]

A related enhancement of PCFGs involves "lexicalization"—i.e., the enrichment of the lexicon with frequency information specific to each lexical entry. This captures various lexical collocation probabilities and phrasal attachment probabilities, which are known to play a role in human parsing (MacDonald et al. 1994; Filip et al. 2002; Jurafsky 2003). In one version of this approach, when a lexical item is brought into a parse as the head of a phrase (the noun in an NP, the verb in a VP, etc.), the information in its lexical entry is passed upwards to the phrasal node (NP, VP….). This takes advantage of the strong probabilistic dependency relations between phrasal heads. But it can also lead a parser astray when the phrasal heads are not in fact related in the ways that the system supposes.

Another reason PCFGs are less predictive than *n*-gram models is that they embody a false assumption to the effect that the probabilities of various rule applications are context-free. This renders a PCFG-based parser blind to some obvious facts—e.g., that the probability of an NP expanding is different if the NP is in subject rather than object position. Note that this problem remains even if the lexicon is enriched in the ways suggested above. This points to a need for probabilistic systems that are sensitive to both surrounding words and syntactic relations.

> The upshot of these observations is that we should be able to build a much better probabilistic parser than one based on a PCFG by better taking into account lexical and structural context. The challenge (as so often) is to find factors that give us a lot of extra discrimination while not defeating us with a multiplicity of parameters that lead to sparse data problems. (Manning and Schütze 2000: p. 420)

Manning and Schütze go on to describe various efforts in this direction, including experimental work with History Based Grammars at IBM (Magerman 1994, 1995) and models in the Data-Oriented Parsing paradigm (Bod 1998).

Computational linguists use probabilistic models in at least three different ways: (i) to guide lexical retrieval—the identification of input words in a noisy stimulus stream; (ii) to improve incremental parsing by removing some of the paths at any given choice point; and (iii) to rank the fully completed parses that come in from a base-grammar analysis. As Clark (2010) observes, this third use entails an enrichment of the tripartite "anatomy" with which we began this chapter. He notes that any comprehension system must answer four types of question:

> (1) What is the grammar which defines the set of legal syntactic structures for a sentence? … (2) What is the algorithm for determining the set of legal parses for a sentence? What data structure is used to represent those parses? (3) What is the model for determining the plausibility of different parses for a sentence? (4) What is the algorithm, given the model and a set of possible parses, which finds the best parse (or the *n* best parses)? The first three questions correspond roughly to the characterization of parsing in Steedman (2000) (although Steedman uses the term 'oracle' for model). The additional fourth component is the decoder. (Clark 2010: 333)

One might have thought that the "decoder" is either trivial or redundant, given the presence of the model/oracle. But, as Clark points out, that isn't so.

> If the possible parses can be efficiently enumerated, then there is a trivial decoding algorithm: simply loop through each parse calculating its score, and return the highest-scoring one. However, as we shall see, the grammars used by statistical parsers are often wide-coverage, producing many parses for some sentences, far too many to enumerate. In this case, a more sophisticated representation of the possible parses, and decoding algorithm, is required. (Clark 2010: 333)

As already noted, when seen as a search procedure, parsing is a matter of navigating a vast search space of phrase structure trees, which can itself be thought of as a branching tree in which legal applications of rules constitute the branch points.

Such a search can be accomplished by what is known as a *stack decoding algorithm*. Here, partial legal parse trees are entered into a priority queue, from which they are retrieved in order of probability, and expanded by applications of the available grammar rules. The results are then entered back into the queue. The process repeats until the queue contains what the system takes to be an optimal parse.

Clark's point in the above quotation is that if we could enumerate all of the parses and their probabilities in a priority queue, then we would be guaranteed the best possible parse. But given the vastness of the search space, this is impossible. To deal with this problem, probabilistic parsing systems use an effective heuristic known as *beam search*, which expands only the *best* partial trees. (We will see definitions of "best" in a moment.) The size of the "beam" is a measure of how many options are pruned at each step; the narrower the beam the more aggressive the algorithm is at pruning the branches in the queue before expanding the remaining ones. "A beam may either be fixed size, or keep all results whose goodness is within a factor α of the goodness of the best item in the beam" (Manning and Schütze 2000: p. 441).

There are, broadly-speaking, three approaches to conducting a beam search: One is called "uniform-cost search," where one uses a PCFG to work out the probability of the steps taken thus far—or, equivalently, the probability of the current partial parse—and then selects the highest-ranked ones for expansion at the next ply. Computing these probabilities is relatively easy, but this approach is blind to how near its selections are to a complete solution. A second approach, "best-first search," selects derivations on the basis of an estimate of how close they are to a solution.

> [H]ow many steps should be expected to finish parsing from search node n? … Perhaps the most straightforward answer to this question would be to say that estimates about how sentences are going to end should be based on experience. This idea can be formalized by simulating the action of a given parsing strategy on a corpus sample and recording how far away particular states were from actually finishing. (Hale 2011: p. 411)

The third approach, known as A* search (Hart et al. 1968), combines these two methods, yielding an optimal method that selects derivations on the basis of *both* the steps already taken *and* the work still left to do.

> Working out the probability of the steps already taken is easy. The tricky part is working out the probability of the work still to do. It turns out that the right thing to do is to choose an optimistic estimate, meaning that the probability estimate for the steps still to be taken is always equal to or higher than the actual cost will turn out to be. If we can do that, it can be shown that the resulting search algorithm is still complete and optimal. Search methods that work in this way are called A* search algorithms. A* search algorithms are much more efficient because they direct the parser towards the partial derivations that look nearest to leading to a complete derivation. Indeed, A* search is optimally efficient meaning that no other optimal algorithm can be guaranteed to explore less of the search space. (Manning and Schütze 2000: p. 441-2)

Picking up on this notion of an "optimistic estimate," recall the passage from Hale (2011), quoted in the previous section, in which he recasts what appears to be an effect of Minimal Attachment (MA) as "a side effect of optimistic expectations about the way sentences typically end" (414). We can now see that the optimistic expectations Hale has in mind are precisely those driving the A* search that constitutes a key feature of his model. There is a worry, though, that the coincidence between the effects of MA and of A* search is merely superficial, or confined to only short sentences. As Manning and Schütze (2000) note,

> PCFGs have certain biases, which may not be appropriate. All else being equal, in a PCFG, the probability of a smaller tree is greater than a larger tree. This is not totally wild—it is consonant with Frazier (1978) Minimal Attachment heuristic—but it does not give a sensible model of actual sentences, which peak in frequency at some intermediate length. For instance, … the most frequent length for Street Journal sentences is around 23 words. A PCFG gives too much of the probability mass to very short sentences. (387)

Hale is alive to this worry, and addresses it by suppressing the effects of the probability of the current parse (which he labels 'g(n)' in the passage quoted below), to see if the MA-like effect—i.e., the "same pattern of search behavior"—persists. Observing that it does, Hale concludes that pattern is due at least in part to the probability estimate of the work left to be done (labeled 'h(n)').

> A critic might argue that the proposed comprehension model is a notational variant of Frazier's (1979) Garden Path model in the sense that g(n) prioritizes the exploration of phrase structures that have fewer nodes first. By setting the step-costs identically to zero, the role of g(n) can be suppressed. The fact that the same pattern of search behavior still arises confirms that h(n) has an important role too. This term of the A* heuristic value equation has no parallel in Garden Path theory. This demonstration rules out the possibility that the proposal is a notation variant of the Garden Path model. (Hale 2011: p. 436, fn. 6)

Hale's talk of "pattern of search behavior" tells us that his explanatory target is not merely the available behavioral or neurocognitive data concerning attachment preferences and processing difficulties. Hale seeks to explain why the HSPM traverses the grammatically-conditioned search space in a way that, by and large, minimizes the number of nonterminal nodes in each of the structures it considers. This answers our earlier question: Hale's account is an attempt to subsume a high-level generalization, MA, cast in the vocabulary of a hierarchical grammar, under a more basic theory about the determinants of the parser's decisions. Other theorists have seen heuristics and statistics as *complementary* aspects of a model, not in competition with one another.

> [S]ome of the discovered features of the human parser have already been exploited in the NLP system implementation. For example, a number of preference heuristics (like Late Closure (Kimball 1973) or Minimal Attachment (Frazier & Fodor 1978)), devised in psycholinguistic environments to describe some aspects of the human parser, have been implemented to solve local ambiguities (see, e.g., (Hobbs & Bear, 1990)), and, after the emergence of the corpus-based approaches, NLP systems often involve a combination of heuristics and statistics to guide the parser decisions. (Costa et al. 2000: 1)

Of course, there are also many probabilistic parsers in computational linguistics that are designed for practical applications in commercial settings—e.g., the SPANNER parser at IBM (Magerman 1995). The goal of these systems is not to model detailed aspects of human parsing performance, such as attachment preferences or patterns of processing difficulty. Insofar as they seek to emulate human parsing, it is only in respect of the sheer speed, general accuracy, and width of coverage. Thus, when a computational linguist who favors nonhierarchical dependency grammars argues that "much of the superstructure of a phrase structure tree is otiose: it is not really needed for constructing an understanding of sentences" (Manning and Schütze 2000: p. 430), it may well be that fine-grained psycholinguistic data are not her concern, and that psychological plausibility is no part of her objective.

This of course raises the question: Do probabilistic parsers have a strong claim to psychological plausibility? Brants and Crocker (2000a, b) review a number of experiments designed to "counter… the possible criticism that probabilistic parsers are too powerful and resource intensive to be considered as the basis of a cognitively plausible model" (2000b: p. 666). They conclude that

> wide-coverage, probabilistic parsers do not suffer impaired accuracy when subject to strict cognitive memory limitations and incremental processing. Furthermore, parse times are substantially reduced. This suggests that it may be fruitful to pursue the use of these models within computational psycholinguistics, where it is necessary to explain not only the relatively rare 'pathologies' of the human parser, but also its more frequently observed accuracy and robustness. (2000a: p. 117)

To take just one example, Hale (2001, 2003) presents evidence for the psychological plausibility of an Earley parser that draws on an internally represented PCFG, showing how such a model predicts two kinds of processing difficulties that the HSPM is known to exhibit. The first has to do with the main-clause/relative-clause ambiguity:

(35) The horse raced past the barn fell.
(36) The ship floated down the river sank.
(37) The dealer sold the forgeries complained.
(38) The man sent the letter cried.

The second processing difficulty has to do with the asymmetry between subjects and objects in unreduced relative clauses. The examples below are taken from Gibson (1998).[33]

(39) [*S* The reporter [*S*' who [*S* the senator attacked]] admitted the error].
(40) [*S* The reporter [*S*' who [*S* attacked the senator]] admitted the error].

[33] Gibson cites the results of a number of psycholinguistic experiments that establish the reality of this processing difficulty: "The object extraction is more complex by a number of measures including phoneme monitoring, on-line lexical decision, reading times, and response-accuracy to probe questions (Holmes 1973; Hakes et al. 1976; Wanner and Maratsos,1978; Holmes and O'Regan 1981; Ford 1983; Waters et al. 1987; King and Just 1991). In addition, the volume of blood flow in the brain is greater in language areas for object-extractions than for subject-extractions (Just et al. 1996a, b; Stromswold et al. 1996), and aphasic stroke patients cannot reliably answer comprehension questions about object-extracted RCs, although they perform well on subject-extracted RCs (Caramazza and Zurif 1976; Caplan and Futter 1986; Grodzinsky 1989; Hickok et al. 1993)" (p. 2).

The kind of explanation that Hale provides is importantly different from that the Garden-Path model, which appeals to the least-effort parsing principles, MA, LC, and MCP. The latter is committed to a *serial* processing model, on which the HSPM incorporates new input incrementally by constructing only one parse tree at each successive step. The processing difficulties associated with garden-path sentences are then accounted for by the computational costs of a *reanalysis* operation. By contrast, Hale (2001, 2003) assumes that the HSPM is a *parallel* parser, which constructs *multiple* parses at each successive step—one for each possibility consistent with the grammar (above a given probabilistic threshold). These are then ranked in terms of their probability of success. Processing difficulties are explained, on a parallel model, by appeal to a computationally costly operation of *re-ranking* these analyses. The lower the probability of an analysis at step n, the more costly it is to rank that analysis higher upon encountering new input at step $n + 1$.[34] For instance, when the word 'fell' comes in after the first several words of sentence (35), what initially seemed like a highly probable analysis must be demoted, while what initially seemed like highly improbable analysis must be ranked highest. Hale shows that it is precisely at this point that an Earley parser operating with a PCFG encounters serious difficulty, thereby simulating human performance.[35]

Hale (2011) continues this work, incorporating both A* search and intelligent left-corner parsing strategies. We noted above (Sect. 8.3.3) that left-corner parsers enjoy a variety of benefits over purely top-down and bottom-up parsers. Their advantages are even more striking in the context of probabilistic systems. To begin with, there is a unique left-corner derivation for any given tree, which allows for a "probabilistically proper" language model—i.e., one in which the probabilities for all of the rules with the same left-hand side sum to 1. Perhaps more importantly, the behavior of the parser at any given step is conditioned by two sources of probabilistic information: (i) the probability of the current parse, and (ii) the local goal. This takes into account the fact that "the probability of a certain expansion of NP can be different in subject position and object position, because the goal category is different" (Manning and Schütze 2000: 427). Manning and Carpenter (1997) conducted an early exploration of probabilistic left-corner grammars (PLCGs) and found that they significantly outperform parsers that operate with a basic PCFG. Likewise, Roark and Johnson (1999) write:

> We have examined several probabilistic predictive parser variations, and have shown the approach in general to be a viable one, both in terms of the quality of the parses, and the efficiency with which they are found. We have shown that the improvement of the grammars with non-local information not only results in better parses, but guides the parser to them much more efficiently, in contrast to dynamic programming methods. Finally, we have shown that the accuracy improvement that has been demonstrated with left-corner

[34] Resolving the dispute between advocates of serial and parallel models would clearly yield a deeper understanding of the HSPM. For discussion, see Crocker et al. (2000), and references therein.

[35] Of course, a serial parser can likewise make use of statistical information contained in PCFGs, e.g., to determine which rule to apply or which lexical item to select in cases of ambiguity. Jurafsky and Martin (2008: Chap. 14).

> approaches can be attributed to the non-local information utilized by the method. This is relevant to the study of the human sentence processing mechanism insofar as it demonstrates that it is possible to have a model which makes explicit the syntactic relationships between items in the input incrementally, while still scaling up to broad-coverage. (p. 426)

In the models developed by Hale (2001, 2003), the rules of a grammar are explicitly represented—i.e., stored in a separate data structure and consulted as data, in accordance with the Earley algorithm. This is in line with Hale's commitment to what he calls the Strong Competence Hypothesis:

> What is the relation between a person's knowledge of grammar and that same person's application of that knowledge in perceiving syntactic structure? The answer to be proposed here observes three principles. Principle 1 *The relation between the parser and grammar is one of strong competence. Strong competence holds that the human sentence processing mechanism directly uses rules of grammar in its operation, and that a bare minimum of extragrammatical machinery is necessary.* This hypothesis, originally proposed by Chomsky (Chomsky 1965, page 9) has been pursued by many researchers (Bresnan 1982) (Stabler 1991) (Steedman 1992) (Shieber and Johnson 1993), and stands in contrast with an approach directed towards the discovery of autonomous principles unique to the processing mechanism.

However, as noted earlier, even though the Earley algorithm is implemented this way in conventional stored-program computers, there is no reason why the algorithm could not be "hardwired," in which case the grammar would be merely embodied, without being represented.

Thus, while Hale's model appears at first to lend support to **REP-GRAM-DATA**, it in fact leaves open whether the HSPM's "direct use" of the rules of a grammar is a matter of explicitly *representing* those rules or *embodying* them in a procedural fashion. Indeed, some of the remarks in Hale (2011) are strongly suggestive of **EMB-GRAM-PROC**.

> On the top-down strategy, to 'use' 'apply' or equivalently 'announce' a rule means to exchange a prediction for one category, e.g., the left-hand side symbol VP, for a sequence of other categories such as V and NP. This predictive aspect of top-down parsing makes it an attractive to cognitive modelers since there is evidence for these sorts of anticipatory representations … The alternative to top-down parsing is bottom-up parsing. A bottom-up parser exchanges evidence about a sequence of daughter categories for a conclusion about a parent category. (404)

Still, I think that the psychological plausibility of Hale's models, and others like them, provides reason to think that **EMB-GRAM-PROC** is the *weakest* position consistent with we already know about human parsing; the weaker **GRAM-CONFORM** leaves mysterious the range of data that such models explain.

More generally, as the discussion above makes clear (and as Fig. 8.13 illustrates), a parser whose oracle draws on frequency information must in one way or another incorporate a grammar. Thus, issues pertaining to the psychological reality of grammatical rules and principles arise regardless of whether one opts for the resource-based or the frequency-based approach to ambiguity resolution. This point is sometimes obscured by the fact that computational linguists working in the statistical parsing tradition—arguably the dominant approach these days—are often much less concerned with what grammar generates the search space of phrase markers than with the statistical techniques that serve to prune that space.

> … a statistically-minded linguist will not be much interested in how many parses his parser produces for a sentence. Normally there is still some categorical base to the grammar and so there is a fixed finite number of parses, but statistically-minded linguists can afford to be quite licentious about what they allow into their grammar, and so they usually are. What is important is the probability distribution over the parses generated by the grammar. (Manning and Schütze 2000: 411–12)

As such, they often choose grammars on grounds that have little to do with psychological plausibility or faithfulness to the latest advances in formal syntax.

The grammars in such systems are typically impoverished in their descriptive resources, often lacking basic theoretical constructs, such as gaps/traces or the complement/adjunct distinction. (Collins 1999 is an exception.) Indeed, some of the most popular grammars employed in this area look very little like anything a generative linguist would endorse—e.g., the dependency grammars mentioned earlier, which eschew hierarchical constituency structures in favor of dependency relations between individual words.[36] This difference is actually superficial, in view of the isomorphisms that hold between various kinds of dependency grammar and corresponding types of phrase structure grammar (Manning and Schütze 2000: 429–30). But even such superficial differences can encourage the incorrect view that statistical parsing doesn't "really" require a grammar. In truth, the only statistical parsing techniques for which this strikingly strong claim would be plausible are ones that deal *solely* in *n*-gram probabilities—co-occurrence frequencies defined over linear strings of *n* words at a time. Given that the HSPM almost certainly builds mental phrase markers (Chaps. 5 and 6), such models have no psychological plausibility, however useful they may be in commercial applications.

Thus, for our purposes here, the most interesting question is not whether the HSPM defines its statistical generalizations over the rules, principles, or structures of some grammar—this is virtually guaranteed—but whether that grammar receives independent motivation from formal linguistics.[37] To address this question, I devote the next chapter to historical tour through the evolution of grammars in the generative tradition, starting with transformational grammars and gradually moving into the Principles & Parameters framework, which encompasses Government and Binding Theory and Minimalism.

[36] Though, see Clark (2010) for a probabilistic parser that makes use of the sophisticated Combinatory Categorial Grammar (CCG). It seems likely that probabilistic extensions of such sophisticated grammars will emerge in the coming years.

[37] An interesting wrinkle: Manning (2003) argues that formal linguistics should itself take the statistical turn, so to speak, and cast its grammars in a probabilistic formalism. Thus, one way of securing a tight relation between the linguist's descriptive grammar and the psycholinguist's mental grammar is to adjust the former, not the latter.

Chapter 9
The Psychological Reality of Syntactic Principles

Abstract In this chapter, I survey a variety of grammars that have played a role in psycholinguistics, tracing the coevolution of theories in formal syntax and the computational parsing models that they inspired. In Chomsky's "Standard Theory" the output of context-free rules is fed into the transformational component of a grammar. Many incorrectly interpreted early psycholinguistic experiments as shedding doubt on the psychological reality of transformational operations. These arguments, based on the Derivational Theory of Complexity, ultimately fail. But transformational parsers were rejected anyway, on computational grounds. Augmented Transition Networks (ATNs) rose to prominence, offering a promising framework for describing the surface syntax of natural language, as well as a natural implementation of the grammar as a parsing model. ATN parsers thus serve as a clear example of how grammatical rules can be viewed as procedural dispositions. A strong criticism of the ATN architecture, due to Lyn Frazier and Janet Fodor, relied heavily on the fact that ATNs do not explicitly represent the rules of a grammar in a separate data structure. Frazier and Fodor's argument faces difficulties, but it vividly illustrates the kind of explanatory payoff that a model might derive from explicitly representing a grammar. Finally, I turn to principle-based parsers, which implement the principles of the Government and Binding (GB) theory as either generators or filters of syntactic analyses, yielding compact, efficient, wide-coverage systems. More recently, computational linguists have built parsers that use the syntactic principles of the Minimalist program. Indeed, Amy Weinberg has argued that parsing is "the incremental satisfaction of grammatical constraints" imposed by Minimalist grammars. If successful, her proposal would constitute the strongest argument for the psychological reality of Minimalist principles.

Keywords Standard theory • Context-free grammar (CFG) • Transformational grammar • Transformational rules • Backwards transformations • S-structure • D-structure • Derivational Theory of Complexity (DTC) • Augmented Transition Network (ATN) • Representation • Computational efficiency • Principles-and-Parameters (P&P) • Government and Binding theory (GB) • Principle-based parsing • Ambiguity resolution • Mental syntactic principles (MSPs) • Minimalism • Minimalist grammar • The Minimalist program • Joan Bresnan • Lexical-functional grammar (LFG) • Precomputed lexical realization • Chronometric data • Incremental parsing • Covering grammar • Strong Competence Hypothesis • Feature unification

D. Pereplyotchik, *Psychosyntax*, Philosophical Studies Series 129,
DOI 10.1007/978-3-319-60066-6_9

• Lexicalist grammars • The Chomsky hierarchy • Canonical sentoid strategy • Fodor Bever and Garrett (FBG) • Heuristics-based parsing • Procedural embodiment • The Sausage Machine model • Preliminary phrase packager (PPP) • Sentence structure supervisor (SSS) • Chunking • X-bar theory • Binding theory • Case assignment • The empty category principle • Theta criterion • Theta-roles • Case filter • Command • Frequency-based parsing • Economy conditions • Merge • Move • Spell-Out • Lexical features • Feature checking • Robert Berwick • Henk Harkema • Amy Weinberg

9.1 Introduction

The comprehension models that psycholinguists and computational linguists have developed can differ along a number of dimensions. One key difference has to do with what type of grammar a given model employs. In the previous chapter, we considered context-free grammars and their probabilistic extensions. In this chapter, we survey a wider variety of grammars, focusing on those that have played a prominent role in psycholinguistics: transformational grammars, ATN grammars, Government and Binding Theory, and Minimalism. The main goal is, of course, to assess claims about the psychological reality of any such grammar. But a subsidiary goal is to contribute to the history of science by tracing the co-evolution of theories in formal linguistics and the parsing models that they inspired in computational psychology. To further both ends, I provide introductions to the basic theoretical machinery of each of the abovementioned frameworks. My discussion of their descriptive and explanatory resources will, perforce, be rather brief. It is not my aim to establish the superiority of one or another approach to grammar. My task here is to illustrate the main distinctions between them, and to evaluate the prospects of including any of them in a realistic account of human language processing.

Context-free rules form the base of early versions of transformational grammar. In the Standard Theory (Chomsky 1957), the output of these rules is fed into a component of the grammar consisting of *transformational* rules. After introducing transformations in Sect. 9.2, I discuss the experiments that Fodor et al. (1974) interpreted as shedding doubt on the psychological reality of transformational operations. Although Fodor, Bever, and Garrett's argument turns out to be unsuccessful, we will see that parsing models based on the Standard Theory were found to be unworkable anyway, though on purely computational grounds.

In the wake of these findings, the framework of Augmented Transition Networks (ATNs) rose to prominence in the 1970s. The ATN formalism offered a promising alternative for describing the surface syntax of natural language. Moreover, it suggested a natural implementation of the grammar as a parsing model—a top-down search procedure over the space of syntactically possible structures. In Sect. 9.3, we will examine ATN parsers, which serve as a clear example of how grammatical rules can be seen as procedural dispositions—represented or embodied. Section 9.3 concludes with a discussion of the psychological plausibility of ATN parsers. A strong criticism of the ATN architecture, due to Lyn Frazier and Janet Dean Fodor, relies heavily on the fact that ATNs do not explicitly represent the rules of a grammar in a separate data structure. Frazier and Fodor's argument faces difficulties, but it vividly illustrates the kind of explanatory payoff that a model might derive from explicitly representing a grammar.

In Sect. 9.4, I turn to principle-based parsers—models whose internal operations implement the principles of Government and Binding (GB) theory. Principle-based models constitute a particularly exciting application of the Parsing as Deduction (PAD) approach discussed in the previous chapter. Here, GB principles, are formalized as either generators of, or as logical constraints on, possible syntactic analyses. Modern principle-based parsers are sophisticated wide-coverage systems that offer a number of advantages over traditional rule-based systems, in respect of compactness, modularity, and efficiency.

In recent years, the Principles and Parameters approach has moved away from the GB formalism. Seeking a yet more elegant and compact account of syntactic structure, Chomsky (1995, 2001) and others have developed a variety of Minimalist grammars, in which the basic operations of Merge and Move serve to forge hierarchical relations between lexical items, which are themselves conceived of as sets of features. Unsurprisingly, computational linguists have undertaken the task of building parsing models that incorporate these grammars. In Sect. 9.5, I examine one such effort, focusing on the algorithms developed by Harkema (2001). Harkema employs the PAD approach in constructing bottom-up, top-down, and mixed-direction algorithms for parsing with Minimalist grammars. Finally, I examine the proposal of Weinberg (1999), which attempts to account for a number of psycholinguistic results by viewing parsing as "the incremental satisfaction of grammatical constraints" imposed by Minimalist grammars. If successful, Weinberg's proposal would constitute the strongest argument for **REP-GRAM-PROC**. We would, that is, have excellent grounds for maintaining that Minimalist principles are explicitly represented in the minds/brains of competent language users. I conclude, however, that the more parsimonious position, **EMB-GRAM-PROC**, cannot be ruled out.

9.2 Transformational Grammar

In an effort to explain a variety of linguistic generalizations that context-free grammars can only stipulate, Chomsky (1957) introduced transformational grammars, which operate at two distinct levels of description. At the first level, a syntactic structure is assigned to a string of words in accordance with, e.g., a context-free grammar.[1] Following standard nomenclature, I will use the label "D-structure" to

[1] Other possibilities are certainly imaginable. See Chomsky (1965) for a proposal to enrich the base of the grammar.

denote the level at which these base rules yield a set of syntactic descriptions. At the second level, transformational rules are brought to bear, mapping these D-structure descriptions into what Chomsky called "S-structure" descriptions. (Informally, 'D' is a mnemonic for 'deep' and 'S' for 'surface'.) Transformational rules thus serve to map entire syntactic structures into *other* syntactic structures. For instance, the sentences in (2)–(8) can be generated by an application of rules (9)–(15) to the syntactic structure of the active declarative sentence in (1).[2]

(1)	**Morris solved the problem.**	**Active**
(2)	**The problem was solved by Morris.**	**Passive**
(3)	**Morris didn't solve the problem.**	**Negative**
(4)	**Did Morris solve the problem?**	**Question**
(5)	**Was the problem solved by Morris?**	**Passive question**
(6)	**The problem wasn't solved by Morris.**	**Negative passive**
(7)	**Didn't Morris solve the problem?**	**Negative question**
(8)	**Wasn't the problem solved by Morris?**	**Negative passive question**

(9)	**NP_1 V + ed NP_2 ⇔ NP_2 was V + ed by NP_1**	**Passivization**
(10)	**NP_1 V + ed NP_2 ⇔ NP_1 didn't V NP_2**	**Negation**
(11)	**NP_1 V + ed NP_2 ⇔ Did NP_1 V NP_2**	**Question-formation**
(12)	**NP_1 V + ed NP_2 ⇔ was NP_2 V + ed by NP_1**	**Passive question-formation**
(13)	**NP_1 V + ed NP_2 ⇔ NP_2 wasn't V + ed by NP_1**	**Passive negation**
(14)	**NP_1 V + ed NP_2 ⇔ Didn't NP_1 V NP_2**	**Negative question-formation**
(15)	**NP_1 V + ed NP_2 ⇔ wasn't NP_2 V + ed by NP_1**	**Negative passive question-formation**

The transformations in (9)–(15) map active declarative structures into a variety of other sentence structures, and vice versa, thereby capturing the systematic relations between them. They also illustrate how various transformations can yield novel syntactic structures on the basis of a kernel structure—in this case, the active declarative structure of (1). Using these transformations as generators of novel structures, we can even eliminate from our grammar all of the context-free rules that generate the structures that appear on the right-hand sides of (9)–(15). Once the context-free rules assign a structure to some kernel sentence, the remaining structures can be generated directly by the transformation rules, without appeal to the context-free

[2] The examples in (1)–(15), adapted from Fodor, Bever, and Garrett (1974: pp. 97–98). The discussion here is oversimplified in an important respect: Chomsky did not posit separate transformations for each linguistic phenomenon. Some were dealt with by the ordered, serial application of several transformational rules. For instance, negative questions would have undergone two transformations: Negation and Question-formation. We return to this point below.

rules that form the "base" of the grammar. This eliminates redundancy, making for a more compact formalism.

The symbol '⇔', and the attendant notion of a "mapping," should be treated with care. The transformation rules, *qua* structure rules, embody claims about the relations between sentences. The rule in (9), for instance, is a claim to the effect that if a sentence has the syntactic structure NP_1 *V + ed* NP_2—as dictated by the underlying context-free rewrite rules—then there is another sentence in the language that has the syntactic structure NP_2 *was V + ed by* NP_1.[3] As always, without additional argument, no part of this explanatory framework entails anything about psychological or computational operations.

Nor do the compactness or explanatory power of transformational grammars count as reasons for ascribing to them any psychological reality (Chaps. 2, 3, and 4). Certainly, from the point of view of a syntactician, compactness is a desirable property. But, as Stabler (1984) points out, compactness is not automatically a desideratum for the psycholinguist, nor even for a computational parsing theorist. Moreover, the mere fact that transformations explain more linguistic generalizations than simple context-free grammars do is not *by itself* a reason for asserting that transformations are psychologically real operations. There has been some confusion on this point. For instance, Wanner and Maratsos (1978) interpret Chomsky as denying this claim.

> Chomsky has consistently assumed that the psychologically real representation of linguistic knowledge will be just that grammar that captures linguistically significant generalizations; according to Chomsky, 'we have [such] a generalization where a set of rules about distinct items can be replaced by a single rule about the whole set…' (Chomsky 1965: p. 42). This assumption is fundamental to Chomsky's claim for the psychological reality of transformational rules; it is by accumulating clear cases in which the transformational notation permits a set of nontransformational rules to be replaced by a single transformation that Chomsky builds a psychological case for transformations. (p. 158)

In defense of these authors, it must be noted that the corpus of Chomsky's work in the 1960s and 1970s provides somewhat contradictory clues about his commitments on the psychological reality issue (Chap. 2; see also Devitt 2006a: ch. 4). Let's now look at the kinds of empirical results that *are* relevant to that topic.

9.2.1 The Psychological Reality of Transformations and the Derivational Theory of Complexity

In a highly influential discussion, Fodor et al. (1974) reviewed a number of psycholinguistic studies that they interpreted as evidence *against* the psychological reality of transformational operations. The model that they contemplated construed the

[3] In the examples above, the sentences share many of their lexical items, but differ in meaning. Fodor, Bever, and Garrett (1974: ch. 3) discuss restrictions on the relationship between transformationally related sentences. I pass over these complications here.

process of syntactic analysis as involving two phases. In the first phase, the HSPM computes an S-structure representation of the linguistic input. In the second phase, this representation is transformed into a D-structure representation, from which semantic information would then be recovered. The computation of the D-structure representation was to be accomplished by the *reverse* application of transformational rules. We will refer to these as "backwards transformations".

The studies reviewed by Fodor, Bever, and Garrett (henceforth, FBG) were premised on the assumption, inherent in the *Standard Theory* of transformational grammar (Chomsky 1957, 1965), that transformations must apply in a particular order. If this is correct, then the reverse process should proceed in a certain order as well**.** For instance, negative passive questions would have to go through more backwards transformations than negative declarative sentences in the course of comprehension. Similarly, passives would have to go through one more backwards transformation than actives. (See FBG, pp. 228 for details.)

FBG's discussion takes aim at what they called the *Derivational Theory of Complexity* (DTC). According to the DTC, the transformations of the Standard Theory are psychologically real procedural operations. That is, the HSPM is seen as literally transforming syntactic representations (MPMs) in the ways that the rules specify. The DTC is often attributed to Miller and Chomsky (1963). Berwick and Weinberg (1984) claim that Miller and Chomsky "identified rules of the grammar with computational operations of the parser in a one-to-one fashion," and that "[t]his identification led to specific behavioral predictions, collapsing grammatical with processing complexity" (pp. 38–9). They quote the following telling passage from Miller and Chomsky (1963):

> The psychological plausibility of a transformational model of the language user would be strengthened, of course, if it could be shown that our performance on tasks requiring the appreciation of the structure of transformed sentences is some function of the nature, number, and complexity of the grammatical transformations involved" (p. 481).

Berwick and Weinberg (1984) go on to state the core idea in a way that avoids commitment to Chomsky's Standard Theory and, more generally, to any specific grammar, transformational or otherwise. Promoting it to the status of a methodological principle, they write:

> Miller and Chomsky's original (1963) suggestion is really that grammars be realized more or less directly as parsing algorithms. We might take this as a methodological principle. In this case, we impose the condition that the logical organization of rules and structures incorporated in a grammar be mirrored rather exactly in the organization of the parsing mechanism. We will call this *type transparency* (1984: p. 39).

This principle goes by other names; many have called it the *strong competence hypothesis* (Bresnan and Kaplan 1982; Steedman 2000; Hale 2003, 2011), the label I'll adopt here. The strong competence hypothesis is at the core of the positions we have called **REP-GRAM-PROC** and **EMB-GRAM-PROC**. As many authors have noted (e.g., Phillips 1996), the failure of the DTC would not automatically translate into an argument against the strong competence hypothesis. The latter is logically weaker than the former. Given that the DTC concerns a particular type of transformational grammar, and presupposes a particular kind of parsing architecture, adopting a different grammar or parsing architecture can preserve the more general strong competence hypothesis. That being said, the history of the controversy surrounding the DTC should serve as a cautionary tale for anyone seeking to confirm or disconfirm any claim about the psychological reality of a grammar. It is in many ways a perfect instance of what philosophers of science call the Duhem-Quine thesis, according to which a piece of recalcitrant data can be accommodated by a variety of competing theories, each of which rejects one or another of the auxiliary hypotheses that led to the prediction in the first place. Let's look at how this all played out.

Shortly after the formulation of the Standard Theory, Slobin (1966) conducted a study that focused on the processing of active and passive sentences, of which half were semantically reversible and half nonreversible, as illustrated in (16)–(19).[4]

(16)	**John saw Bill. / Bill saw John.**	**active, reversible**
(17)	**John was seen by Bill. / Bill was seen by John.**	**Passive, reversible**
(18)	**John ate the apple. / #The apple ate John.**	**Active, nonreversible**
(19)	**The apple was eaten by John. / #John was eaten by the apple.**	**Passive, nonreversible**

Slobin asked his subjects to match a picture to each such sentence, and measured their reaction times on this task. He found that in the nonreversible case, processing the active form in (18) and the passive form in (19) took an equal amount of time. FBG saw this as a disconfirmation of the DTC. They reasoned as follows: Each step in the comprehension process—i.e., each transformation—would have to take a determinate amount of time, and from this it seems to follow that reaction times on comprehension tasks should track the number of backwards transformations that each sentence would have to undergo before being successfully comprehended. On the strength of data such as those in the Slobin study, FBG concluded that the DTC is probably false and that "there exist no suggestions about how a generative grammar might be concretely employed as a sentence recognizer in a psychologically plausible model" (p. 368).

FBG's argument against the DTC prompted a number of reactions. Berwick and Weinberg (1982, 1983, 1984) point out that FBG's argument consists in displaying an inconsistent set and opting to give up one of the members. Here are the four claims they take to be in play:

[i] The correct theory of syntax is directly realized as a model of sentence processing.
[ii] The Standard Theory of transformational grammar is the correct theory of syntax.

[4] The sentences marked with a '#' exhibit a semantic defect or pragmatic infelicity. Following the policy of Chap. 5, the present discussion does not presuppose any particular way of drawing the semantics/pragmatics distinction.

[iii] Active and passive sentences are processed in the same amount of time.
[iv] The time it takes to process a sentence is an accurate measure of processing complexity.

FBG suggest that the weakest link is claim [i], and conclude that the DTC has been disconfirmed. But Berwick and Weinberg point out that we may, instead, hold on to claim [i] and reject any of the others—a shining example of the Quine-Duhem thesis. For instance, Bresnan (1978) gives up claim [ii], opting for an alternative theory of syntax.[5] Bresnan famously argued that the experimental results show that the Standard Theory of transformational grammar must be replaced by a different grammar—one that *can* be put to use in a psychologically plausible processing model. On these and other grounds, she proposed a new formalism, Lexical Functional Grammar (LFG), in which the effects of some transformations are offloaded to the lexicon. For instance, whereas the Standard Theory had a transformation for turning actives into passives, LFG posits two distinct lexical entries for the active and the passive forms of a verb.[6] This appears to have consequences for psycholinguistics. If LFG were implemented in a real-time processing model, such information would be *precomputed* and stored in lexical memory. And because accessing one lexical item is just as quick as accessing another, active and passive sentences are predicted take an equal amount of time to process, in accordance with Slobin's experimental result.[7]

Another strategy for resisting FBG's argument is taken by Forster and Olbrei (1973), who challenge claim [iii], presenting data that conflicts with Slobin's results. Using an experimental paradigm known as rapid serial visual presentation (RSVP), Forster and Olbrei found that nonreversible passives took significantly *longer* to process than nonreversible actives. They also argued that Slobin's experimental paradigm—matching sentences to pictures—does not reveal anything about sentence processing *per se*. In support of this, they pointed out that subjects were slower at matching passive sentences to pictures even when there was a 3-second delay between the presentation of the sentences and the pictures. Given that the HSPM takes far less than 3 seconds to process a sentence, active or passive, a more plausible explanation of the slower response times that Slobin observed would appeal to processing effects *outside* of the HSPM, perhaps in the visual system or at the interface between language and vision.

Berwick and Weinberg (henceforth, BW) develop a yet different sort of response FBG's argument. They concentrate on claim [iv], which concerns the relationship between reaction-time data and processing complexity. They point out that FBG's

[5] Berwick and Weinberg (1984) note that the revisions need not be as extensive as Bresnan envisions. Minor revisions—e.g., eschewing the so called "whiz-deletion" transformation—might do the trick.

[6] We saw a version of this strategy in our discussion of the models developed by Schank and his colleagues (Chap. 6).

[7] Note that this prediction only holds when the active and passive sentences are matched in all other relevant respects, e.g., sentence length and frequency of lexical items, n-grams, etc. I ignore this complication in the main text.

argument assumes the HSPM conducts its operations *serially*, first computing a *full* S-structure representation and only later computing the D-structure. BW argue that this assumption is groundless. In support of this, they sketch a parsing model, closely resembling that of Marcus (1980), on which S-structure and D-structure representations are computed simultaneously, *in parallel*.[8] On such a model, actives and passives would processed in an equal amount of time (*ceteris paribus*, fn. 7).

BW conclude that chronometric behavioral data can have a bearing on claims about the psychological plausibility of specific grammars *only when the parsing architecture and complexity measure have been fixed*. Given that we do *not* know what these parameters are in the case of the HSPM, it is unwarranted to conclude that transformational grammar is not at the core of a psychologically plausible parsing model. In light of such considerations, many psycholinguists came to see the DTC untestable, and hence irrelevant. As one critic put it:

> Psycholinguists are entitled to ignore the DTC not because they showed it to be untrue, but because the experiments that have been performed are irrelevant to its truth, and it is difficult to see how to devise experiments that would test it without having to make arbitrary assumptions about the other parts of the language understanding system. (Garnham 1983)

BW also challenged a pair of assumptions made by Bresnan (1978). Bresnan's argument for the psychological plausibility of an LFG-based parsing model rested in part on the claim that computing a structure on-line is more costly than accessing precomputed information. BW point out that, in a parallel processing architecture, computation is cheap. And, lacking a detailed knowledge of the specific neural mechanisms underlying language use, we simply cannot tell whether the HSPM has devoted its resources to memory access or to online computation.

Bresnan also assumes that *only* a parser that draws on LFG can make use of precomputed structural information. In response, BW note that a parser can instead draw on a modified version of transformational grammar. For instance, in such a parser, "bounded movement"—i.e., the kind of NP-movement associated with passive and raising constructions (Chap. 5)—has been precomputed and stored in the mental lexicon. "Transformational grammar … provides the necessary principles to distinguish among rule types so as to permit some rules to be handled by memory retrieval and others by active computation, all within a machine architecture like that assumed in the DTC or in Bresnan 1978" (BW 1984: p. 75). If this strategy were adopted, then the grammar used for parsing would not be *transparently* related to the transformational grammar advocated by Chomsky and others. But BW claim that the resemblance between them would still be substantive. And the grammar used for parsing would differ in important ways from Bresnan's LFG.

> [Bresnan assumes] that the transformations postulated by EST [the *Extended Standard Theory* of transformational grammar] should be thought of as "active," time-consuming computations. However, the necessity of this assumption is unclear unless there is also

[8] NB: This is just one kind of parallelism; it is *not* the same as the kind discussed above, in connection with the Earley parser. Nor does 'parallel' mean the same here as it does in the label 'Parallel Distributed Processing'. There are, then, at least three distinct kinds of parallelism. See Vasishth and Lewis (2006: p. 419) for discussion.

> insistence upon some strong version of transparency and we state the Move NP rule and its parsing correlate in exactly the same form. If transparency is relaxed, it is possible to embed an ST [Standard Theory] or EST based parser in a serial computational model. We could do this by "precomputing" the effects of the transformational component and storing the results in the lexicon. … Given the assumptions of transformational grammar … one should note that the precomputed lexical templates associated with NP movement would necessarily be governed by purely structural principles, unlike the templates of extended lexical grammar. (BW 1984: pp. 73–4).

Berwick and Weinberg's proposal in this passage yields a novel perspective on the psychological role of grammars, which I'll call *precomputed lexical realization* (PLR).

(PLR) The principles of a grammar, G, are realized in a parser in virtue of the fact that the lexical entries on which parser draws encode precomputed structures that are in accord with those principles.

Berwick and Weinberg note that PLR posits a "less than straightforward" relation between the grammar and the parser, but they remain hopeful that it is sufficient to capture an important sense in which transformational rules might have psychological reality: "Note that this in no way disconfirms transformational grammar either as a grammar or a central component of a parsing model; it says merely that the way in which a grammar may be embedded as a part of the model of language use is less than straightforward" (BW 1984: pp. 73–4). One way of determining whether BW are correct about this is to ask what place PLR would occupy in the taxonomy of positions that we saw at the outset of this chapter.

The answer, I submit, is that PLR is best seen as a version of the trivially true GRAM-CONFORM, which places no substantive constraints on the relation between grammar and language processing. To see this, consider a more general claim that BW advance, of which PLR is a special case.

(CGR) The principles of grammar G_1 are realized in the parser in virtue of the fact that its operations consist of sequence of derivational steps, each of which are governed by the principles of a grammar G_2 ($\neq G_1$), and G_2 is a **covering grammar** for G_1.

I call this position *covering-grammar realization* (CGR) on account of its reliance on the notion of a covering grammar, which we can define as follows: G_2 is a covering grammar of G_1 just in case (i) G_1 and G_2 are weakly equivalent—i.e., they generate the same set of surface strings, though not necessarily the same internal structures—and (ii) we can find the structural descriptions that G_1 assigns to sentences by parsing the sentences using G_2 and then applying a "simple" or easily computed mapping to the resulting output.

> Intuitively one grammar can be said to *cover* another if the first grammar can be used to easily recover all the parses that the second grammar assigns to input sentences. This being so, the first grammar can be used instead of the second grammar itself to parse sentences of the language generated by the second grammar. (Berwick and Weinberg 1984: p. 78)

These characterizations of the notion of a covering grammar make plain that the "covered" grammar plays no actual role in the operations of the parsing system. Only the covering grammar is involved.[9] The relation between the covered grammar and the covering grammar is a formal one, but it plays no role in processing linguistic input. BW appear to hold that the covered grammar (which they sometimes refer to as the "competence grammar"), is nevertheless psychologically real, on account of providing a superior description of the target language.

> [T]he cover relation provides a rich stock of cases where two grammars generate trees that do not necessarily look very much alike. Yet one grammar can serve as the "true" competence grammar for a language because it generates the proper structural descriptions while the other can be used for efficient parsing because of certain special structural characteristics of the trees it generates.

I pointed out earlier that, contrary to BW's assumption, a grammar's generating the "proper" structural descriptions of the sentences of a language does not constitute sufficient grounds for declaring it to be psychologically real. Stabler (1984) argues convincingly that syntacticians place constraints on descriptive grammars that may not be applicable in psycholinguistics. The compact and simple grammar that's useful for the explanatory practices of a syntactician may not have the right formal properties for efficient use in a parsing system. BW do provide compelling grounds for keeping in mind various "nontransparent" relations between grammars, of which "covering" is an instance.[10] But this methodological point cannot establish the psychological reality of a grammar that plays no actual role in the causal mechanisms of language processing.

---

[9] Steedman (2000) considers a position according to which the covering grammar and the covered grammar *both* play a role in on-line language processing. He argues convincingly, that this is implausible on basic evolutionary grounds: "[C]onsiderations of parsimony in the theory of language evolution and language development … might also lead us to expect that, as a matter of fact, a close relation is likely to hold between the competence grammar and the structures dealt with by the psychological processor, and that it will in fact incorporate the competence grammar in a modular fashion. One reason that has been frequently invoked is that language development in children is extremely fast and gives the appearance of proceeding via the piecemeal addition, substitution, and modification of individual rules and categories of competence grammar. Any addition of, or change to, a rule of competence grammar will not in general correspond to a similarly modular change in a covering grammar. Instead, the entire ensemble of competence rules will typically have to be recompiled into a new covering grammar. Even if we assume that the transformation of one grammar into another is determined by a language-independent algorithm and can be computed each time at negligible cost, we have still sacrificed parsimony in the theory and increased the burden of explanation on the theory of evolution. In particular, it is quite unclear why the development of either of the principal components of the theory in isolation should confer any selective advantage. The competence grammar is by assumption unprocessable, and the covering grammar is by assumption uninterpretable. It looks as though they can only evolve as a unified system, together with the translation process. This is likely to be harder than to evolve a strictly competence-based system" (pp. 227–228).

[10] "This approach allows us to hold the structural descriptions of a grammar fixed and then consider variations in parsing methods. The theory of grammar will limit the class of possible parsers to just those that cover the original competence grammar. This is possibly a strong limitation, hence of

Before leaving this topic, I would like to point out an important difference between BW's proposal with regard to transformational grammar and an approach to language processing that was inspired by Bresnan's work on LFG. Many current parsing models are based on a *lexicalist* grammar—a formalism that attempts to capture syntactic phenomena by enriching the information stored in the lexicon. The grammars whose primary formal operation is *feature unification* (Jurafsky and Martin 2008: ch. 15) all fall into this category. Constructing such grammars has become a dominant approach in syntactic theory, and parsers that incorporate them have been developed by Jurafsky (1993, 1996) and Vosse and Kempen (2000, 2009).[11] The important point for our purposes is that the relation between the parser and the grammar in these models is *not* correctly described by PLR.

Transformational grammars were initially proposed in a non-lexicalist formalism. BW pointed out that a parser can be governed by a covering grammar that precomputes the effects of some transformational rules and stores those effects in the lexicon. By contrast, unification-based grammars, such as Head-Driven Phrase Structure Grammar (HPSG), were developed in a lexicalist formalism from the very beginning. No modification is necessary in order to implement them in a unification-based parsing model. Hence, their relation to the parser is more direct than PLR. Indeed, it may well turn out that a lexicalist theory provides the most elegant and successful approach to syntactic issues *and* that the HSPM will be shown to implement a unification-based algorithm. In that case, the grammar used for parsing and the grammar preferred by the syntactician would be identical, thus vindicating the strong competence hypothesis and either **REP-GRAM-PROC** or **EMB-GRAM-PROC**. Although we will not explore the prospects for unification-based parsing

potential interest to parsing theory. Such cases provide real examples of the existence of nontransparent ways to incorporate grammars into models of language use … As far as parsing is concerned, both the theory and practice of parser design have made considerable use of a nontransparent relation between grammar and parser, that of grammatical cover. But why should the notion of a covering grammar play a role at all? That is, given that the mapping between grammar and parser can be quite abstract, why should we connect them at all? Why not just build a possibly nonlinguistically based parser? The answer is that keeping levels of grammar and algorithmic realization distinct, it is easier to determine just what is contributing to the discrepancies between theory and surface facts. For instance, if levels are kept distinct, then one is able to hold the grammar constant and vary machine architectures to explore the possibility of a good fit between psycholinguistic evidence and model. Suppose these results came to naught. We can then try to covary machine architecture and covering mappings, still seeking model and data compatibility. If this fails, one could then try different grammars. In short, modularity of explanation permits a corresponding modularity of scientific investigation. For a complex information processing system like the language faculty, this may be the investigative method of choice" (Berwick and Weinberg, 1984: p. 78–80).

[11] While both models were constructed for the explicit purpose of accounting for a wide range of psycholinguistic data, Vosse and Kempen's model has the additional virtue of explaining the processing difficulties characteristic of aphasia. As they point out, this is a significant feature of their approach, given that most models aim to account only for the processing difficulties that statistically normal speakers encounter on garden-path sentences and other local ambiguities. A notable exception to this trend is Grodzinsky (2000), who argues that Broca's aphasics lack the operations that theorists in the GB tradition (e.g., Haegeman 1994) treat as "A-movement".

models here, I suspect that many of the conclusions we draw concerning Minimalist parsers in Sect. 9.5 will carry over without significant modification.

9.2.2 *The Failure of Old-School Transformational Parsers*

Whatever the case about its fit with psychological data, the Standard Theory of transformational grammar was taken up by computational linguists who attempted to use it in writing parsing algorithms (e.g., King 1983). According to Harkema (2001), these typically worked in four steps.

> First, the parser will generate a set of possible surface trees for the sentence, using a so-called covering grammar. Next, the parser will reconstruct a set of possible base trees from the set of possible surface trees by applying the transformational rules of the grammar in reverse. After this step, the parser will discard those base trees that cannot actually be generated by the base component of the grammar. Finally, the parser will feed the remaining base trees into the transformational component to establish which base trees can in fact produce the sentence to be parsed. The set of base trees surviving this last test will be returned by the parser as the result of the parsing process. (Harkema 2001: p. 6)

Such models ran into two kinds of problems. The first type is familiar from our discussion in the previous chapter: inefficiency. Early versions of transformational grammar were not formally restricted. It was clear from the outset that they were stronger than context-free, but later research (Peters and Ritchie 1973) showed that they were actually capable of generating recursively enumerable languages. Recursively enumerable languages are type-0 languages on the Chomsky hierarchy. A grammar that generates them is virtually unrestricted, and stronger than even a context-sensitive grammar.[12] The search space generated by reverse transformations was correspondingly vast.

If inefficiency were the only issue, theorists might be justified in holding out hope that clever search strategies and suitable heuristics would eventually solve the problem (Chap. 8, Sect. 9.5). There was, however, a more principled difficulty. Some transformations in the Standard Theory were not reversible, on account of the fact that they deleted material from trees.

> Applying these rules in reverse can be very problematic, because the tree to which the reverse transformational rule is to apply may not always contain enough information to reconstruct the content and position of the deleted material. Because of these computational problems, early transformational parsers were not considered very successful. (Harkema 2001: pp. 6–7; cf. Fodor et al. 1974: pp. 315–316)

This is perhaps the principal reason why, by late 1970s, many were convinced that transformational grammars were not only at odds with psycholinguistic data, but also unusable as computational models of human language comprehension. The preceding discussion shows why the former claim is naïve; transformational

[12] A grammar strong enough to generate natural languages is widely believed to be mildly context-sensitive.

grammars were not disconfirmed by psycholinguistic data.[13] However, the failure of the computational parsing models based on the Standard Theory was quite real.

The Standard Theory itself suffered from a number of problems. Its unrestricted expressive power and its proliferation of language- and construction-specific rules came to be seen as its most serious defects. This led to a number of revisions, thus giving rise to the *Extended Standard Theory* and the *Revised Extended Standard Theory*. At last, the appeal to construction-specific ordered transformations was abandoned altogether, in favor of what is now widely known as the *Principles and Parameters* (P&P) approach, within which the Government and Binding theory (GB) and the Minimalist grammars were formulated. The distinction between D-structure and S-structure was maintained (in various forms) throughout these developments, to be abandoned only with the rise of Minimalism. We will examine the central tenets of GB in Sect. 9.4 and of Minimalism in Sect. 9.5.

The failure of parsing models based directly on the Standard Theory gave rise to a host of proposals regarding how the HSPM assigns syntactic structure to linguistic input. Fodor et al. (1974) suggested that it does so by using a battery of construction-specific heuristic strategies, such as the "canonical sentoid strategy," which dictates that N-V-N structure be interpreted as having an S-O-V form. This proposal has been roundly rejected. J. D. Fodor (1998a: p. 287) observes, it wouldn't work at all for verb-initial languages (e.g., Irish), or verb-final languages (e.g., Japanese). Of course, it may be that parsing heuristics differ from language to language. But a uniform (i.e., language-neutral) account of parsing preferences would be vastly more preferable. In any event, Pritchett (1992: pp. 22–26) provides a range of data that are at odds with FBG's proposal.

The heuristics-based approach is unsatisfactory for another reason: it posits different strategies for what are essentially the same linguistic phenomena. This complicates both the parsing theory and the theory of language acquisition. Wanner and Maratsos (1978) express this important point in the following passage.

> With the empirical failure of psychological models that incorporate transformational rules, however, many psychologists have abandoned not only transformational grammar but also the assumption about the psychological reality of the generalizations on which transformational grammar is based. The heuristic strategies schemes advanced by Bever, Fodor, and others make no claim to capture linguistically significant generalizations. For example, Fodor and Garrett (1967; also Fodor, Garrett, and Bever 1968) propose one heuristic strategy to handle object relative clauses and an entirely different heuristic strategy to handle subject relative clauses (Fodor et al. 1974). Presumably, their system of strategies can only be expanded to handle the full range of relative clauses by adding more heuristic strategies. Such a system does not class relative clauses together in any way, nor does it represent their relation to declarative clauses. If a child acquired this type of grammatical representation of relative clauses, he would face the formidable task of having to learn each relative clause strategy independently of all the others, and none of his knowledge of declarative clauses would transfer to relative clauses.

[13] A similar position is advocated by Phillips (1994, 1996).

Wanner and Maratsos take these considerations to motivate an approach based on what they call Augmented Transition Network grammars, for which a number of processing algorithms were developed in the 1970s. They write:

> In our view, … the process of comprehension places psychological constraints on the form in which grammatical generalizations can be maintained. Thus, whatever the listener knows about the structure of his language, he must represent it in a form that can be used in a comprehension process that extracts semantically relevant grammatical information in a single, beginning-to-end pass over the surface structure of the sentence. The virtue of ATN notation is that it gives us a precise way of representing grammatical generalizations in just this form, thus permitting us to make strong tests of our assumptions about comprehension. (Wanner and Maratsos 1978: pp. 158–159)

It is important to recognize that the ATN framework is by no means the only conceivable solution to the problems that Wanner and Maratsos identify in Fodor, Bever and Garrett's heuristics-based approach.[14] Still, it proved influential and shed additional light on the psychological reality debate. In the next section, we continue our historical tour of the parsing literature, examining both the details of the ATN approach and the objections that have been leveled against it.

9.3 Augmented Transition Networks

9.3.1 Parsing with an ATN

The general framework of Augmented Transition Networks (ATNs) can be conceived of as a set of notational devices for constructing recursive phrase structure grammars that differ in some ways from the context-free grammars described in Chap. 8. ATN grammars consist of a set of networks, each of which characterizes a kind of phrase. There is a network for full sentences, another for noun phrases, a third for verb phrases, and so on. Fig. 9.1 illustrates two very simple networks. (We will see more complex examples below.)

The task of the network in Fig. 9.1a, at the top of the diagram, is to construct a full sentence. It does so by first looking for a noun phrase and then for a verb. After the verb is found, it checks for another noun phrase. If it finds one, then it ends its

[14] J. D. Fodor (1998a) writes: "Though the failure of the derivational theory of complexity caused a stir, before long a way was found to cope with derivational operations algorithmically and on-line, by folding them into the phrase structure assignment operations: Establish a transformational dependency between two sentence positions just in case it is needed in order to reconcile surface derivations from base phrase structure rules. (See the HOLD hypothesis of Wanner and Maratsos 1978; the Superstrategy of J. D. Fodor, 1980). As a result, parsing even with a transformational grammar could be seen as "left to right," systematic and incremental, effected by a precise program that faithfully applies the mental grammar. There was no need any longer for heuristics…" (p. 289). In this passage, Fodor identifies the HOLD hypothesis—an important feature of ATN parsers, to be discussed below—as one of two competing ideas for avoiding Fodor, Bever and Garrett's heuristics; the other is her own Superstrategy.

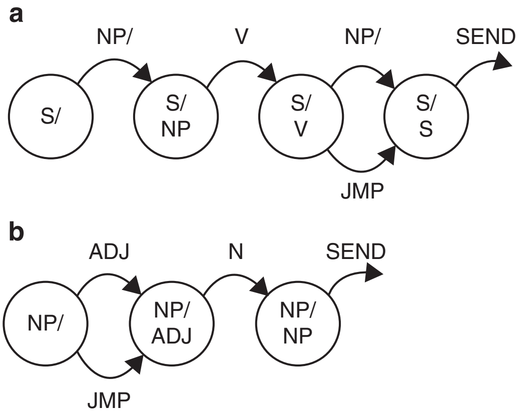


Fig. 9.1 (**a**) A network for constructing simple sentences. (**b**) A network for constructing simple noun phrases (*Source*: Bates 1978)

analysis of the sentence. If it doesn't, then it ends the analysis anyway. The network in Fig. 9.1b, at the bottom of the diagram, is designed to build noun phrases. It begins by checking for an adjective. If it doesn't find one, it looks for a noun. When a noun is found, it ends its analysis of the noun phrase. At this point, whatever process initially called for a noun phrase receives the go-ahead to resume.

The circles in the diagram represent states, linked to one another by arcs. The arcs are labeled with conditions that must be satisfied before a transition can be made. State **S/** in the network at the top of the diagram initiates the search for a full sentence. This state activates the **NP/** arc, which triggers a search for a noun phrase; unless a noun phrase can be found, processing cannot continue. Building noun phrases is the job of the network in Fig. 9.1b, so the activation of the **NP/** arc is an instruction for the activation of that network. By activating the **NP/** state, the sentence network has in effect predicted that the first phrase in the sentence will be a noun phrase.[15] Once this prediction is made, the noun phrase network sets to work. Let's imagine that the sentence it receives as input is (21).

(21) Accomplished artists command respect.

The **NP/** state has two arcs linking it with the **NP/ADJ** state. The arc at the top is labeled **ADJ** and the one at the bottom **JMP**. By convention, the action associated with the higher of two arcs will take priority. In this case, then, the **ADJ** arc is attempted first, initiating a search of the lexicon to determine whether the first word of the input is an adjective. The noun phrase network determines that 'Accomplished' is indeed an adjective. If it had not, the **JMP** arc would have been activated, which

[15] Recall that making such predictions prior to receiving any input is a characteristic of top-down parsers. The ATN system described here is therefore an instance of the top-down approach, though one that relies on backtracking rather than parallel processing in order to deal with local ambiguity resolution.

would allow the network to move to the next state without having satisfied any condition—a sort of "free pass." Once the **ADJ** condition is satisfied, the network transitions into the **NP/ADJ** state, which has only one arc to the **NP/NP** state. This arc, labeled **N**, imposes the condition that a noun must be found before the process can continue. The network thus initiates another search of the lexicon, and determines that 'artists' is indeed a noun. This allows for a transition into the **NP/NP** state, from which the only arc is **SEND**. The **SEND** arc instructs the network to complete the noun phrase by storing in memory the structure it has built thus far—viz., [$_{NP}$ [$_{ADJ}$ Accomplished] [$_{N}$ artists]]. Control is then ceded back to the process that initiated the noun phrase network.

Returning, then, to the sentence network in Fig. 9.1a, we see that the condition of the **NP/** arc has been satisfied, allowing for a transition from the **S/** state to the **S/NP** state. The only arc leaving the **S/NP** state imposes the condition that a verb must be found. A lexical search reveals that 'command' is a verb. The network transitions into the **S/V** state, from which there are two arcs, hence two possibilities. One is to find another NP, which involves reactivating the noun phrase network. Because the **NP/** arc is the higher of the two, this action is tried first.[16] The input is the word 'respect', so the lexical search concludes that the next item in the input is not an adjective. Hence, the noun phrase network uses the **JMP** arc to get to the **NP/ADJ** state. The **N** arc initiates another lexical search, this time successful—'respect' is a noun. The **SEND** arc then "pops" the processing back up to the sentence network, which is now in the **S/S** state—its final state. The **SEND** arc from the **S/S** state completes the sentence, outputting the parse in (22).

(22) [$_{S}$ [$_{NP}$ [$_{ADJ}$ Accomplished] [$_{N}$ artists]] [$_{VP}$ [$_{V}$ command] [$_{N}$ respect]]].

Wanner and Maratsos (1978) consider how an ATN would parse relative clauses—a process that involves finding positions occupied by *wh*-traces.[17] As the examples in (23)–(25) illustrate, the *wh*-trace can appear in subject, object, and adjunct positions.

(23) … the girl who <u>*wh*-trace</u> talked to the teacher about the problem …
(24) … the teacher whom the girl talked to <u>*wh*-trace</u> about the problem …
(25) … the problem that the girl talked to the teacher about <u>*wh*-trace</u> …

The antecedent of each *wh*-trace is always the NP that appears before a relative pronoun—i.e., before the complementizers 'who', 'whom', and 'that', in the examples above. The parser's task is to locate the gap and co-index it with the correct NP.

In order to accomplish this, the ATN parser must incorporate a feature Wannaer and Maratsos call the HOLD list—a memory bank in which items are stored and

[16] Notice that imposing an uniform order on the operations of the ATN can be a way of implementing parsing principles like Late Closure (Wanner 1980). See the discussion of Fodor and Frazier (1980) for a decisive critique of this approach.

[17] See Chap. 5 for a discussion of *wh*-traces and their effects on sentence processing.

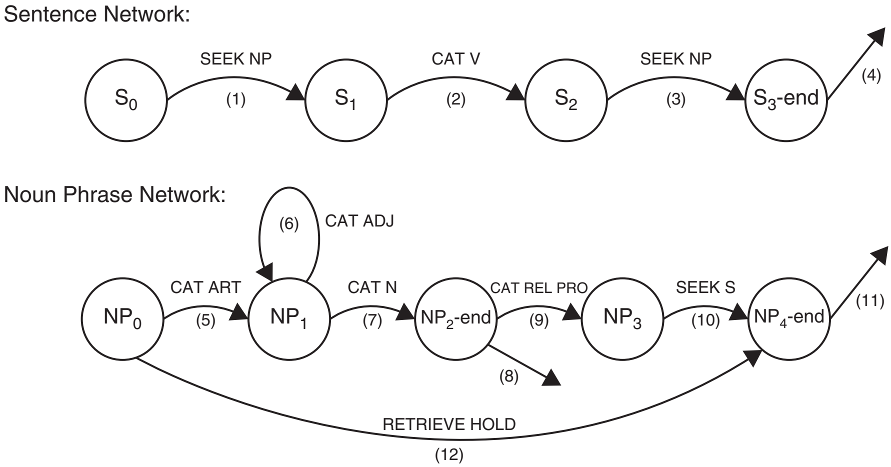


Figure 9.2 The labeling conventions employed here are slightly different from those in Fig. 9.1. The arcs carry SEEK and CAT labels. The label SEEK NP is equivalent to the **NP/** label in Fig. 9.1. It indicates that the noun phrase network should be activated before the transition to the next state can be made. The CAT V label is equivalent to the **V** label in Fig. 9.1. It indicates that the lexicon should be consulted to determine whether the next word in the input is a verb. The arc labels ART and REL PRO have been added. These call for a lexical decision regarding whether the next word in the input is an article ('a', 'the', etc.) and a relative pronoun ('who', 'which', etc.), respectively. Since adjectives can be repeated indefinitely ("the *awesome shiny new expensive* car"), the ADJ arc can lead back into the state from which it originated, so as to allow for additional activations of that very same arc. Finally, the arcs have been numbered. The numbers do *not* correspond to the order in which the actions are carried out. They serve as an expository device in Wanner and Maratsos's discussion of the thematic role assignments that are carried out in the course of the parsing process.

retrieved at later points in the parsing process. The addition of this feature allows the ATN to parse all three types of relative clauses exemplified in (23)–(25). Let's walk through a processing sequence, to illustrate this concept. Following Wanner and Maratsos, we will consider the object-extracted relative clause in (26). (A diagram of an ATN that can be used to parse this sentence appears in Fig. 9.2.)

(56) The old train that the boy watched left the station.

As before, the process begins with a prediction to the effect that the first part of the sentence is a noun phrase. The noun phrase network is activated. It runs through the CAT ART, CAT ADJ, and CAT N arcs, finding first the article 'The', then the adjective 'old', and then the noun 'train'. The network then attempts to traverse the CAT REL PRO arc, thus checking for a relative pronoun. If there were no relative pronoun in the input, then the noun phrase would be completed and processing would "pop" back up to the sentence network. In the case we're considering, though, the network finds the relative pronoun 'that'. This automatically places the noun phrase that has so far been constructed (i.e., 'The old train') onto the HOLD list.

The system's working assumption is that this noun phrase will be reactivated when the position of the *wh*-trace in the relative clause is discovered. The network now enters state NP_3, from which the only arc is a directive to activate the sentence network. The logic behind this move is that relative clauses are basically just declarative sentences with a missing noun phrase. For instance, the relative clause bracketed in (27) is just the declarative sentence (28), except lacking a subject at the position occupied by the *wh*-trace.

(27) … the girl who [*<u>wh</u>*<u>-trace</u> talked to the teacher about the problem] …
(28) The girl talked to the teacher about the problem.

Hence, in the course of constructing the relative clause (construed by the network as a declarative sentence), the ATN will find that some noun phrase is needed but not present in the input. This will prompt the ATN to check the HOLD list, to see if there are any noun phrases waiting to be incorporated. Here is how this happens.

The sentence network, now activated for the purpose of constructing the relative clause, begins by traversing the SEEK NP arc—i.e., seeking a noun phrase.[18] This re-activates the noun phrase network. The words 'the' and 'boy' are identified as a noun phrase and processing pops back up to the sentence network. The SEEK V arc is successfully traversed when the word 'watched' is identified as a verb. At this point, the sentence network attempts to locate the next NP—i.e., the object of 'watched'. The lexical search indicates that that 'left' is not a noun phrase. So the network checks the HOLD list and finds a noun phrase there. This noun phrase satisfies the condition imposed by the SEEK NP arc between states S_2 and S_3. The relative clause, treated as a declarative sentence, has been processed. The network goes on to work on the verb phrase 'left the station' in the familiar way, which we need not rehearse here.

Wanner and Maratsos note that the ATN just described, if taken to be a psychologically plausible model of the HSPM, makes predictions about the memory costs involved in human sentence processing. In particular, when a relative pronoun is found, the NP that has been constructed ('the old train', in the case just discussed) must be placed on the HOLD list, thus exacting a toll on short-term memory resources. They go on to present the results of two experiments that they conducted to show that human listeners do indeed exhibit processing effects indicative of increased memory load after the relative pronoun. The details of the experiments need not concern us here; suffice it to say that the results bode well for the psychological plausibility of the ATN model.

Wanner and Maratsos go on to compare their model with the heuristics-based approach advanced by Fodor et al. (1974). On that approach, three separate heuris-

[18] Notice that the ATN described here, unlike the simple one discussed at the outset of this section, makes it possible for the sentence network to be activated in the course of an operation initiated by that very network. As Fig. 9.2 illustrates, the sentence network can activate the noun phrase network, which can subsequently activate the sentence network. This allows for the embedding of sentences within sentences. This and similar formal devices render the ATN recursive, hence at least as powerful as a recursive context-free grammar.

tics would need to be consulted and applied in the course of processing the three types of relative clause exemplified in (23)–(25).

> In an ATN notation such a system of independent strategies would appear as a large set of distinct arc sequences following the head noun arc in the noun phrase network. But recall that this proliferation of arc sequences is exactly what is avoided in the HOLD model, and avoiding it is possible only because the HOLD model preserves a generalization about the class of all relative clauses—namely, that all such clauses appear to be declarative clauses with a phrase missing. Our evidence for the HOLD hypothesis, then, is evidence that the language user's comprehension system does incorporate at least this one generalization about the relation between clause types. As everyone knows, this is precisely the kind of generalization that can be captured by transformations. Thus, our evidence indicates that the comprehension system may employ the kinds of generalizations captured by transformational grammar, although not in transformational form. (p. 159)

9.3.2 The Bearing of ATNs on the Psychological Reality Issue

The networks just described can be thought of as simple grammars. Although the presentation above casts each of them as a parsing model, there is no obligation to view them that way. An ATN can be seen simply as an alternative way of describing the surface syntax of a language—i.e., for stating the optional and obligatory relations (both linear and hierarchical) between syntactic categories. Instead of stating such relations in a declarative format, as context-free grammars do, an ATN grammar encodes these relations in *procedural* instructions—rules for what to *do* in a certain context. Still, although an ATN can be seen as nothing more than a grammar, it is nevertheless quite natural to interpret it as a processing model. The only declarative data structures that need to be appended in order to make it operate as a parser are a lexicon and a memory bank for storing the partial structures that are assembled at intermediate points in the process.

This duality of the ATN formalism has significant implications for the psychological reality of grammar. It provides a vivid example of the way in which the rules of a grammar can *govern* processing, in a sense that comports with the position that I've been calling **EMB-GRAM-PROC**. The very simple ATNs discussed above are perhaps not the best aid to our thinking about this issue. Consider a more robust ATN grammar—the one used in the LUNAR parser (Woods 1973), depicted in Fig. 9.3 below.[19] Imagine a hardwired electrical circuit whose states correspond, one-by-one, to the states of this grammar (denoted by the circles in the diagrams in

[19] I owe these beautiful diagrams to Bates (1978), whose comprehensive discussion of ATNs is second to none.

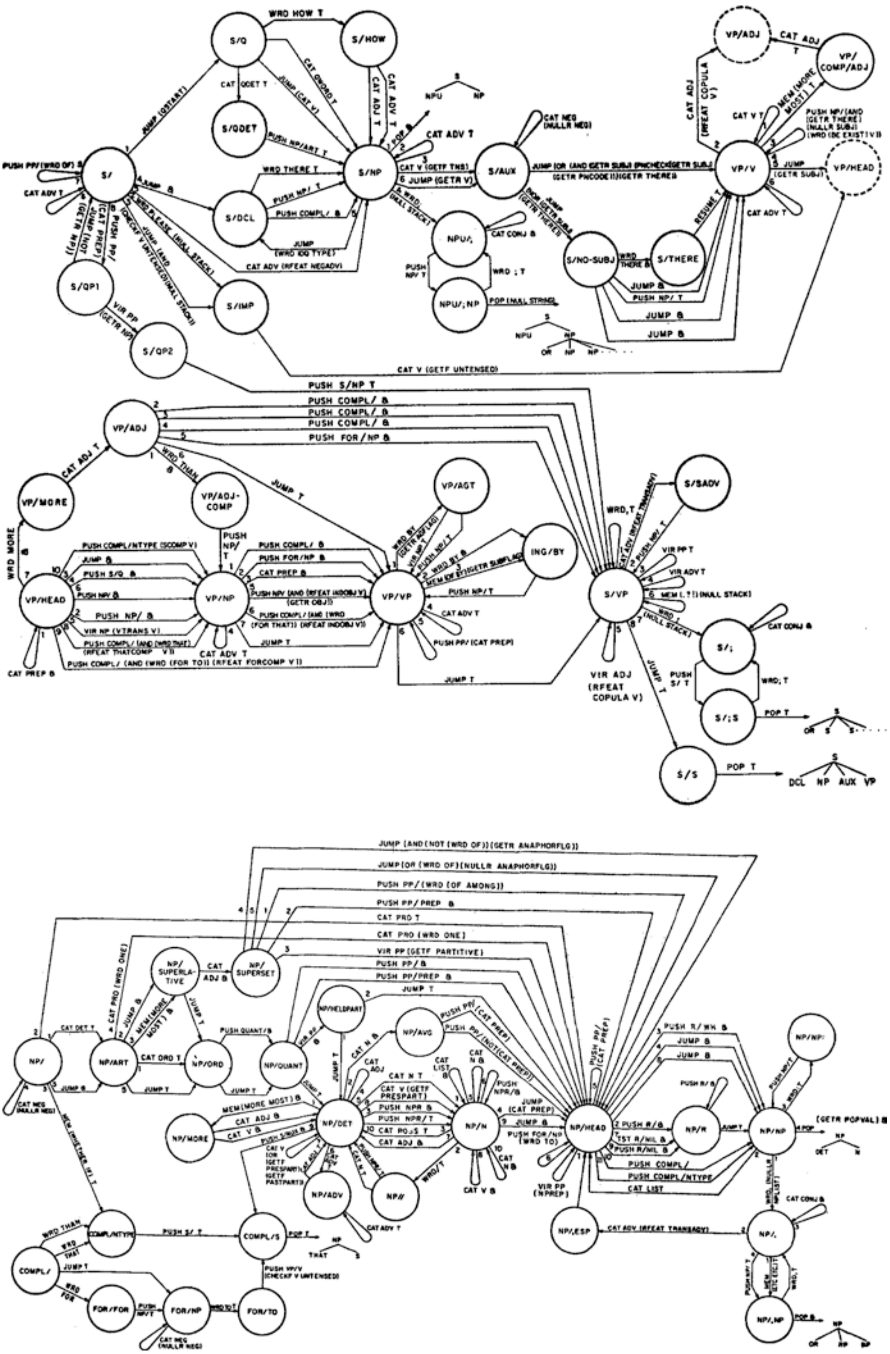


Fig. 9.3 The ATN grammar employed by the LUNAR system (Woods 1973) (*Source*: Bates 1978)

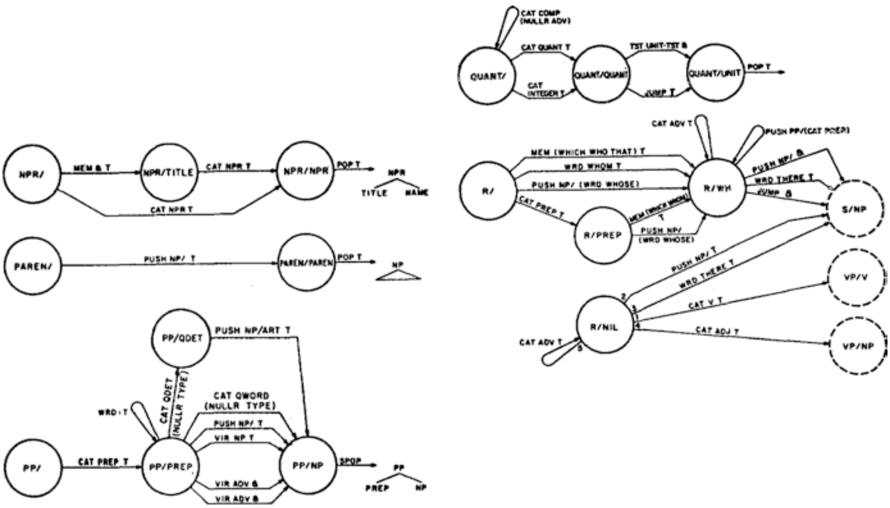


Fig. 9.3 (continued)

Fig. 9.3).[20] The circuit can be embedded in an object that resembles an ordinary pocket calculator, except that the buttons each stand for a word, rather than a number, and the output on the screen is a phrase marker rather than a numerical solution to a calculation query. When linguistic input is entered, the flow of current through the circuit constitutes state-to-state transitions—precisely those encoded in the arcs of the ATN grammar. The result is a parse of the input sentence.

The ATN circuit that we are now contemplating can be said to *embody* an ATN grammar without explicitly representing it. The grammar is not stored in a declarative format, as a separate data structure awaiting to be read by the system at an opportune time. Rather, the transitions between circuit states—i.e., the actions of the parser—are "in accordance with" the ATN grammar, not merely in the weak sense that they produce the phrase markers that the grammar assigns to strings of the language, but in a much stronger sense: The device could not output those phrase markers if its internal organization did not *embody* the grammar—i.e., if the states of the circuit did not correspond, one by one, to the nodes of the grammar. An alteration in the circuit would be an alteration in the grammar that the circuit embodies. If the altered grammar were not *true of* the expressions of English, then the parser *wouldn't work*. It would either

[20] Bates (1978) discusses various ways in which ATNs can be compiled and implemented. Her treatment of the issue demonstrates that the hardwired circuit described in the main text is by no means the only possible implementation of ATN parsing routines. Indeed, to my knowledge, such a circuit has never been constructed. ATN grammars are typically encoded as data, just like the context-free grammars that we considered earlier. This accords with the position I've labeled rep-gram-data. But, as noted above, this may well be an artifact of the techniques that computational linguists find most convenient in building their models. No inference can be drawn about how the human brain would implement an ATN system.

halt when presented with perfectly well-formed inputs, or it would process the input in a way that would lead to systematic misconstruals at the level of semantics.

9.3.3 Problems for the ATN Framework

Given the success of ATN models, both as descriptions of surface syntax and as parsing systems, one might wonder why the approach fell out of favor in subsequent decades. There are several reasons, which we briefly review here. The first few are pretty weak; I leave the more weighty objections for last.

First, ATN parsers are “brittle,” in the sense that they cannot parse ungrammatical input. In contrast to human listeners, who can understand sentences like (29) with relative ease, an unadorned ATN simply grinds to a halt when confronted with such input.

(29) *The boys is running.

This is by no means fatal. The ATN architecture can be modified in various ways to deal with the problem. For instance, upon encountering (29), a more sophisticated ATN parser can treat its failure as a signal to back up and revise its lexical decisions, treating ‘boys’ as ‘boy’ or ‘is’ as ‘are’, depending on the details of the design. Other modifications are imaginable. It has been remarked that these strategies are *ad hoc* (Berwick 1991b). But it’s conceivable that experimental evidence would reveal that human hearers employ such strategies in dealing with ungrammaticality. Concerning the relevant psychological experiments, Berwick writes:

> [S]imple agreement violations seem to cause more problems than they ought to. Given this, Crain and Fodor (1985) argue that the real problem has to do with confusability: subjects are looking for a near-miss grammatical examples, and that’s possible with agreement violations but not possible with violations such as *Who did you see Picasso’s picture of.* Therefore, in the agreement violation case, processing will be longer as the subject searches for a possible alternative; in the second, constrain violating example, the subject gives up rapidly. Freedman and Forster (1985) attempt to lay this matter to rest by trying to show that correctability does not account for all such results, but from my point of view the entire matter still seems unclear. (Berwick 1991b: p. 182)

The second problem for ATNs has to do with familiar efficiency issues. ATN parsers are top-down systems, which are known to be less efficient than their bottom-up counterparts. Marcus (1980: ch. 2) discusses some of the problems that plague ATNs. He notes that these parsers are *nondeterministic*, relying on a backtracking strategy in the face of an error in mid-sentence ambiguity resolution. Given that the parser may have to backtrack arbitrarily far, checking all of the decisions it made at every choice-point, this strategy can quickly become inefficient.

As we noted above, other nondeterministic systems invoke parallel processing to solve this problem, essentially relying on expanded memory resources to store multiple parses whenever an ambiguity presents itself before the end of a sentence.

Again, though, the expansion of memory resources can quickly spiral out of control, and in parallel systems that rank candidate parses, computational resources are needed to revise the rankings on the fly.[21] Some theorists suggested using the HOLD feature to delay parsing decisions until the system reaches a point in the input where an ambiguity can be resolved without the risk of error—a "minimal commitment" model. But this is psychologically implausible; there is ample experimental evidence that the HSPM does not delay its decisions in this fashion (Chap. 5).

Marcus (1980) invoked these considerations in rejecting the nondeterministic models that were available at the time, including ATN parsers, and in motivating his own deterministic parser.[22] However, the problems are not as severe as Marcus's discussion would lead one to believe. As discussed earlier, a variety of methods have been developed to increase the efficiency of top-down parsers. The use of grammar-based and resource-based processing heuristics, as well as probabilistic techniques (Chap. 8), may well be able to resolve these difficulties. Indeed, in an article championing ATN models, Wanner (1980) makes precisely these suggestions.[23]

Let's now turn to the strongest arguments against the ATN framework. Fodor and Frazier (1980) argue that the empirically well-confirmed processing principle, Right Association (a.k.a Late Closure.), is not realizable in an ATN architecture in a principled way.[24] More damning, Fodor and Frazier point out that ATNs cannot implement Minimal Attachment (MA) *at all*. In a moment, will see how their competing model, the "Sausage Machine," implements these principles. Presently, let us examine their argument for the claim regarding MA.

Wanner (1980) suggested that MA can be implemented in an ATN by a scheduling routine that enforces the following procedure: Traverse all CAT arcs before attempting to traverse a SEEK arc. However, as Fodor and Frazier (1980) point out, the CAT-before-SEEK strategy is *not* equivalent to MA. The clinching example is the sentence fragment in (30), which can be construed as the beginning of multiply-conjoined NP, as in (31), or as a reduced relative clause, as in (32).

---

[21] For a discussion and defense of ranked parallel models, see Gibson (1991).

[22] Historically, the exploration of ATN parsers gave way to a research program launched by Marcus (1980) into the capabilities of *deterministic* parsing models. I will not discuss deterministic parsers here. Despite their intrinsic interest, I do not see that they shed any new light on the psychological reality debate. In addition, there are compelling grounds for rejecting both the original Marcus parser and its "D-theoretic" offspring as viable models of human parsing (Pritchett 1992: pp. 44–51).

[23] The majority of Wanner's discussion is devoted to examining how Minimal Attachment and Late Closure (which he calls Right Association) can be incorporated into an ATN system. The only mention of probabilistic strategies is at the very end of the paper (p. 224, fn. 8).

[24] Right Association (RA) is a principle originally proposed by Kimball (1973). According to this principle, "terminal symbols optimally associate with the lowest nonterminal node" (Kimball, 1973: p. 24). A version of this principle survives today under the label 'Late Closure' (Frazier 1979; Frazier and Fodor 1978; Fodor 1998a). Kimball's commitments about the architecture of the parser differ slightly from Fodor and Frazier's, making it hard to say that RA and LC are identical claims clothed in different words. But the claims are certainly very similar.

(60) The man the girl...

(61) The man the girl and the dog played in the park.

(62) The man the girl dislikes was walking in the park.

Evidently, the HSPM prefers the conjunction analysis (31) over the relative-clause analysis (32).[25] But, while MA predicts this preference, the CAT-before-SEEK strategy does not, as the reader can verify by scrutinizing the ATN grammar in Wanner (1980).

Fodor and Frazier argue that these problems for the ATN models stem from their not *explicitly* representing the rules of a grammar in a separate data structure. By contrast, Fodor and Frazier's model, the Sausage Machine, stores the grammar in what they call a "rule library."[26]

> We have argued (in connection with Minimal Attachment) that the grammatical rules for the language should be stored separately from the specification of the computational operations involved in the application of those rules. They will reside in a special 'rule library', and will be accessed by the executive component of the parser as needed. (p. 451)

This is a significant difference between the ATN and the Sausage Machine. In the terms that we've developed here, it's the difference between **EMB-GRAM-PROC**. and **REP-GRAM-DATA**. Fodor and Frazier point out that, while both models can implement the principle of Late Closure (LC), only their model can implement MA and give a *principled explanation* of the fact that the parsing preferences of the HSPM are governed by LC. The explanation is principled because it follows directly from the architecture of the Sausage Machine—including, crucially, the existence of a separate rule library—whereas the ATN model would have to stipulate the relevant scheduling routines in an *ad hoc* fashion. Let us briefly examine the Sausage Machine, to see how it incorporates MA and RA/LC.

The Sausage Machine (henceforth, SM) has two components—the preliminary phrase packager (PPP) and the Sentence Structure Supervisor (SSS). The PPP examines the first five or six words of the input and builds a partial mental phrase marker for those words. This material is then erased from its memory and passed on to the SSS, where it is temporarily stored, awaiting integration with further input. When the PPP passes along the next partial phrase marker, the SSS puts it together with the prior one. If necessary, the process repeats several times, eventually yielding a complete phrase marker for the entire sentence.

---

[25] Frazier and Fodor do not discuss the lack of commas in (30)–(31). But this is not a problem for their claim that the HSPM prefers the reading in (31) over the one in (32). Indeed, it *bolsters* their claim. For, lacking a comma, the fragment in (30) is more likely to be interpreted as introducing a reduced relative clause than a conjunctive list. The fact that the HSPM nevertheless prefers the conjunctive analysis is therefore even *more* noteworthy.

[26] See Frazier and Fodor (1978) for a detailed description and motivation of the Sausage Machine. McRoy and Hirst (1990) extend the model in interesting ways. Criticisms can be found in Pritchett (1992: pp. 30–40).

The motivations for this dual-component structure had to do with memory limitations. As is well known, only a small number of "items" can be held in short-term working memory. But this limitation can be transcended, to some extent, by "chunking" several simple representations into a complex structure and treating the result as a single "item". The cost of this strategy is that the system becomes blind to the internal structure of the complex representation, and must expend computational resources to "look within" that representation if it becomes necessary to do so. It follows that if the SSS is unable to integrate the outputs of the PPP, then those representations have to be "unpacked," so to speak, before structural revision can take place. This gives rise to the processing principle known as Revision as a Last Resort (RALR), thus explaining why many kinds of reanalysis are computationally costly.

In the SM, the input triggers a sort of competition between whatever rules might be applicable. The rules are "accessed in parallel and selected in terms of the outcome of a 'race'—the first rule or combination of rules that successfully relates the current lexical item to the phrase marker dominates subsequent processing" (Fodor and Frazier 1980: p. 434). A particular analysis of the input is generated by whatever set of rules wins the race. And since the application of each rule takes time and computational resources, the winning set of rules is always going to be the *minimal* set required by the input. For example, suppose that the model is working with the following simple grammar:

(33)	*S → NP VP*	(35)	*NP → Det N*
(34)	*VP → V NP (PP)*	(36)	*NP → NP PP*

In this case, the input in (37) would be parsed as (38), not as (39).

(37) David read the letter to Aoife.
(38) David [$_{VP}$ read [$_{NP}$ the letter] [$_{PP}$ to Aoife]].
(39) David [$_{VP}$ read [$_{NP}$ [$_{NP}$ the letter [$_{PP}$ to Aoife]]].

The reason is that modifying a verb requires applying only one rule—viz., (34)—whereas modifying a noun requires applying two—viz., (35) and (36). Thus, SM predicts this particular parsing preference. Notice also that the structure in (38) requires attaching the PP directly to the VP, whereas the structure in (39) requires attaching it to the NP, thus necessitating the creation of an extra NP node. The creation of extra nodes violates MA. Thus, as Fodor and Frazier point out, the "race" aspect of the SM not only predicts an isolated parsing preference, but also explains why, quite generally, the preferred attachments conform to MA.[27]

[27] Another strength of the model is that it explains why MA sometimes does *not* apply. For instance, consider the sentence *Joe read the newspaper article, the card, the telegram, and the letter to Mary*. MA dictates that 'to Mary' should be attached to the verb, analogously to (38) above. But, in this case, most readers prefer that this preposition be attached to 'the letter'. Fodor and Frazier's explanation is this: By the time the HSPM encounters the preposition 'to Mary', the verb 'read' and the structure surrounding will have already been passed from the PPP to the SSS. Therefore, when the preposition is encountered, it the PPP will have no choice but to attach it to the NP 'the letter'.

From this, Fodor and Frazier conclude that the SM is superior to the competing ATN models, and that the rules of a grammar must be *explicitly represented* by the HSPM.

> [W]hen making its subsequent decisions, the executive unit of the parser refers to the geometric arrangement of nodes in the partial phrase marker that it has already constructed. It then seems unavoidable that *the well-formedness conditions on phrase markers are stored independently of the executive unit, and are accessed by it as needed*. That is, the range of syntactically legitimate attachments at each point in a sentence must be determined by *a survey of the syntactic rules for the language*, rather than being incorporated into a fixed ranking of the moves the parser should make at that particular point… (Frazier and Fodor 1978: 322n, emphasis added).

Fodor and Frazier's argument yields support for the position that we have dubbed **REP-GRAM-DATA**. It constitutes the beginning of an answer to the kinds of concerns that Devitt (2006a) voices in the following passage.

> Frazier and Fodor take their theory to support the view that a hearer mentally represents the well-formedness conditions of her language … So they are proposing a version of position (ii): rules of the language are represented and used as data. [fn. 35: Note also Fodor's later commitment to the processor being "transparent" in the following sense: it "makes use of information about linguistic structure in the form in which the mental ("competence") grammar provides it, so that statements from the grammar do not have to be modified, translated, or 'pre-compiled' before they can be applied to sentences" (1989: 177).] But it is unclear why the well-formedness conditions have to be represented and surveyed. Why could not the rules governing the "subsequent decisions" be embodied but unrepresented rules that respect (in my technical sense) the well-formedness conditions? It is hard to see what "payoff" there is in having the conditions represented. Pylyshyn's Razor counts against our supposing that they are. (*Ignorance of Language*, p. 239)

If Fodor and Frazier are right, then the explanatory payoff that we get from positing represented rules is this: The "race" aspect of the SM model accounts for the fact the attachment preferences of the HSPM conform to MA. If the rules were embodied but *un*represented, as they are in the ATN, then we would not be able to explain this important finding. Its reliance on the explicit representation of rules gives the SM model an advantage over the procedural approach of the competing ATN model.

This argument is, however, by no means conclusive. An opponent of **REP-GRAM-DATA** can challenge it by pointing out that the hardwired ATN circuit is not the only competitor to the SM model. Indeed, Devitt (2006a) holds out hope that rival models that eschew a commitment to **REP-GRAM-DATA** will eventually prove themselves superior:

> Where does this brief discussion leave our preliminary assessments? Most importantly, it does not undermine them. And I think that it gives further support to our negative view of [the Representational Thesis] RT. Clearly constraint-based connectionist approaches support our assessment that the future will be fairly brute-causal, hence support our fourth tentative proposal [that the speedy automatic language processes arising wholly, or at least partly, from linguistic competence are fairly brute-causal associationist processes that do not operate on metalinguistic representations of the syntactic and semantic properties of linguistic expressions]. These approaches look promising but it is too early in the study of language comprehension to be confident that they are right. (240)

In Chap. 6, we reviewed the difficulties facing Devitt's brute-causal associationist proposal. Still, it is possible that a procedural model can do what the ATNs failed to do—i.e., to explain the HSPM's conformity to Minimal Attachment. There are, moreover, well-known objections to the SM model (Pritchett 1992: pp. 30–40), which must be addressed before we can confidently declare Fodor and Frazier's argument for **REP-GRAM-DATA** a success.

One such objection has to do with Fodor and Frazier's commitment to a "resource-based" conception of what we earlier called the "oracle". We saw in the previous chapter that rival accounts of the HSPM's ambiguity-resolution strategies appeal to frequency information and, most intriguingly, to independently motivated syntactic principles. If these alternative accounts are correct, then the principal arguments in favor of the SM model collapse. A second objection concerns the fact that Fodor and Frazier's SM made use of the now-outdated context-free grammars. More recent versions of the SM made use of a more up-to-date grammar (McRoy and Hirst 1990), but even these have been superseded in recent decades by grammars in the Principles and Parameters framework. As we will see in the sections that follow, the parsing models that make use of such grammars can likewise underwrite an argument for **REP-GRAM-DATA**, though different from the one that Fodor and Frazier advanced.

9.4 Principles and Parameters in Syntax and Parsing

9.4.1 Government and Binding Theory

The Standard Theory of transformational grammar posited transformational rules over and above context-free phrase structure rules. These rules were construction-specific, in the sense that each separate linguistic construction—passive, raising, question, etc.—was associated with a distinct transformational rule. In addition, many such rules were language-specific. This proliferation of rules came to be seen as a serious problem. From the perspective of the syntactician, a grammar that posits such rules ranks low in elegance and simplicity. From the perspective of an acquisition theorist, a formalism that requires very different sets of rules for distinct languages makes it harder to explain how a child could ascertain the rules of the local language in a relatively short amount of time. Finally, as we saw, the "backwards-transformation" models of comprehension based on the Standard Theory were unsuccessful.

Subsequent work, eventuating in what is now known as the *Principles and Parameters* (P&P) approach heralded a striking revision of transformational grammar throughout the 1980s. The most prominent grammar to emerge from that approach was the Government and Binding (GB) theory (Chomsky 1981; Haegeman 1994). Though this theory retained the distinction between D-structure and S-structure, it replaced the myriad transformational rules of the Standard Theory with a single rule: *Move-α*. This rule says, quite literally, that any constituent appearing at D-structure can be moved anywhere in the phrase-structure tree on the way up to S-structure. Left unconstrained, this rule would generate a great many S-structure descriptions that fail to correspond to anything one finds in natural language. The substantive task, then, is to eliminate these unwanted descriptions. In the P&P model, this is accomplished by invoking a set of syntactic *principles*, which operate

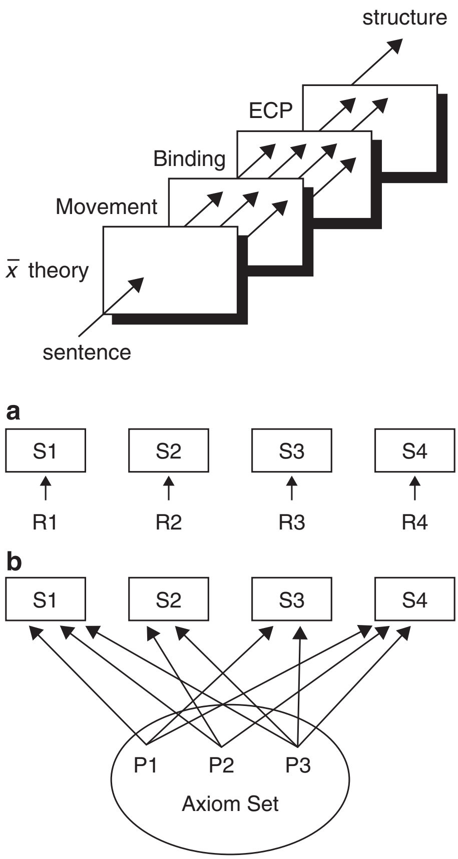


Fig. 9.4 Filters and generators. **(a)** Rules vs. **(b)** Principles (*Source*: Berwick 1991b)

in the manner of filters, ruling out ill-formed structures and retaining only the well-formed ones, as shown in Fig. 9.4.

The distinction between Standard Theory *rules* and GB *principles* is illustrated in Fig. 9.4. The principles are few in number and compact in formulation, bearing an obvious resemblance to the axioms of a deductive system.

The principles are, moreover, parameterized, typically in a binary fashion. Each distinct way of setting the parameters allows the grammar to generate the syntactic structures characteristic of one particular language. The only differences between distinct languages, then, are to be found in the lexicon and the parameter settings. Quite apart from its elegance and compactness, the resulting formalism provides guidance to the acquisition theorist, who can now view acquisition as a process of *setting parameters* in response to linguistic exemplars.[28]

[28] This, of course, requires the parameters to be either innate or acquired in some way by the child. Moreover, parameters cannot be set until the child has acquired a relatively rich lexicon, which specifies (at the very least) each word's syntactic category. See Chap. 4 for discussion.

The conception of D-structure also underwent revision. Leading up to the rise of GB, it gradually came to be understood that the syntactic descriptions pitched at D-structure need not describe strings that are actually spoken, written, or signed. That is, the phrase structure rules that form the base of the grammar are free to generate D-structure descriptions that fail to correspond to any string we might find in a corpus. The phrase marker in (40) illustrates this new kind of D-structure description.[29]

(40) $[_S$ $[_{S'}$ was $[_{VP}$ believed $[_{NP}$ the story] by the villagers]]]

Plainly, (40) is not a sentence of English. Rather, it encodes an *aspect* of the syntactic structure of a sentence—pitched at just one level of description. The full sentence is actually a 4-tuple: {Phonological Form (PF), D-structure (DS), S-structure (SS), and Logical Form (LF)}. As we saw in Chap. 5, D-structure descriptions reflect, among other things, the argument structures of verbs. For instance, the D-structure in (40) reflects the fact that the object of the verb 'believes' is the noun phrase 'the story'. Hence, D-structure descriptions do not tell us what surface forms we will find in a corpus. Rather, they tell us about the argument structure of the verbs that enter into such sentences.

Once they are generated by the base rules of the grammar, D-structure phrase markers like (40) are mapped, via the transformational rule *Move-α*, to S-structure descriptions, like (41).

(41) $[_S$ $[_{NP}$ The story$]_i$ $[_{S'}$ was $[_{VP}$ believed $[_{NP}$ NP-trace$_i$] by the villagers]]]. S-structure

As (41) illustrates, when *Move-α* takes the noun phrase from its canonical position in D-structure to its new position in S-structure, an NP-trace is left behind. This trace is co-indexed with the noun phrase, thus providing information about what semantic role the noun phrase plays in the sentence.

To summarize, the story is now this: A single sentence has *two distinct aspects of syntactic structure*—D-structure and S-structure. The phrase structure rules that generate D-structures are designed to describe an *aspect* of sentences like (42). The D-structure in (43) is then mapped onto *another* aspect of the syntactic structure of (42), viz., its S-structure (44).

(42)	**The story was believed by the villagers.**	**English sentence**
(43)	**$[_S$ $[_{S'}$ was $[_{VP}$ believed $[_{NP}$ the story] by the villagers]]]**	**The D-structure of (42)**
(44)	**$[_S$ $[_{NP}$ the story$]_i$ $[_{S'}$ was $[_{VP}$ believed $[_{NP}$ NP-trace$_i$] by the villagers]]].**	**The S-structure of (42)**

The transformation rule *Move-α* defines a relation between these two aspects of a single sentence. This stands in contrast to the transformational rules that comprised

[29] This example is adapted from Haegeman (1994), p. 306.

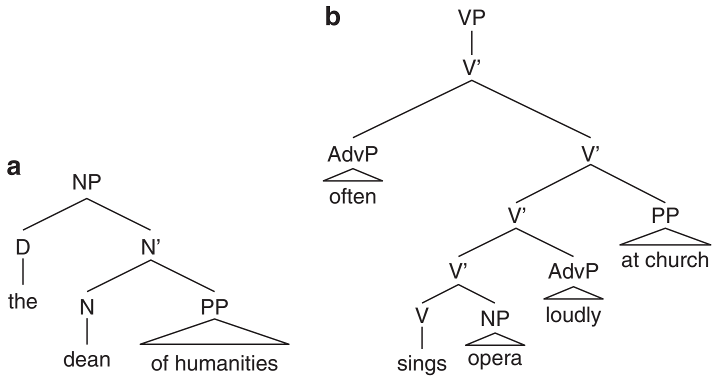


Fig. 9.5 Examples of X-bar structures for verbs and nouns. In **(a)**, the determiner 'the' is a specifier, the noun 'dean' is the head of the phrase, and the prepositional phrase 'of humanities' is the complement of the noun. In **(b)**, the adverbial phrase 'often' is the specifier, the verb 'sings' is the head of the phrase, the noun 'opera' is the complement of the verb, and the phrases 'loudly' and 'at church' are adjunct phrases.

the earliest versions of the Standard Theory, which described relations between two distinct sentences, e.g., an active and a passive.

In older versions of transformational grammar, ordinary context-free phrase structure rules generated the D-structure representations. In the P&P framework, phrase structure is assigned by set of principles, collectively known as X-bar theory. These principles can be rendered in the familiar context-free format, as the rules below illustrate. (Parentheses denote optional elements.)[30]

NP → (D) N′	VP → V′	AdjP → Adj′	PP → P′
N′ → (AdjP) N′	V′ → (AdvP) V′	Adj′ → ({AdvP/AdjP}) Adj′	P′ → (AdvP)
N′ → N′ (PP)	V′ → V′ ({AdvP/PP})	Adj′ → Adj′ (PP)	P′ → P′ (PP)
N′ → N (PP)	V′ → V (NP)	Adj′ → Adj (PP)	P′ → P ({NP/PP})

Figure 9.5 illustrates the X-bar structure of an NP and a VP. Fig. 9.6 makes use of a somewhat more abstract notation to illustrate the various types of structures generated by X-bar theory.

What is distinctive about X-bar theory is that "bar nodes," like N′ and V′ in Fig. 9.5, introduce additional structure into phrases. And the positions defined in terms of these structures can be used to distinguish between *phrasal heads*, *specifier phrases*, *adjunct phrases*, and *complement phrases* (Fig. 9.6). This allows X-bar theory to correctly predict various kinds of constituency relations within phrases. For instance, the structure in Fig. 9.6b allows for a natural explanation of the fact

[30] Berwick (1991b: pp. 195–203) discusses the virtues of parsing with X-bar theory "directly," i.e., without first translating X-bar principles into a set of context-free rules. Using this technique, Berwick claims to be able to achieve a psychologically plausible *deterministic* parse, *without lookahead*—plainly a vast improvement over the original deterministic parser presented in Marcus (1980).

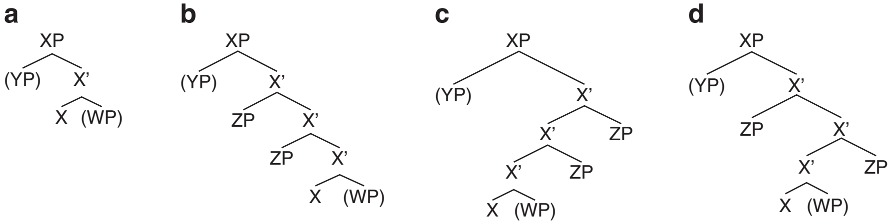


Fig. 9.6 The phrase structure type in **(a)** is the most basic X-bar structure. It has slots for the specifier phrase (YP) and a complement phrase (WP). The structures in **(b)-(d)** have additional slots for adjunct phrases (ZP). The differences between **(b)**, **(c)**, and **(d)** all have to do with the location of the adjunct phrases (ZPs).

that various parts of the verb phrase in sentence (45) can be replaced by 'does so' in sentences (46)–(48).[31]

(45)	**Alia often [sings opera loudly at church].**	The VP ('sings opera loudly at church') is bracketed.
(46)	**Erika [does so], too.**	'Does so' replaces 'sings opera loudly at church'.
(47)	**Jennifer often [does so] at sporting events.**	'Does so' replaces 'sings opera loudly'
(48)	**Oriana [does so] quietly at work.**	'Does so' replaces 'sings opera'

Each phrasal head—i.e., each noun, verb, adverb, adjective, and preposition—projects an X-bar structure like those in Figs. 9.5 and 9.6. These are then linked together to generate phrase structure trees for entire sentences, yielding D-structures descriptions. As discussed above, the transformational rule *Move-α* applies to a phrase marker at D-structure, moving around any element of the structure, thus generating an array of S-structure phrase markers. A variety of principles then serve to filter out some of these and retain others, as illustrated in Fig. 9.4.

One such principle is the *Empty Category Principle* (ECP), which specifies the structural configurations in which empty categories can appear in well-formed phrase markers. Another set of principles constitutes the *Binding Theory*, which specifies the permissible relations between pronouns and their antecedents. These serve to rule out structures like (49)–(50).

(49)	***Jeremy$_i$ believes Dan likes himself$_i$.**	Coindexing 'Jeremy' and 'himself' violates Binding Theory.
(50)	***Who$_i$ do you think that *wh*-trace$_i$ will like me?**	Coindexing 'what' and '*wh*-trace' violates the ECP.

[31] See Haegeman (1994) for a motivation of X-bar theory and the details of various constituency tests, like the 'does so' test that is demonstrated in the main text.

Another principle is the *Theta-Criterion*. This principle ensures that the argument structure (i.e., the "θ-grid") of each verb is satisfied at D-structure and beyond. For instance, the verbs 'sleep', 'wear', and 'put' take one, two, and three arguments respectively. In the terminology of θ-theory, these verbs have one, two, and three "θ-roles" in their respective "θ-grids." This is demonstrated by the following examples.[32]

(51) **David is sleeping. / *David is sleeping the blanket. / *David is sleeping the blanket on the table.**
(52) ***David is wearing. / David is wearing the shirt. / *David is wearing the shirt on the table.**
(53) ***David is putting. / *David is putting the book. / David is putting the book on the table.**

The θ-Criterion rules out the ill-formed structures, starred above. Finally, a principle known as the Case Filter ensures that all overt nouns are assigned Case. According to this principle, a noun can only be assigned Case in certain structural configurations. The Case Filter rules out structures like (54), in which the subject is not assigned Case.

(54)	***It is possible Eugene to play soccer.**	The noun 'Eugene' cannot receive Case in this configuration.

Figure 9.7 summarizes the GB principles that assign syntactic structures to sentences, beginning with the initial structural assignments allowed by X-bar theory, proceeding through the various filters described above, among many others, and eventuating in a well-formed phrase marker. The principles are partially ordered—the application of some requires the prior application of others. For instance, as Fig. 9.7 illustrates, the Case filter cannot apply until movement operations have applied.[33]

A grammar consisting of X-bar theory, movement operations, and filtering principles can be stated in a compact axiomatic form (Johnson 1989, 1991; Berwick 1991a,b; Stabler 1992), This accounts for its elegance and increased explanatory power over older transformational grammars. However, to repeat an earlier point, neither the compactness nor the explanatory power of a grammar can constitute grounds for ascribing psychological reality to the rules or principles of that grammar (Stabler 1984). Independent arguments must be supplied even for seeing the

[32] The examples given here gloss over a great many of subtleties. In the interest of brevity, I shall not enter into a digression concerning the syntactic arguments for the distinction between arguments and adjuncts, nor the tests for deciding whether a seemingly complete sentence is elliptical for a one that contains additional arguments in its overt form. These are delicate matters and a full discussion would take us well beyond the scope of the present work.

[33] Here, I am ignoring so-called "inherent" case-marking, which was hypothesized to take place at D-structure.

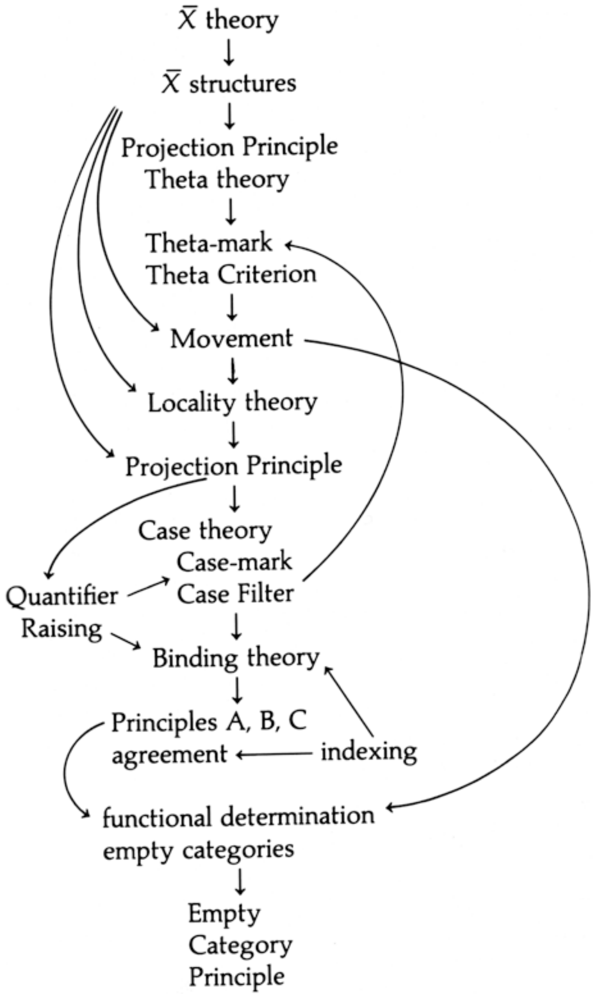


Fig. 9.7 The GB principles that assign syntactic structures to sentences, beginning (at the top) with the initial structural assignments allowed by X-bar theory, proceeding through the various filters described above (Binding, ECP, Theta Criterion, etc.), and eventuating in a well-formed phrase marker. The principles are partially ordered, such that the application of some requires the prior application of others. *Source*: Berwick (1991b).

phrase markers that are generated by a grammar as psychologically real. It should be even more evident that the *rules or principles* of a successful grammar are not automatically "in the mind".

In Chap. 5, we saw reasons to believe that S-structure descriptions should be seen as psychologically real representations. By contrast, D-structure descriptions did not fare so well. Recall the experiment reported in Bock et al. (1992), the results of which cast doubt on the claim that the HSPM constructs D-structure representations for the purposes of comprehension and production. The principle-based parsers described in the following section comport with this conclusion. In describing his model, Johnson (1989) rejects the claim that a "transparent implementation" of GB theory in a parser requires constructing representations of the D-structure of the input.

> Use of knowledge of a level of representation does not entail the explicit construction of that representation, as I demonstrate by exhibiting a parser that uses knowledge of D-structures in the parsing process yet avoids the explicit construction of this level of representation. (p. 105)

In connection with the psychological reality issue, we must ask whether principles-based parsers incorporate representations of the GB rules and principles, and if so, how. We will be in a position to address these questions after we've familiarized ourselves with their main features and their algorithmic operations. Let's now turn to those.

9.4.2 Principles-Based Parsing

With the rise of the P&P approach to language acquisition, and the descriptive successes of GB theory, it was only natural that computational linguists would embark on the project of implementing these novel ideas in parsing models. In what follows, I present the main features of a number of efforts in this direction.

One striking feature of principle-based parsers is their flexibility with regard to different languages. Yang and Berwick (1996) point out that "[t]raditional parsing technologies utilize language-particular, rule-based formalisms, which usually result in large and inflexible systems (Marcus 1980)." By contrast, the principle-based parser that Yang and Berwick designed can be used to assign structure to sentences from a variety of typologically distinct languages. To switch from one language to another, all the parser needs to do is swap out one lexicon for another and set the parameters to the values characteristic of the target language. For instance, an X-bar parameter can be changed from the head-initial setting (as in, e.g., English) to a head-final setting (as in, e.g., Japanese). Similarly, the Case-filter can be set so that only the members of some desired class of phrases are capable of assigning Case (in particular configurations).

> It is believed that languages are constrained by a small number of universal principles, with linguistic variations largely specified by parametric settings. The merit of principle-based parsing is two-fold. As a tool for linguists, it is directly rooted in grammatical theories. Therefore, linguistic problems, particularly those that involve complex interactions among linguistic principles, can be cast in a computational framework and extensively studied by drawing directly on an already-substantiated linguistic platform. It is designed from the start to accommodate a wide range of languages — not just 'Eurocentric' Romance or Germanic languages. Japanese, Korean, Hindi and Bangla have all been relatively easily modeled in PAPPI (Berwick and Fong 1995, Berwick forthcoming). Differences among languages reduce to distinct dictionaries, required in any case, plus parametric variation in the principles. (Yang and Berwick 1996)

This kind of flexibility is particularly impressive when the languages are radically different from one another.

Berwick (1991b) discusses the advantages of a principle-based parser for the free word-order languages, Warlpiri and Japanese. According to him, attempts to build parsers and translation systems for these languages on the basis of context-free phrase structure rules were discovered to be inefficient and ultimately unsuccessful (pp. 163–170). The problem was that context-free rules fuse together information about two distinct linguistic properties—hierarchical dominance and linear precedence. Though X-bar theory was presented above as a kind of context-free phrase structure grammar, Berwick discusses ways of implementing principle-based parsers without appealing to context-free rules (see below). He also demonstrates that the principle-based parser designed for Warlpiri performs much better than older models, which were designed to use context-free grammars. The following results should give an idea of the impressive progress that principle-based systems have made in achieving broad coverage within an interesting variety of languages.

> We are able to parse sentences with the range of structures including Wh-movement, the Binding Theory, Quantifier Scoping, the BA-construction to complex NP (clausal, possessive, and numeral/ classifier). All testing sentences are correctly analyzed: LF logical form representations are computed for the grammatical sentences and the ungrammatical ones are ruled out ones with linguistic principle violation(s) shown. Each parse takes no more than 2 seconds on a Sparc10 workstation. Overall, excluding a hand-wired dictionary, less than 100 additional lines of Prolog are required. Because the PAPPI system implements its model linguistic theory faithfully, adapting new languages is expected to be quite minimal, as our implementation shows. Additionally, it provides a platform on which linguists can experiment theoretical proposals extensively and also cross-linguistically, without having to know much about the internal design of the parser. Furthermore, principle-based systems output very rich and accurate structural descriptions, including logical form representations, that assist in more engineering-oriented NLP tasks that go beyond parsing (Yang and Berwick 1996: p. 370).

Besides their flexibility with respect to distinct languages, principle-based parsers exhibit a tolerance for ungrammatical input. Indeed, a binary notion of grammaticality is difficult to maintain in the P&P framework. On its way from X-bar analysis to the ultimate assignment of structure, linguistic input passes through a series of filters (Figs. 9.4 and 9.7). Almost every input is assigned *some* interpretation as it runs this gauntlet. Throughout the process, the parser takes note of any failure on the part of the input to satisfy one or another principle of the grammar. In the human case, this would give rise to variable-strength judgments of ungrammaticality; the more principles are violated, the more ungrammatical the input will be judged to be. This stands in contrast with the behavior of parsers based on context-free grammars and ATN grammars, discussed above.

For our purposes, one of the most interesting facts about principle-based parsers is that they make heavy use of the Parsing as Deduction (PAD) approach. And, as the theorists who develop these models are eager to point out, the combination of GB theory and the PAD yields a satisfying interpretation of Chomsky's talk of "knowledge of language."

> the human language processor uses its knowledge of a language to obtain knowledge of the utterances of that language in the "same way" that a theorem-prover "uses" axioms to deduce their consequences. … My primary goal here is to demonstrate that the Parsing as Deduction hypothesis does in fact provide a viable explanation for how it is that knowledge of a language of the form attributed to human language users by modern GB theory

> (Chomsky 1981; Van Riemsdijk and Williams 1986) can actually be used to obtain knowledge of the utterances of that language. (Johnson 1989: pp. 105–106)

> As Johnson's paper argues, deductive inference is still perhaps the clearest way to think about how to 'use' knowledge of language. In a certain sense, it even seems straightforward. The terms in the definitions like [that of the Case filter] have a suggestive logical ring to them, and even include informal quantifiers like *every*; terms like lexical NP can be predicates, and so forth. In this way, one is led to first-order logic or Horn clause logic implementation (Prolog) as a natural first choice or implementation, and there have been several such principle-based parsers written besides those described by authors of this volume who have built Prolog implementations of principle-based theories (see Sharp 1985; Kolb and Thiersch 1988). Parsing amounts to using a theorem prover to search through the space of possible satisfying representations to find a parse… (Berwick 1991a)

In the latter passage, Berwick refers to the idea of formalizing GB principles in first-order logic. Understanding how this can be done is crucial to making sense of how the PAD approach can be applied to GB theory. Let's look at a concrete example.

One of the "filters" described earlier is the Empty Category Principle (ECP), which can be stated as follows: *Empty categories must be properly governed.* Formulated in this way, the ECP contains technical terms that stand in need of definition. Passing over some complex issues (Haegeman 1994; Carnie 2010), we can make do with the following:

<u>Proper government</u>:	X properly governs Y if and only if X governs Y and X is a lexical category (i.e., N, V, P, Adv, Adj).
<u>Government</u>:	X governs Y if and only if Y is contained in the maximal X-bar projection of X (X-max), and X-max is the smallest maximal projection containing Y, and X c-commands Y.
<u>C-command</u>:	X c-commands Y if and only if the first branching node dominating X also dominates Y, and X does not itself dominate Y.

When made precise, the relations defined above can be expressed in first-order logic, as follows. (Here, the symbol '⇔' expresses the material biconditional.)

<u>Proper government</u>:	*properly-governs(X, Y) ⇔ governs(X, Y) & lexical-category(X)*
<u>Lexical Category</u>:	*lexical-category(X) ⇔ Noun(X) ∨ Verb(X) ∨ Prep(X) ∨ Adv(X) ∨ Adj(X)*
<u>Government</u>:	*governs(X, Y) ⇔ maximal-projection(X, Xmax) & dominates(Xmax, Y) & least-maximal-projection(Y, Xmax) & c-commands(X, Y)*
<u>C-command</u>:	*c-commands(X, Y) ⇔ first-branching-node-from(X, BrNode) & dominates (BrNode, Y) & ~ dominates(X, Y) & ~ dominates(Y, X)*

This, in turn, allows us to formulate the Empty Category Principle as follows:

(ECP) *∀Y[empty-category(Y) . ∃X(properly-governs(X, Y))]*

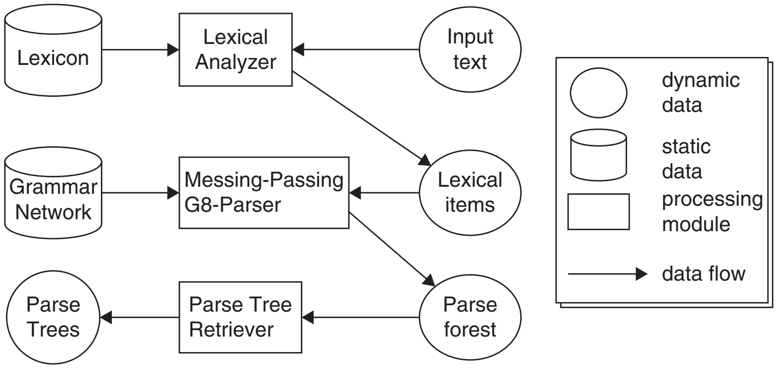


Fig. 9.8 An information-flow diagram of the PRINCIPAR system developed by Lin (1993, 1994). Lin annotates it as follows: "Sentence analysis is divided into three steps. The lexical analyser first converts the input sentence into a set of lexical items. Then, a message passing algorithm for GB-parsing is used to construct a shared parse forest. Finally, a parse tree retriever is used to enumerate the parse trees" (1994, p. 482). Note the explicit characterization of the grammar as a static data structure, which is consulted/accessed by the parser. This conforms to the view that we've labeled **REP-GRAM-DATA**

With this definition in hand, developers of principle-based parsing systems can use a Prolog or LISP script to scan a set of candidate phrase markers generated by the X-bar module, and to assign partial ungrammaticality to those that fail to satisfy the definition. Berwick provides another example of this strategy, using a definition of the Case filter, stated directly in Prolog.

> How then can a theory written in English be mapped into logic? Here for example, following Fong (1991, forthcoming) is a typical statement of the Case filter: "At S-structure, every lexical NP needs Case." How are we to implement this statement as part of a parsing program? … We must actually translate the English connectives such as needs or every and the required feature checking (the property of being *lexical* on NPs), into logical formulas. But in fact this can readily be done. (Berwick 1991a: pp. 14–15)

These examples of the ECP and the Case Filter illustrate what theorists like Yang and Berwick (1996) have in mind when they claim that "A principle-based parser *transparently reflects* the structure of the contemporary linguistic theory, the Principles and Parameters framework" (p. 363, emphasis mine).

We are now in a position to examine the computational role of GB principles in the kinds of parsers that have been implemented thus far. Figure 9.8 illustrates the flow of information through PRINCIPAR, a principles-based parsing system developed by Lin (1993 1994). Inspecting the diagram, we see that Lin characterizes the grammar as a *static data structure*, which is consulted or accessed by the parser. This conforms to the view that we've been calling **REP-GRAM-DATA**, according to which the structure rules of a language are represented and used *as data* by the processing rules governing the parser.

Complicating matters somewhat, Berwick (1991b) discusses a number of ways in which the GB grammar can be related to a principles-based parser. He writes:

> It seems, then, that there is at least a strong and a weak sense of how a system might "use" a set of principles. In a strong sense, the principles might be literally represented as such, individually, and used directly online as data structures for parsing. For example, this is roughly how the rules of a context-free grammar might be used in, say, Earley's algorithm… In a weaker sense, the principles might be *interpreted*.

In this passage, Berwick lends support to the idea that the Earley parsing algorithm is strongly suggestive of **REP-GRAM-DATA**. However, his talk of a "weaker sense" in which the principles are "interpreted" might seem to cast doubt on this idea. Further investigation reveals that it does not. For, when Berwick goes on to discuss what might be meant by "interpreted," his primary example is a logical procedure known as partial evaluation. A partial evaluation of the predicate '*properly-governs(X, Y)*' is given in (PG).

(PG) *properly-governs(X, Y) ⇔ maximal-projection(X, Y) & dominates(Xmax, Y) & least-maximal-projection(Y, Xmax) & first-branching-node-from(X, BrNode) & dominates (BrNode, Y) & ~ dominates(X, Y) & ~ dominates(Y, X).*

It should be clear that "partial evaluation" is nothing more than the logical combination (in this case, simple conjunction) of various distinct principles and relations defined by the grammar. Hence, Berwick's distinction between the strong and weak senses in which a parser can use a grammar has no bearing on whether the rules or principles are explicitly represented and used as data. It has to do, instead, with the *logical format* in which they are represented in the parser, and the extent to which this corresponds to how they are represented in linguistic textbooks.

Experiments in computational linguistics confirm that the logical format of the data structure in which the GB principles are represented can have significant consequences for the efficiency of the parser. Using a slightly different terminology to express the same ideas, Merlo (1995) discusses in detail her experimental efforts to increase the efficiency of a principle-based system by "precompiling" the grammar, thus "multiplying out" the interactions of the principles. She writes: "By compilation … I mean off-line computation of some general property of the grammar, for example the off-line computation of the interaction of principles, using partial evaluation or variable substitution" (p. 517, fn. 2). Merlo concludes that (pre)compilation is useful up to a point, but the more of it one does, the more rules one creates. And the more rules the system has to apply, the longer it takes to return a parse. So there is a point at which compiling actually does more harm than good. The PRINCIPAR system discussed above offers a novel approach to the efficiency problem, which, incidentally, provides support to the claim that MPMs are themselves syntactically structured (ch. 6). Lin (1994) writes:

> Principle-based grammars, such as Government-Binding (GB) theory (Chomsky 1981; Haegeman 1994), offer many advantages over rule-based and unification-based grammars, such as the universality of principles and modularity of components in the grammar. Principles are constraints over X-bar structures. Most previous principle-based parsers, e.g., (Dorr 1991; Fong 1991; Johnson 1991), essentially generate all possible X-bar structures of a sentence and then use the principles to filter out the illicit ones. The drawback of this approach is the inefficiency due to the large number of candidate structures to be filtered out. The problem persists even when various techniques such as optimal ordering of

> principles (Fong 1991), and co-routining (Dorr 1991; Johnson 1991) are used. … This paper describes an efficient, broad coverage, principle-based parser, called PRINCIPAR. The main innovation in PRINCIPAR is that it applies principles to *descriptions of X-bar structures rather than the structures themselves*. X-bar structures of a sentence are only built when their descriptions have satisfied all the principles. (p. 482, emphasis added)

Unsurprisingly, recent approaches to principle-based parsing seek to increase efficiency by drawing on probabilistic information. Berthouzoz and Merlo (1997) conducted a series of comparative studies, which yielded the following results:

> We have presented some experiments in evaluating the performance of a symbolic parser for English (IPS). Due to its massively parallel nature, IPS has good recall on the set of sentences that we tested, but precision is very low, showing that overgeneration of constructions needs to be reduced. We have tested two methods. In the structural method, both the generating and the filtering device are based on knowledge stored in a linguistically-defined grammar. In the hybrid method, the generating device is structure-based but the filtering is done based on statistical information extracted from a corpus. We observe that the statistical method outperforms the structural approach, that it is less computationally expensive, and it is easier to define and develop. (p. 185)

A detailed discussion of these approaches would take us to far afield. Suffice it to say that issues of efficiency are being addressed head-on.

Let us sum up: P&P grammars are, in many respects, improvements over previous transformational grammars. In addition to their formal virtues, they suggest a novel framework within which to approach questions about both the acquisition and the processing of language. In the processing domain, they have been used in systems that deliver fast, reliable, and detailed parses of an impressively broad range of natural-language constructions, in a variety of typologically distinct languages. These parses are, moreover, fine-grained enough to induce degrees of grammaticality, thereby replicating a key feature of human performance. Finally, the systems in question make use of the PAD framework in fleshing out the oft-repeated claim that knowledge of syntax is put to use in language processing. The most common way in which they do so is by explicitly representing grammatical rules and principles and drawing on them as data, in accordance with the position that we've been calling **REP-GRAM-DATA**.

We now turn to the most recent development in the P&P framework—i.e., what Chomsky (1995) has called the Minimalist program. Like the Government and Binding theory, Minimalism has had a profound effect on psycholinguistics. We will see that parsing in accordance with Minimalist grammars is often a matter of directly applying the ground-level operations of *Merge* and *Move*. As with the ATN framework discussed in Sect. 9.3, Minimalism provides fresh support for **REP-GRAM-PROC** and **EMB-GRAM-PROC**.

9.5 Minimalist Grammars

9.5.1 Minimalism: The Basics (Features, Merge, Move, and Spell-Out)

The Minimalist Program, as the name suggests, is an attempt to achieve the explanatory power of GB, but with more modest theoretical machinery—ideally, the bare minimum. The central notions of GB, viz., D-structure and Government, are argued to be superfluous. Their explanatory functions are absorbed by components of the grammar that *all* theorists must posit—i.e., the lexicon and a recursive structure-building operation. I will not attempt to motivate the central claims of Minimalist grammars here. My aim is simply to sketch those claims, and proceed directly to a discussion of their role in parsing theory.

As noted above, Minimalist grammars give a prominent role to the lexicon. The grammar conceives of the lexicon as the starting point of a *derivation*—a process that begins when a batch of lexical entries are delivered to the syntactic component of the grammar. The syntax is seen as nothing more than a conduit between the lexicon and both the phonological component (PF) and the "conceptual-intentional" component (CI) of the grammar. Lexical entries are treated as *bundles of features*. Some features can only be "read" by PF, others only by CI. PF and CI thus impose "legibility conditions" on the features of an expression; in order to be read by those components, the lexical features must be "interpretable" by them. Other features, however, are only interpretable within the *syntactic* component—hence *un*interpretable by PF or CI. The syntactic component of the grammar must, therefore, operate so as to output *only* the interpretable features to PF and to CI. It is the only component of the grammar that is a theory-internal posit, PF and CI being obviously necessary to characterize vocalization and thought. For this reason, parsimony dictates that its role be minimized to the extent possible. This gives rise to the "Economy Conditions" discussed below.

How, then, can the syntax rid expressions of the purely syntactic features that they get from the lexicon? The answer is: *feature checking*—a process wherein lexical items that "need to receive" some feature enter into a relation with lexical items that "need to get rid of" that very feature. Once "checked," a purely syntactic feature is deleted, hence no longer problematic for the PF or CI components. By contrast, purely syntactic features that are left unchecked engender problems at PF or CI, causing the derivation to crash.

The kinds of features that a lexical item might have include its parts of speech (complementizer, tense marker, verb, determiner, noun, adjective, etc.), its subcategorization and argument-selecting theta-features, Case-assigning or Case-requiring features, agreement features for person, number, and gender, and others. In symbolic form:

Part of speech:	{c, t, v, d, n, a,…}
Argument selecting features:	{=x, x=, =X, X=}[34]
Licensors (the providers):	{+case, +wh, + focus…}
Licensees (the needy):	{−case, −wh, −focus…}

Here, then, are some concrete examples of lexical items[35]:

Word	Feature Bundle	Interpretation of the feature symbols
The	{d, =n, − case}	(determiner, selects a noun, needs case)
student	{n}	(noun)
wrote	{*v*, =d, + case, =d}	(verb, selects 2 DPs, assigns case to one DP)
a	{d, =n, − case}	(determiner, selects a noun, needs case)
dissertation	{n}	(noun)

Now that we have a lexicon in place, we need to implement the feature-checking operation discussed above. This must be an operation that relates the features of two lexical items, and it must be recursive, so as to account for the productivity of natural language. The simplest possible operation that meets these conditions is *Merge*, which takes two lexical items X and Y, and places them in a structure, {X, Y}, that inherits the features of its members, some of which can then enter into a checking relationship. Because *Merge* is recursive, it can take as input its own outputs, combining {X, Y} with W to form the object {W, {X, Y}}, and then with Z to form {Z, {W, {X, Y}}}, and so on. In this way, hierarchical structures are generated. (We can represent these, as usual, with binary-branching trees, as in Fig. 9.9.) Here is a simple example, using the lexical items displayed above.

	Words	**Feature Structures**
Input to *Merge*:	'wrote', 'a'	{v, =d, +case, =d}, {d}
Output of *Merge*:	'wrote a'	{v, =d}

When the two feature bundles (i.e., words) are merged, the {d} feature of the determiner 'a' "checks" the {=d} feature of the verb 'wrote'. Informally, it's as though 'wrote' calls out and says "I need a determiner," and the determiner 'a' comes along to supply that feature. The result is a structure that has features {v, =d}, i.e., a verb that "needs" one more determiner. The example in Fig. 9.9 illustrates the same idea,

[34] Note that the direction of the symbol '=' indicates whether the item needs its features checked on its left or on its right. Note also that the features rendered in capital letters are "Strong," while those rendered in lower-case are "Weak"—a distinction that will play no role in subsequent discussion. To use language that we have not yet introduced, the Strong features, unlike the Weak ones, require *phonological* features to be displaced within the structure. (I discuss displacement shortly.)

[35] I am grateful to an anonymous reviewer for pointing out that the categories in the box are Stabler's, where "wrote" is VSO and the SVO order is derived by Kaynesian LCA movement, whereas the unboxed symbols are Harkema's notation, where "wrote" starts out as SVO.

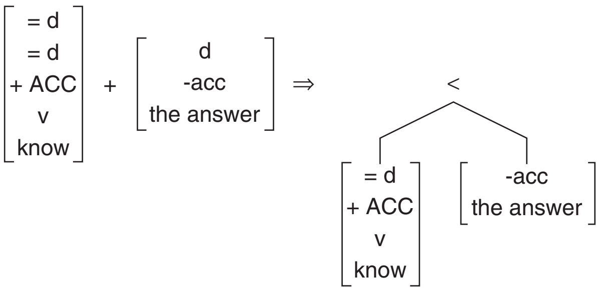


Fig. 9.9 The two feature bundles associated with 'know' and 'the answer' are merged, creating a tree-structure whose head is 'know' (as indicated by the direction of the '<' pointer). *Merge* is triggered by the need to check the {=d} feature on 'know' and the {d} feature on 'the answer'. *Source*: Gerth and beim Graben (2009)

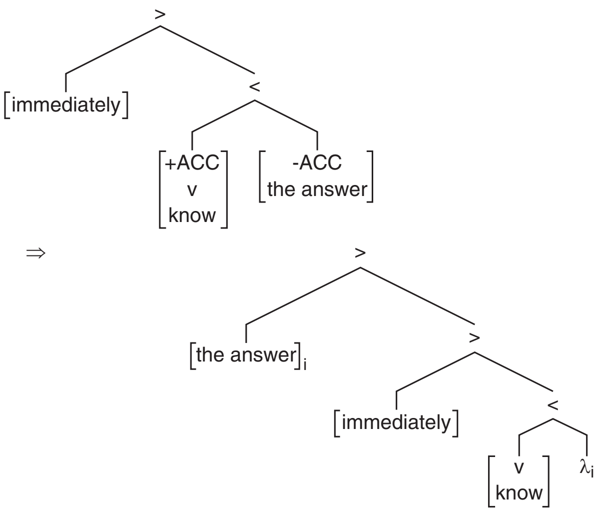


Fig. 9.10 The expression 'the answer' is moved from its most-embedded in the tree position (top panel) to its new least-embedded position (bottom panel), leaving behind a trace (λ_i) in its initial position. The movement results in a checking of the feature {−ACC} on 'the answer' and the feature {+ACC} on 'know'. (Gerth and beim Graben 2009)

except also introduces the pointers '>' and '<' to indicate the location of the head of the phrase.

Minimalist grammars posit another operation, *Move*, whose function is to displace the most-embedded element in the structure, in order to check its syntactic features before any further *Merge* operations can apply. Figure 9.10 provides an example in which the phrase 'the answer' moves up in the tree, in order to check its Accusative Case features {−acc} against the feature {+ACC} on 'know'.

A central tenet of the Minimalist Program is its insistence on what Chomsky calls "Economy Conditions." In GB, the rule *Move-α* applied to D-structure trees indiscriminately, generating a batch of structures, most of which are then filtered out by the principles of the grammar. From the Minimalist perspective, this is a profligate system. Minimalist grammars take the opposite approach: nothing is generated or operated upon unless something else makes it absolutely necessary. And, even then, the most economical operation is the only one that applies. (This is the principle of Last Resort.) Economy thus dictates that *Merge* and *Move* are always and only triggered by the need to check features.

Another example of economy constraints has to do with how the syntactic component of the grammar outputs material to PF—an operation known as *Spell-Out*. Given that we speak, sign, and write in linear structures that do not explicitly reflect hierarchical relations, the output of the syntax to PF is a syntactically unstructured string. As Weinberg (1999) puts it, "Spell-Out turns a syntactic structure with relevant constituent relationships into a string ready for phonological interpretation" (p. 287). The question arises: When in the derivation does Spell-Out take place? And how often? Uriagereka (1999) argues that Spell-Out applies whenever neither *Merge* nor *Move* can apply in the course of a derivation.[36] Weinberg (1999) capitalizes on this idea in providing an account of ambiguity resolution preferences, which we will examine below. Her account relies crucially on the following idea:

> Within the context of the Minimalist Program, Spell-Out is a grammatical operation, on a par with movement transformations. As such, it is governed by conditions on transformations, in particular by the economy conditions discussed above. These conditions establish a preference for derivations that utilize the fewest operations possible. An operation is applied only to satisfy some independent grammatical condition. In this case, this means that the system spells out or linearizes only when it cannot otherwise establish a chain of precedence. (p. 288)

We have now introduced the basic components of Minimalist grammars: the syntax, PF, CI, legibility at the interfaces, lexical features, feature checking, *Merge*, *Move*, Economy conditions, and Spell-Out. We end this crash course by noting that the P&P component of the Minimalist grammars consists of various principles regarding feature checking, and language-specific parameters that specify which functional categories—Tense Phrase (TP), Agreement Phrase (AgrP), and so forth—will carry which features. Given the prominent role that features play in

---

[36] Weinberg (1999) summarizes Uriagereka's argument as follows: "Uriagereka uses Spell-Out as a repair mechanism to retain one-to-one correspondence between dominance and precedence. He assumes that both precedence and dominance must be established between terminal elements at all points of the derivation. Precedence implies merger, and merger is only possible when a chain of dominance can be established. When merger is not possible, the string is linearized (turned into an unstructured string where only previously established precedence relations are preserved). Since the elements that have been linearized are invisible in the syntax, precedence does not have to be established between them and other items in the structure. Thus, when two categories cannot be combined through merger or movement (the only syntactic operations) to form a dominating category, the material that has been given structure so far is "spelled out" or linearized… This idea preserves the one-to-one mapping between precedence and dominance, but only at the cost of never building single phrase markers. Instead, the system builds blocks…where all elements stand in a c-command relation to each other. When this c-command relation is interrupted, the unit is spelled out…" (pp. 287–8).

determining hierarchical structure, it turns out that it may well be possible to characterize the differences between all languages in just these terms—a very exciting prospect.

In formal syntax, the Minimalist Program was underway in the early 1990s. As with the other frameworks that we've examined in this chapter, it took about a decade for Minimalist grammars to make an impact on theories of sentence processing. In what follows, we will look first at a simple application, due to Berwick (1997). Berwick sees *Merge* as corresponding to the reduction step in shift-reduce parsers. It seems clear that this idea will be the core of any Minimalist parser. But this doesn't amount to an explicit parsing algorithm or tell us much about how to deal with phrasal movement and gaps. It also sheds no light on our central question, regarding how (or whether) the grammar is mentally represented. To fill in these details, we look to Harkema (2001), who provides a detailed algorithmic treatment, within the PAD framework that we discussed above. Harkema specifies a top-down, a bottom-up, and an Earley parser for Minimalist grammars, and proves that these can run in polynomial time.[37] Finally, we consider a proposal due to Weinberg (1999), who appeals to Minimalist principles in explaining a range of psycholinguistic data pertaining to the HSPM's ambiguity-resolution preferences. If successful, this proposal would constitute the most powerful argument for the psychological plausibility of Minimalist parsers and would go a long way toward establishing either **REP-GRAM-PROC** or **EMB-GRAM-PROC** with respect to Minimalist grammar.

9.5.2 Parsing with Minimalist Grammars

Berwick (1997) treats *Merge* as corresponding to the reduction step in shift-reduce parsers. Recall that such parsers consist of a working memory stack that holds a small number of input items, as well as the results of prior parsing operations—i.e., prior "reductions," which merge two items from the input stream. Berwick illustrates this scheme in Fig. 9.11.

Although this is not represented in Fig. 9.11, each of the lexical items is a feature bundle. The parser reduces the items in the memory buffer in accordance with their feature structures, checking syntactic features with every merger. Unfortunately, Berwick does not discuss the details of the algorithm, and says nothing about how the parser would handle input that contains gaps—e.g., wh-questions or relative-clause constructions.

Harkema (2001) supplies these details, working within the Parsing as Deduction (PAD) framework. As we've seen, the first step in applying this framework is to

[37] Although Harkema does not implement his algorithms, Fong (2011) and Gerth and beim Graben (2009) discuss up-and-running systems that make use of Minimalist grammars. Fong's parser is described, with visual aids, here: http://dingo.sbs.arizona.edu/~sandiway/mpp/mons2011.pdf Gerth and beim Graben's parser is a connectionist system. They even go some way toward demonstrating the psychological plausibility of their model.

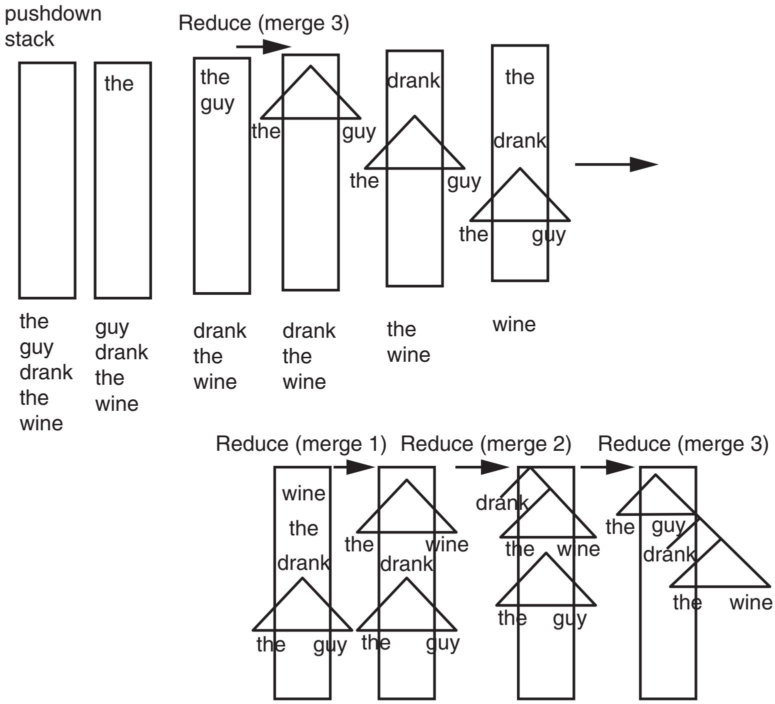


Fig. 9.11 A bottom-up parse of the sentence *The guy drank the wine*. The vertical bars constitute the contents of the system's working memory buffer at each step of the parse. At first, the buffer is empty. In the first two steps, the items 'the' and 'guy' are shifted onto the stack. In the next step (top panel), the operation *Merge* applies, creating a simple hierarchical structure. In the final step, two such structures are merged, thus completing the parse

formalize the rules of a grammar as inference rules for the purposes of formal deduction. In doing so, Harkema draws a distinction between different varieties of the transformational rules *Merge* and *Move*, yielding the five transformation types illustrated in Fig. 9.12. This division of two operations, *Merge* and *Move*, into five operations constitutes a departure from the fully "transparent" implementation of Minimalist grammar. But the relation between Harkema's grammar and what we might call "pure" Minimalism is nevertheless very tight.

The inference rules in Fig. 9.12 rules take lexical items as input and yield others as output. To take an example, the transitive verb *praise*, which has a vt feature and a –v agreement feature, merges with the noun *Lavinia*, to check their respective {=d} and {d} features. The result is what we see in Fig. 9.13.

Repeated applications of Merge and Move comprise the derivation encoded in the tree diagram in Fig. 9.14, which Harkema discusses in the following passage.

> The bottom-up, top-down, and Earley-style parsers that will be presented in the remainder of this dissertation are defined on derivation trees of the sort given in [Fig. 9.14]. A bottom-up parser will start from the lexical items at the leaves of the tree, which are provided by the words of the sentence to be parsed, and try to derive the expression at the root of the tree by repeatedly applying the functions Merge-1 through Move-2. A top-down parser will start from the expression at the root of the derivation tree and, by repeatedly applying the func-

Merge-1:

$$\frac{[(\mathrm{p},\ \mathrm{q}){:}{=}\mathrm{x}\gamma]_s \qquad [(\mathrm{q},\ \mathrm{v}){:}\mathrm{x},\ \alpha_1, \ldots, \alpha_k]_t}{[(\mathrm{p},\ \mathrm{v}){:}\gamma,\ \alpha_1, \ldots, \alpha_k]_c}\ \text{Merge-1}$$

Merge-2:

$$\frac{[(\mathrm{p},\ \mathrm{q}){:}{=}\mathrm{x}\gamma,\ \alpha_1, \ldots, \alpha_k]_c \qquad [(\mathrm{v},\ \mathrm{p}){:}\mathrm{x},\ \iota_1, \ldots, \iota_l]_t}{[(\mathrm{v},\ \mathrm{q}){:}\gamma,\ \alpha_1, \ldots, \alpha_k,\ \iota_1, \ldots, \iota_l]_c}\ \text{Merge-2}$$

Merge-3: $(\delta \neq \emptyset)$

$$\frac{[(\mathrm{p},\ \mathrm{q}){:}{=}\mathrm{x}\gamma,\ \alpha_1, \ldots, \alpha_k]_{t_1} \qquad [(\mathrm{v},\ \mathrm{w}){:}\mathrm{x}\delta,\ \iota_1, \ldots, \iota_l]_{t_2}}{[(\mathrm{p},\ \mathrm{q}){:}\gamma,\ \alpha_1, \ldots, \alpha_k,\ (\mathrm{v},\ \mathrm{w}){:}\delta,\ \iota_1, \ldots, \iota_l]_c}\ \text{Merge-3}$$

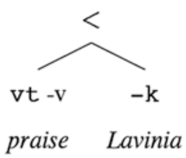
Move-1:

$$\frac{[(\mathrm{p},\ \mathrm{q}){:}{+}\mathrm{y}\gamma,\ \alpha_1, \ldots, \alpha_{i-1},\ (\mathrm{v},\ \mathrm{p}){:}{-}\mathrm{y},\ \alpha_{i+1}, \ldots, \alpha_k]_c}{[(\mathrm{v},\ \mathrm{q}){:}\gamma,\ \alpha_1, \ldots, \alpha_{i-1},\ \alpha_{i+1}, \ldots, \alpha_k]_c}\ \text{Move-1}$$

Move-2: $(\delta \neq \emptyset)$

$$\frac{[(\mathrm{p},\ \mathrm{q}){:}{+}\mathrm{y}\gamma,\ \alpha_1, \ldots, \alpha_{i-1},\ (\mathrm{v},\ \mathrm{w}){:}{-}\mathrm{y}\delta,\ \alpha_{i+1}, \ldots, \alpha_k]_c}{[(\mathrm{p},\ \mathrm{q}){:}\gamma,\ \alpha_1, \ldots, \alpha_{i-1},\ (\mathrm{v},\ \mathrm{w}){:}\delta,\ \alpha_{i+1}, \ldots, \alpha_k]_c}\ \text{Move-2}$$

Fig. 9.12 The five inference rules of the deductive parsing scheme of Harkema (2001). The cumulative effects of these rules simulate the *Move* and *Merge* operations in Minimalist syntax. Each bracketed item is a feature bundle. The material above a line represents the input to the deductive step; the material below the line, the output.



$$[(1,2){:}\mathtt{vt}\ {-}\mathtt{v},(3,4){:}{-}\mathtt{k}]_c$$

Fig. 9.13 The result of a merger. The head of the phrase is *praise*, as indicated by the pointer '<', and *Lavinia* still needs Case {−k}. The structure can be represented as a tree (*top*) or in bracket notation (*bottom*) (*Source*: Harkema 2001)

tions Merge-1 through Move-2 in reverse, will try to break it down into lexical items at the leaves matching the words of the sentence to be parsed. The Earley-style parser essentially is a top-down parser with some additional book-keeping machinery which makes it applicable to a wider range of grammars than a basic top-down algorithm.

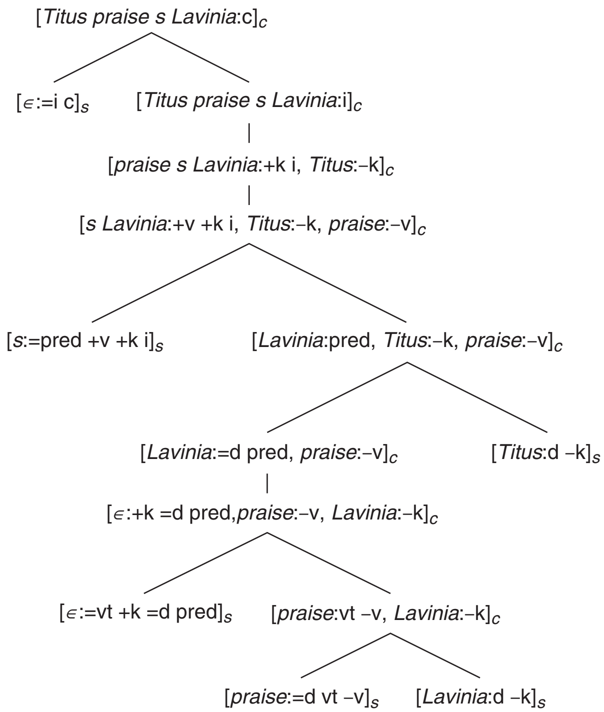


Fig. 9.14 A derivation tree. Note that this is *not* a phrase marker. In this tree, every node is a structured item, which is itself representable by a tree, but is collapsed here using bracket notation, as in Fig. 9.13. The derivation tree represents the hierarchical order of the Merge and Move operations. A bottom-up algorithm will traverse this tree from the leaves up, whereas and top-down algorithm will start at the top and attempt to break the input down into individual lexical items (*Source*: Harkema 2001)

The process that Harkema describes can be represented as a deductive proof procedure:

Step	Lexical Item	Features	Interpretation
1.	Lavinia:	{d, −k}	*Lavinia* is a determiner (on this analysis) and needs case
2.	Titus:	{d, −k}	*Titus* is a determiner (on this analysis) and needs case
3.	Praise:	{=d, vt, −v}	The transitive verb *praise* needs a determiner and a verbal agreement feature.
4.	−s:	{=pred, +v, +k, i}	The inflectional suffix *-s* needs a predicate to attach to, has a verbal agreement feature, and assigns case.
5.	ε:	{=i, c}	An empty complementizer that needs an inflectional phrase
6.	ε:	{=vt, +k, =d, pred}	An empty category that needs a determiner to assign case to

Step	Lexical input to the operation	Operation	Steps that serve as inputs
7.	[praise: vt −v, Lavinia:−k]	Merge-3	(3, 1)
8.	[ε:+k =d pred, praise:-v, Lavinia:−k]	Merge-3	(6, 7)
9.	[Lavinia:=d pred, praise:−v]	Move-1	(8)
10.	[Lavinia:pred, Titus:−k, praise:−v]	Merge-3	(9, 2)
11.	[s Lavinia:+v + k i Titus:−k, praise:−v]	Merge-1	(4, 10)
12.	[praise s Lavinia:+k i, Titus:−k]	Move-1	(1)
13.	[Titus praise s Lavinia:i]	Move-1	(12)
14.	[Titus praise s Lavinia:c]	Merge-1	(5, 12)

This derivation amounts to a bottom-up parse of *Titus praises Lavinia*, which can also be represented as an upward ascent along the derivation tree in Fig. 9.14.

Note that Harkema's proposal makes an explicit provision for various empty categories ('ε') that serve to facilitate the recognition of movement operations. For instance, the empty category that we see in step 8 of the derivation, where *Lavinia* moves to pick up Case, is what facilitates Case assignment. The empty category thus gives rise to the movement; other kinds of movement are triggered by different features. Wh-movement is triggered by the need to check the {−wh} and {+wh} features that some lexical items have in their "bundles." Notice, finally, that this gives rise to a need for the algorithm to recognize the presence of empty categories. For instance, the very first step that a top-down algorithm must make in traversing the tree in Fig. 9.14 is to posit an empty complementizer that needs an inflectional phrase.

Harkema proves that the top-down, bottom-up, and Earley algorithms can parse Minimalist languages in polynomial time—a significant result, given the mild context-sensitivity of Minimalist grammars. Moreover, he points out (pp. 211–214) that the algorithms are compatible with a wide variety of approaches to ambiguity resolution, both in ranked parallel models (e.g., Gibson 1991) and in serial models (e.g., Frazier and Fodor 1978). Depending on the details of the architecture, a Minimalist parser like the ones described here can be made to conform to Minimal Attachment and Late Closure, and it can be supplemented with statistical frequency information for the purposes of probabilistic parsing.

As before, the PAD framework allows the rules of the grammar to be explicitly represented and used as data. But Harkema's presentation does not assume this. Given that the inference rules in Fig. 9.12 can be seen as procedural imperatives, the algorithm can also be implemented by a hardwired circuit, whose operations carry out the inference rules directly, without needing to consult them in a separate data structure. The internal states of such a circuit would explicitly represent only its lexical input (steps 1–6 in the derivation above) and the complex tree structures in steps 7–14. Harkema's algorithm, therefore, leaves open the question of which position psychological reality issue we should adopt. In the next section, we examine a line of reasoning that appeals to the virtues of Minimalist parsing model in supporting the claim common to REP-GRAM-PROC. and EMB-GRAM-PROC.—viz. that the principles of a grammar are of the right sort to govern language processing.

9.5.3 Minimalist Grammar as an Ambiguity Resolution Strategy

Steedman (1985) characterized the Competence Hypothesis as follows:

> For a theory of grammar to be psychologically explanatory, it must, under the Competence Hypothesis, allow for a very direct relation of grammatical rules to the operations of a processor. Ideally, it should require nothing more to turn it into a complete theory of processing than the addition of a mechanism for local ambiguity resolution, to tell it which rule of the grammar to apply at a given point in analysis or generation. (p. 386)

Weinberg (1999) puts forward a proposal that seeks to implement an even stronger version of this idea. Steedman takes a grammar to be psychologically explanatory even if ambiguity resolution preferences have to be added to it in order to turn it into a parsing model. Following the lead of theorists such as Pritchett (1992) and Philips (1996), Weinberg proposes to derive ambiguity resolution preferences *directly from the grammar*, thus maintaining the tightest possible connection between the grammar and the parser.[38] If her proposal works, then it obviates the need for resource-based parsing principles, like MA, LC, and MCP, and reduces the need to rely on frequency information in a probabilistic lexicon or grammar.

Weinberg offers the following specification of a parsing algorithm:

> A derivation proceeds left to right. At each point in the derivation, Merge using the smallest number of operations needed to check a feature on the category about to be attached. If Merger is not possible, try to Move within the current [unit]. If neither merger nor movement is licensed, Spell-Out the [unit]. Repeat until all terminals are incorporated into the derivation.

This subsumes Berwick's proposal, and can be implemented in the ways that Harkema suggests. The interest of Weinberg's proposal is not, however, the algorithm, but, rather, the way in which she derives well-known ambiguity resolution preferences from the independently-motivated syntactic principles of Minimalist grammar.

> Insertion or movement is governed by the Economy Conditions… The preference to attach a category using minimal structure follows immediately from this notion of Economy. At each point a category is inserted using the least number of operations necessary for feature transfer or merger. This ban on unnecessary operations subsumes Frazier and Rayner's (1982), Minimal Attachment and Gorrell's (1995) simplicity condition and has the advantage of following from independently motivated grammatical principles. (p. 289)

Moreover, the application of Economy conditions to the operation of Spell-Out opens up the possibility of using *the grammar itself* to predict well-attested garden-path phenomena, including the difficulty of reanalysis.

---

[38] There is a bit of irony here. Early on, Weinberg was skeptical about the Competence Hypothesis (Berwick and Weinberg 1983, 1984). Perhaps the change of mind came with the shift from old-school transformational grammars to the new-fangled Minimalist formalism, which lends itself more easily to an interpretation as a parsing model.

> Following Uriagereka, we assume that Spell-Out occurs whenever a derivation [cannot proceed by the operations of *Merge* and *Move*]. The Spell-Out conditions thus also provide us with an independently motivated theory of reanalysis. If a preferred reading induces a precedence/dominance mismatch, the category that precedes but does not dominate will be spelled out. Again following Uriagereka, this means that the material inside the spelled out category is linearized and all internal syntactic structure is removed, creating a nondecomposable syntactic word. Given this, reanalysis from the preferred to dispreferred reading that requires either extraction of material from, or insertion of material into this syntactic word, will be impossible. As a lexical item, the spelled out material is an atomic unit, which can no longer be decomposed into its component pieces. If however, reanalysis occurs within a domain where Spell-Out has not applied, then material can be accessed and the preferred reading can be transformed into the dispreferred structure. Incorporating Spell-out and Economy conditions into the grammar also explains the preference for right branching derivations without the need for extra explicit principles which favor this type of derivation. … As a grammatical operation, Spell-out is governed by Economy. Since it does not allow the checking of any features it is an operation of the last resort. As such, it will only be invoked when no other feature checking operation can apply and the minimal number of spell-outs to guarantee satisfaction of the [Linear Correspondence Axiom] will operate at each time step in the derivation. A right branching structure insures that an element that proceeds will also dominate a category and thus minimize the need for Spell-Out. Therefore, right branching structures will be preferred because they economize on the need for Spell-Out. (pp. 289–90)

The core of the idea is not new. We saw, for instance, that the Sausage Machine model of Frazier and Fodor (1978) distinguishes two processing levels, one of which (the SSS) operates on structured material that is, from its perspective, nondecomposable. Fodor and Frazier's explanation of the difficulty of reanalysis consisted in pointing out that it takes additional computational resources to "unpack" and "look inside" these units. Weinberg's proposal is essentially the same, treating the output of the Spell-Out operation as precisely that kind of nondecomposable unit. What is new here is that Spell-Out is an operation of the *grammar*, motivated by formal considerations, not an independent aspect of the parser's computational architecture (e.g., storage space in the memory buffers).

On the basis of the slim resources provided by the Economy conditions on Merge, Move, and Spell-Out, Weinberg's model predicts the human attachment preferences for all the following kinds of ambiguity. (All examples are due to Weinberg.)

Argument/Adjunct attachment ambiguities:

Direct object/complement subject ambiguity.

(56a) The man believed his sister would win the Nobel Prize.
(56b) The man believed his sister.

Preposed object/matrix subject ambiguity.

(57a) After Mary mended the socks fell off the table.
(58b) After Mary mended the socks they fell off the table.

Ditransitive/complex transitive object ambiguity.

(59a) John gave the man the dog for Christmas.
(59b) John gave the man the dog bit a bandage.

Subcategorized PP/NP modifier ambiguities.

(60a) I put the book on the table.
(60b) I put the book on the table into my bag.

Main clause/relative clause ambiguity.[39]

(61a) The horse raced past the barn fell.
(61b) The horse ridden past the barn fell.
(61c) The horse that was raced past the barn fell.

Adjunct/Adjunct attachment ambiguities:

Adverb or particle placement:

(62a) I told Mary that I will come yesterday.
(62b) I called to pick the box up.
(62c) I yelled to take the cat out.

Normally relative clause attachments are dispreferred, but here they are the favored reading.

(63a) Although Erica hated the house she had owned it for years.
(63b) Although Erica hated the house she owned her family lived in it for years.

Suppose Weinberg is right in thinking that Minimalist Economy Conditions can absorb all of the explanatory functions of the least-effort parsing principles MA and LC. Still, as we have seen, there is another competing account of the HSPM's "oracle"—i.e., its ambiguity resolution principles—which appeals solely to frequency information. On this account, "[p]rocessing involves factors such as the frequencies of occurrence and co-occurrence of different types of information, and the weighing of probabilistic and grammatical constraints" (MacDonald et al. 1994: p. 700).

[39] Pritchett (1992) and Stevenson and Merlo (1997) have pointed out that that these types of ambiguities do not cause processing difficulties when the unergative verb 'raced' is replaced by transitive and unaccusative verbs, as in (64). Weinberg claims that her Minimalism-based account predicts these subtle data.

(64a) The student found in the classroom was asleep.
(64b) The butter melted in the pan was burnt.

To illustrate the importance of lexical frequency, consider the fact that verbs like 'decide' occur more frequently with sentential complements, and hence are correctly predicted to violate Minimal Attachment, as in (65).

(65) John decided the contest was fair.

Of course, one wants to know more about how this lexical frequency interacts with other types of frequency information. For instance, the fact that simple sentences occur more frequently in a typical corpus than sentences with embeddings conflicts with the prediction that (65) will be easy to process. Lexical and structural frequencies must be balanced in some principled way. Still, it is plain that the HSPM makes use of frequency information. The question is whether that's *all* it uses.

To address this issue, Weinberg (1999) examines two influential frequency-based proposals (MacDonald et al. 1994; Filip, *et al.* 2002), and shows that exclusive reliance on frequencies leads their models to make false predictions in cases where the grammar-based models make correct ones. On the basis of data concerning the statistical properties of unergative, unaccusative, "ordinary," transitive, intransitive, passive, and active verbs (Stevenson and Merlo 1997), she points out that.

> [e]ven though the class of [ordinary verbs] appears in the transitive construction more frequently than either the ergative or unaccusative class …, it is difficult for speakers to interpret occurrences of these verbs as reduced relatives. If this is correct, it poses a problem for frequency-based approaches, which would predict that this class should be the least difficult to interpret as it most frequently appears in the constructions that relative-clause interpretations presuppose.

In the example she provides, the frequency-based theory incorrectly predicts that (66) will be easier to parse than both (67) and (68). It's actually the hardest, both experimentally and introspectively. Readers encounter trouble at the disambiguation point (underlined).

(66) The dictator *fought* in the violent coup <u>was hated</u> throughout the country.
(67) The dictator *chased* in the violent coup <u>was hated</u> throughout the country.
(68) The dictator *overthrown* in the violent coup <u>was hated</u> throughout the country.

The frequency-based constraint-satisfaction theory in Filip, *et al.* faces additional problems. On their view, "the acceptability of reduced relative clauses, headed by passive participles derived from unaccusative and unergative verbs, increases when the passive participle and the main verb in a matrix clause assign their subject-NPs more proto-Patient and fewer proto-Agent properties" (Weinberg 1999: p. 17). Filip, *et al.* define *proto-Patient* and *proto-Agent* as follows:

Proto-Agent: Volitional, sentient causers of the event described, who move and exist independently of the verb described (*sic*).

Proto-Patient: Undergo a change of state, serve as incremental themes, are causally affected by another event participant, are stationary relative to another event participant, and do not exist independently of the event described.

Filip, *et al.*'s proposal is disconfirmed by the fact that (69) is a strong garden-path for most readers, despite the fact that the verbs *move* and *be afraid* assign all of the following proto-patient properties to the noun phrase *the cattle*: (i) undergo a change of state, (ii) serve as incremental themes, and (iii) are causally affected by another event participant.

(69) The cattle moved into the crowded room were afraid of the cowboys.

Filip, *et al.* incorrectly predict that (69) will cause no trouble for readers. Weinberg takes this result (and others like it) to suggest that "frequency has a role to play but is filtered through grammatically justified constraints" (p. 311). She concludes that

> while speakers may very likely track frequency, *this variable works in tandem with independent grammatical constraints*. If a structure occurs very frequently in a given construction, it can influence the initial preferred analysis, but once an analysis is chosen, based on an amalgam of frequency and grammatical variables, *the grammatically driven reanalysis principles decide what will or will not be a garden path*. (p. 306)

Weinberg's positive proposal constitutes a powerful argument in favor the methodological principle that we encountered earlier, variously known as "type transparency" or the Strong Competence Hypothesis. With regard to the psychological reality issue, it supports both **REP-GRAM-PROC** as **EMB-GRAM-PROC**. But, as with Harkema's algorithmic proposal, Weinberg's view does not settle whether the grammar is explicitly represented or embodied as a set of procedural dispositions. The choice between these options will have to be made on either general principles of parsimony, or by appeal to as-yet-unavailable neurocognitive data.

9.6 Summary and Conclusions

In this chapter, we traced the recent history of two co-evolving fields—formal syntax and computational psycholinguistics. We are now in a position to reap the rewards by providing an empirically informed assessment of the available positions on the psychological reality of syntactic rules and principles. For convenience, I repeat the available positions here:

REP-GRAM-PROC The structure rules of a language—i.e., the syntactic rules or principles that comprise its grammar—are (i) identical with the processing rules that govern the HSPM, and (ii) are *explicitly represented* in the competent speaker's mind/brain.

REP-GRAM-DATA The structure rules of a language (the grammar) are (i) distinct from the processing rules of the HSPM, but (ii) are *explicitly represented* in the competent speaker's mind/brain and used *as data* by the processing rules.

EMB-GRAM-PROC. The structure rules of a language (the grammar) are (i) also the processing rules of the HSPM, but (ii) they are *embodied,* not explicitly represented, in the competent speaker's mind/brain.

GRAM-CONFORM Human language processing *conforms* to a grammar, in the sense that the competent speaker's mind/brain reliably takes sentences that are licensed by the grammar as inputs and produces such sentences as outputs.

Let us now summarize our findings from this chapter and the last.

Having introduced context-free grammars, we surveyed the details of top-down, bottom-up, and left-corner parsers, with a focus on the Earley and CYK algorithms and their probabilistic extensions. We saw that these algorithms, though not themselves grammars, are naturally construed as taking the principles of a grammar *as data*, in line with **REP-GRAM-DATA**. Of course, it may well be that this is an artifact, arising from the need to implement the algorithm on a conventional computer. When these algorithms are cast in the PAD framework, the rules can be seen as (subpersonal analogues of) truth-evaluable declarative statements that are explicitly represented, accessed, and inserted as premises in the deduction—in line with **REP-GRAM-DATA**—but they can also be seen as "implicit inference rules" that guide the deductive procedure *without* being explicitly represented. This latter perspective on the Earley and CYK algorithms would support **EMB-GRAM-PROC**. Doubtless, a device that embodies those inference rules can be built. And the human brain may well be such a device. The situation here is typical; models that seem to support **REP-GRAM-DATA** can always be re-interpreted in such a way as to support the weaker position, **EMB-GRAM-PROC**, which is input-output equivalent, but gets by without the storage, access, and explicit use of represented rules.

We then turned our attention to early transformational grammars, which offered the promise of supporting the even stronger position, **REP-GRAM-PROC**. This period in the history of psycholinguistics was marked by the debate over the Derivational Theory of Complexity (DTC). During this time, the fate of **REP-GRAM-PROC** was seen as resting on the viability of the DTC. Unfortunately, when the latter was found to face serious problems, many psycholinguists abandoned **REP-GRAM-PROC** and began to search for weaker positions on the psychological reality of grammar.

One proposal for "embedding" the Standard Theory of transformational grammar in a parser involved "precomputing" the results of various transformations and storing them in the lexicon. I called this *precomputed lexical realization* (PLR). This proposal, due to Berwick and Weinberg (1984), rests on a relaxed view of what relation can hold between the "competence grammar" and the grammar that actually plays a role in processing, consistent with the "competence grammar" being psychologically real. The relaxation of what Berwick and Weinberg called the "transparency" constraint gives rise to the idea that a grammar can psychologically real even if it is neither represented nor embodied in the parser. This is at the heart of the position that I called *covering grammar realization* (CGR), of which PLR is a special case. According to CGR, a grammar, G_1, can be psychologically real if a distinct grammar, G_2, which is a "covering grammar" G_1, governs the parsing process.

Because the "covering" relation between two grammars has nothing to do with the psychological mechanisms that perform syntactic processing, I argued that CGR and PLR are versions of **GRAM-CONFORM**. As such, they cannot be seen as marking out a substantive notion of psychological reality. Still, even if a grammar, G_1, is not psychologically real in virtue of being logically related to a covering grammar, G_2, the fact remains that G_2 is playing some role in the system's processing. So we can ask about the psychological reality of G_2, at which point all of the available positions come back into play.

The fallout from the debate over the DTC demonstrated that a variety of issues about the architecture of neurocomputational mechanisms must be settled before we can confidently use chronometric data (behavioral or neural) as a guide to the nature of the parser and the role of the grammar within it. Though these developments temporarily put a damper on the prospects for settling the psychological reality issue, new comprehension models—the Augmented Transition Networks (ATNs) of the 1970 and 80's—were soon developed, breathing new life into the debate. Unlike transformational parsers, which relied on the demonstrably unworkable strategy of "running transformations in reverse", ATN parsers provided a realistic example of how the structure rules that comprise a grammar can double as the processing rules governing a parser. In this sense ATNs can be seen as supporting **EMB-GRAM-PROC**.

The main competitor to the ATN approach was the Sausage Machine model, developed by Janet Fodor and Lyn Frazier. The showdown between these two models advanced the psychological reality debate by bringing to bear evidence from garden-path processing. As we have seen, the least-effort parsing principles, Minimal Attachment, Late Closure, and the Minimal Chain Principle, are key ingredients in one of the available solutions to the problem of ambiguity resolution. These principles account for our susceptibility to certain kinds of processing errors—the garden-path effects that we surveyed in Chap. 5. Such data provide a window into the inner workings of the HSPM. Fodor and Frazier argued that the ATN architecture cannot in principle implement Minimal Attachment, and hence cannot explain a variety of data regarding human parsing preferences. This casts doubt on psychological plausibility of ATNs. Fodor and Frazier showed, moreover, that the ability of the Sausage Machine model to implement the least-effort parsing principles is due precisely to its commitment to the existence of a separate rule-library—a data structure of explicitly represented syntactic principles. This is, of course, the core claim of **REP-GRAM-DATA**.

If the argument for the Sausage Model were decisive, then **REP-GRAM-DATA** would emerge as the clear winner; the psychological reality debate would thereby be settled. However, the argument is far from decisive. For one thing, the least effort parsing principles face competition from other theories of "the oracle" that must be embedded in each parser—in particular, from the increasingly popular frequency-based accounts. More importantly, the ATN architecture is not the only way of implementing EMB-GRAM-PROC. An argument against the psychological plausibility of ATNs does not, therefore, undermine the more general claim of EMB-GRAM-PROC.

Continuing to trace the co-evolution of grammars and parsers, we looked next at the Government and Binding (GB) theory and the principles-based parsers of the 1990's. Such parsers treat GB principles as explicitly represented data structures, in accordance with REP-GRAM-DATA. The idea is perhaps most vividly illustrated by models that treat parsing as a species of natural deduction. On this view, a grammar can be treated as a set of declaratively represented axioms, which must be accessed and inserted as a "step" in a formal deduction. This view accords most closely with REP-GRAM-DATA. However, as noted above, the analogy to axioms raises the following possibility: Just as explicitly represented axioms can be simulated by *inference rules* that are "implicit" in a proof procedure, the principles of a grammar may well be *embodied* in the parser, serving as the procedural dispositions in accordance with which the deduction proceeds. This points, once again, to the viability of EMB-GRAM-PROC.

The most recent development in generative grammar is the advent of Minimalism, which incorporates the central tenets of the Principles and Parameters approach but eschews reference to some of the theoretical devices from GB theory. Minimalist grammars are designed to be compact, making use of only the bare minimum of theoretical machinery. To this end, they posit the simple operation *Merge*, which can be made to serve as the primary operation of a parsing model. This idea has been formalized in a variety of computationally efficient algorithms due to Harkema (2001), and has been implemented in both classical computational models (Fong 2011) and connectionist networks (Gerth and beim Graben 2009). Most importantly, it serves as the basis of a theory of sentence processing that derives well-attested processing principles directly from the independently necessary principles of the grammar. Weinberg (1999) argues that independently-motivated conditions on the timing of Spell-Out operations can be used to predict the human parsing preferences. Her proposal appeals to the *grammar* to explain a wide range of behavioral and neurocognitive data, including data on which frequency-based constraint-satisfaction models seem to falter. If this is correct, then the role of frequency information in guiding the construction of mental phrase markers is reduced, and the need for grammar-independent, resource-based parsing principles like MA, LC, and MCP is obviated. The latter consequence would undermine the argument for REP-GRAM-DATA that underlies Fodor and Frazier's criticism of ATN parsers.

Weinberg's argument provides strong support for REP-GRAM-PROC. If sound, it shores up a compelling and psychologically plausible unification of the three elements of language processing—the oracle, the algorithm, and the grammar—and does so by reference to the inner workings of the most cutting-edge syntactic formalism in contemporary generative grammar. Clearly, the prospects for REP-GRAM-PROC are not nearly as dim as was supposed in the heyday of the DTC debate. Still, we do not have here a knockdown argument for REP-GRAM-PROC. As always, we must bear in mind the point established by Stabler (1983) and reiterated by Devitt

(2006a).[40] It is provable that any computation performed by a device that explicitly represents the processing rules of an algorithm can also be performed by a "special-purpose" device that embodies the very same algorithm in the form of hardwired procedural dispositions. The brain may well be such a device. Thus, whatever evidence we find in favor of **REP-GRAM-PROC** is *ipso facto* evidence for **EMB-GRAM-PROC**. Thus, models that are committed to **REP-GRAM-PROC** always face more parsimonious rivals, in which the grammar is not represented but embodied. In view of this, a commitment to **REP-GRAM-PROC** can at this stage be grounded only in the conviction—common in the philosophy of science—that one should adopt the strongest, most falsifiable hypothesis consistent with the available data. But this methodological precept faces two problems.

First, it conflates the strictures on *advancing* a hypothesis with strictures on *accepting* one. While it is rational to advance strong, highly falsifiable hypotheses for the purpose of guiding empirical research, accepting or believing such hypotheses is a different matter. The rationality of research methodology comes apart from the rationality of belief fixation. The second (related) problem is that the methodological precept in question runs up against the desirability of *parsimonious* hypotheses—ones that posit the simplest mechanisms sufficient to explain the data. A commitment to **REP-GRAM-PROC** ranks low on this latter measure. Explicit representation of syntactic rules or principles requires additional machinery for storing, accessing, and "inserting" them at the right point in the parsing procedure. Embodied rules and principles may require larger and more complex circuits, but there is at present no reason to believe that the impressive physiological resources of the human brain aren't up to the task.

The past several decades of psycholinguistic research have shed a great deal of light on the nature of the human sentence processing routines, but I am aware of no experimental results or theoretical considerations that would militate decisively in favor of the idea that the rules or principles of grammar are *represented* in the mind. Thus, a common claim in generative linguistics, i.e., that grammars are represented in the brains of competent language users, must be seen either as loose talk—a conflation of the notions of embodiment and representation—or as a strong but as-yet-ungrounded hypothesis. Still, we have seen that a wholesale skepticism concerning the psychological reality of syntactic principles is unwarranted. The view that I've labeled **GRAM-CONFORM** is weaker than what the available evidence warrants. We may, at present, confidently conclude that the rules or principles of syntax are psychologically real *at least* in the sense of being embodied in the brain. The stronger hypothesis of explicit representation is, however, still in the running. It demands scrutiny on the part of psycholinguists and neuroscientists, and may well be vindicated in the fullness of time.

---

[40] "[S]hould a rule govern a cognitive process, it could be the case that it governs by being embodied without being represented. So, where we have evidence that a certain rule does govern, Pylyshyn's Razor demands further evidence before we conclude that it does so by being represented; we need further evidence that the rule plays its role like a soft-wired rule in a general-purpose computer rather than like a hardwired rule in a special-purpose computer. I suggest that there is a striking lack of this further evidence with human cognitive processing" (p. 204).

References

Abney, S. P. (1991). Parsing by chunks. In R. C. Berwick, S. P. Abney, & C. Tenny (Eds.), *Principle-based parsing* (pp. 257–278). Dordrecht: Kluwer Academic Publishers.

Abney, S. P. (1996). Statistical methods in linguistics. In J. Klavans & P. Resnik (Eds.), *The balancing act*. Cambridge: MIT Press.

Abney, S. P., & Johnson, M. (1991). Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research, 20*(3), 233–250.

Aho, A. V., & Ullman, J. D. (1972). *The theory of parsing, translation, and compiling* (Vol. 1). Englewood Cliffs: Prentice-Hall.

Anderson, M. L. (2003). Embodied cognition: A field guide. *Artificial Intelligence, 149*, 91–130.

Antony, L., & Hornstein, N. (Eds.). (2003). *Chomsky and his critics*. Malden: Blackwell.

Armstrong, D. M. (1968). *A materialist theory of the mind*. London: Routledge.

Auer, P., & Wei, L. (Eds.). (2007). *Handbook of multilingualism and multilingual communication*. Berlin: Mouton de Gruyter.

Aydede, M. (2010). The language of thought hypothesis. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. URL = http://plato.stanford.edu/entries/language-thought/

Bader, M., Meng, M., & Bayer, J. (2000). Case and reanalysis. *Journal of Psycholinguistic Research, 29*(1), 37–52.

Balog, K. (1999). Conceivability, possibility, and the mind-body problem. *Philosophical Review, 108*, 497–528.

Barber, A. (Ed.). (2003). *Epistemology of language*. New York: Oxford University Press.

Barber, A. (2004). Idiolects. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. URL = http://plato.stanford.edu/entries/idiolects/

Bates, M. (1978). The theory and practice of augmented transition network grammars. In L. Bolc (Ed.), *Natural language communication with computers*. Berlin: Springer.

Batterink, L., & Neville, H. J. (2013). The human brain processes syntax in the absence of conscious awareness. *Journal of Neuroscience, 33*(19), 8528–8533.

Bechtel, W. (1987). Psycholinguistics as a case of cross-disciplinary research: A symposium introduction. *Synthese, 72*(3), 293–311.

Bechtel, W., & Abrahamsen, A. (2002). *Connectionism and the mind: Parallel processing, dynamics, and evolution in networks* (2nd ed.). Oxford: Blackwell.

Bechtel, W., & Abrahamsen, A. (2006). Phenomena and mechanisms: Putting the symbolic, connectionist, and dynamical systems debate in broader perspective. In R. J. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. 159–185).

Bermúdez, J. L. (2014). *Cognitive science: An introduction to the science of the mind* (2nd ed.). Cambridge: Cambridge University Press.

D. Pereplyotchik, *Psychosyntax*, Philosophical Studies Series 129,
DOI 10.1007/978-3-319-60066-6

Bermúdez, J., & Cahen, A. (2010). Nonconceptual mental content. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*.

Berthouzoz, C., & Merlo, P. (1997). Statistical ambiguity resolution for principle-based parsing. In *Proceedings of the recent advances in natural language processing* (pp. 179–186). Available at the following URL: http://www.latl.unige.ch/doc/ranlp97.ps

Berwick, R. C. (1991a). Principles of principle-based parsing. In R. C. Berwick, S. P. Abney, & C. Tenny (Eds.), *Principle-based parsing: Computation and psycholinguistics* (pp. 1–37). Dordrecht: Kluwer Academic Publishers.

Berwick, R. C. (1991b). Principle-based parsing. In S. M. Shieber & T. Wasow (Eds.), *P. Sells* (pp. 115–226). Foundational issues in natural language processing: Kluwer Academic Publishers.

Berwick, R. C. (1997). *Syntax Facit Saltum*: Computation and the genotype and phenotype of language. *Journal of Neurolinguistics, 10*(213), 231–249.

Berwick, R. C., & Fong, S. (1995). A quarter century of computation with transformational grammar. In J. Cole, G. M. Green, & J. L. Morgan (Eds.), *Linguistics and computation*. Boston: Kluwer Academic Publishers.

Berwick, R. C., & Niyogi, S. (2005). A minimalist implementation of Hale-Keyser incorporation theory. In A. M. DiSciullo (Ed.), *UG and external systems, language, brain, and computation* (p. 269). Amsterdam: John Benjamins Publishing.

Berwick, R. C., & Weinberg, A. S. (1982). Parsing efficiency, computational complexity, and the evaluation of grammatical theories. *Linguistic Inquiry, 13*, 165–192.

Berwick, R. C., & Weinberg, A. S. (1983). The role of grammars in models of language use. *Cognition, 13*, 1–62.

Berwick, R. C., & Weinberg, A. S. (1984). *The grammatical basis of linguistic performance*. Cambridge, MA: MIT Press.

Berwick, R. C., Abney, S. P., & Tenny, C. (Eds.). (1991). *Principle-based parsing: Computation and psycholinguistics*. Dordrecht: Kluwer Academic Publishers.

Bever, T. G., Carroll, J. M., & Miller, L. A. (Eds.). (1984). *Talking minds: The study of language in cognitive science*. Cambridge, MA: MIT Press.

Black, M., & Chiat, S. (1981). Psycholinguistics without 'psychological reality'. *Linguistics, 19*, 37–61.

Black, E., Jelinek, F. Lafferty, J., Magerman, D. M., Mercer, R., & Roukos, S. (1993). Towards history-based grammars: Using richer models for probabilistic parsing. In *ACL 31* (pp. 31–37).

Block, N. (1978). Troubles with functionalism. *Minnesota Studies in the Philosophy of Science, 9*, 261–325.

Block, N. (1986). Advertisement for a semantics for psychology. In P. A. French et al. (Eds.), *Midwest studies in philosophy* (Vol. X, pp. 615–678). Minneapolis: University of Minnesota Press.

Block, N. (1996). On a confusion about a function of consciousness. In *Behavioral and brain sciences* (Vol. 18, pp. 227–247). Cambridge: Cambridge University Press.

Blumenthal, A. L. (1970). *Language and psychology: Historical aspects of psycholinguistics*. New York: Wiley.

Bock, K. (1989). Closed-class immanence in sentence production. *Cognition, 31*, 163–186.

Bock, J. K., & Kroch, A. S. (1989). The isolability of syntactic processing. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 157–196). Dordrecht: Kluwer.

Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition, 35*, 1–39.

Bock, K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review, 99*, 150–171.

Bod, R. (1998). *Beyond grammar: An experience-based theory of language*. Stanford: CSLI Publications.

Bod, R., Hay, J., & Jannedy, S. (Eds.). (2003). *Probabilistic linguistics*. Cambridge, MA: MIT Press.

Boden, M. A. (1984). Animal perception from an artificial intelligence viewpoint. In C. Hookway (Ed.), *Minds, machines and evolution*. Cambridge: Cambridge University Press.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). *Processing syntax and morphology. A neurocognitive perspective*. Oxford: Oxford University Press.

Braddon-Mitchell, D., & Jackson, F. (1996). *Philosophy of mind and cognition*. Blackwell Publishers.

Brandom, R. (1994). *Making it explicit*. Cambridge, MA: Harvard University Press.

Brandom, R. (2001). *Articulating reasons*. Cambridge, MA: Harvard University Press.

Brandom, R. (2008). *Between saying and doing*. New York: Oxford University Press.

Brandom, R. (2009). *Reason in philosophy*. Cambridge, MA: Harvard University Press.

Bresnan, J. (1978). A realistic transformational grammar. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.

Bresnan, J. (Ed.). (1982). *The mental representation of grammatical relations*. Cambridge, MA: MIT Press.

Bresnan, J. (2001). *Lexical-functional syntax*. Malden: Blackwell Publishers.

Bresnan, J., & Kaplan, R. (1982). Introduction: Grammars as mental representations of language. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. xvii–xlii). Cambridge, MA: MIT Press.

Burge, T. (1979). Individualism and the mental. In P. A. French, T. E. Uehling Jr., H. K. Wettsteinn (Eds), *Midwest studies in philosophy* (Vol. IV) Minneapolis: University of Minnesota Press. Reprinted in *The Nature of Mind*, ed. David Rosenthal. New York: Oxford University Press, 1991, pp. 536–567.

Byrne, A., & Logue, H. (Eds.). (2009). *Disjunctivism: Contemporary readings*. Cambridge, MA: MIT Press.

Cain, M. J. (2010). Linguistics, psychology and the scientific study of language. *Dialectica, 64*(3), 385–404.

Carnie, A. (2010). *Constituent structure* (2nd ed.). Oxford, Oxford University Press.

Carreiras, M., & Clifton, C. (1993). Relative clause interpretation preferences in Spanish and English. *Language and Speech, 36*(4), 353–372.

Carruthers, P. (2000). *Phenomenal consciousness: A naturalistic theory*. Cambridge: Cambridge University Press.

Carruthers, P. (2006). The case for massively modular models of the mind. In R. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. 3–21). Malden: Blackwell.

Carter, M. (2007). *Minds and computers: An introduction to the philosophy of artificial intelligence*. Edinburgh: Edinburgh University Press. chs. 14, 16.

Chalmers, D. (1990). Syntactic transformations on distributed representations. *Connection Science, 2*, 53–62.

Chalmers, D. (1993). Connectionism and compositionality: Why Fodor and Pylyshyn were wrong. *Philosophical Psychology, 6*, 305–319.

Chalmers, D. (1993/2011). A computational foundation for the study of cognition. Special issue of the *Journal of Cognitive Science*, G. Piccinini (Ed.), 12(4).

Chalmers, D. (1995). On implementing a computation. *Minds and Machines, 4*, 391–402.

Chalmers, D. (1996). *The conscious mind*. New York: Oxford University Press.

Charniak, E. (1993). *Statistical language learning*. Cambridge: MIT Press.

Charniak, E. (1996). Tree-bank grammars. In *Proceedings of the thirteenth national conference on artificial intelligence* (pp. 1031–1036).

Charniak, E. (1997). Statistical parsing with a context-free grammar and word statistics. In *Proceedings of the fourteenth national conference on artificial intelligence* (pp. 598–603).

Charniak, E., Goldwater, S., & Johnson, M. (1998). Edge-based best-first chart parsing. Available at the following URL: http://www.aclweb.org/anthology-new/W/W98/W98-1115.pdf

Chesi, C. (2004). *Phases and cartography in linguistic computation: Toward a cognitively motivated computational model of linguistic competence*. PhD thesis, Università di Siena.

Chesi, C. (2007a). An introduction to phase-based minimalist grammars: Why *move* is top-down from left-to-right. *Studies in Linguistics (CISCL WP), 1*, 49–90.
Chesi, C. (2007b). Five reasons for building phrase structures top-down from left to right. *Nanzan Linguistics*, Special Issue 3.
Chomsky, N. (1957/2002). *Syntactic structures* (2nd ed.). Berlin: De Gruyter Mouton.
Chomsky, N. (1959). Review of *verbal behavior*. By B. F. Skinner. *Language, 35*, 26–58.
Chomsky, N. (1961/1964). Degrees of grammaticalness. In Fodor and Katz (Eds.), *The structure of language: Readings in the philosophy of language*. Excerpted from "Some Methodological Remarks on Generative Grammar," in Word, 17, 1961 (pp. 219–239).
Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
Chomsky, N. (1967). Reply to Harman. In S. Hook (Ed.), *Language and philosophy: A symposium*. New York University Press: New York.
Chomsky, N. (1969a). Linguistics and philosophy. *Hook, 1969*, 51–94.
Chomsky, N. (1969b). Comments on Harman's reply. *Hook, 1969*, 152–159.
Chomsky, N. (1972). *Language and mind* (Enlarged ed.). New York: Harcourt Brace Jovanovich.
Chomsky, N. (1975a). Knowledge of language. In K. Gunderson (Ed.), *Minnesota studies in the philosophy of science, volume VII: Language, mind, and knowledge* (pp. 299–320).
Chomsky, N. (1975b). *Reflections on language*. New York: Pantheon Books.
Chomsky, N. (1975c). Quine's empirical assumptions. In D. Davidson & J. Hintikka (Eds.), *Words and objections: Essays on the work of W. V. Quine* (pp. 53–68). Dordrecht: Reidel.
Chomsky, N. (1977). *Essays on form and interpretation*. Amsterdam: Elsevier North Holland.
Chomsky, N. (1978). On the biological basis of language capacities. In G. A. Miller & E. Lenneberg (Eds.), *Psychology and biology of language and thought*. New York: Academic.
Chomsky, N. (1980a). *Rules and representations*. New York: Columbia University Press.
Chomsky, N. (1980b). Rules and representations. *Behavioral and Brain Sciences, 3*, 1–14.
Chomsky, N. (1980c). "Author's response" to peer commentary on 1980b. *Behavioral and Brain Sciences, 3*, 42–58.
Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris Publications.
Chomsky, N. (1986). *Knowledge of language*. New York: Praeger.
Chomsky, N. (1988). Language and problems of knowledge. In A. P. Martinich (Ed.), *The philosophy of language*, 4th ed., ch. 39.
Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge, MA: Cambridge University Press.
Chomsky, N. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A life in language*, *Current studies in linguistics 36* (pp. 1–52). Cambridge, MA: MIT Press.
Chomsky, N. (2003). Reply to Rey. In Antony and Hornstein (Eds.), pp. 274–287.
Churchland, P. M. (1979). *Scientific realism and the plasticity of mind*. Cambridge, MA: Cambridge University Press.
Churchland, P. M. (1985/1989). The ontological status of observables: In praise of super-empirical virtues. In P. M. Churchland & C. A. Hooker (Eds.), *Images of science*. Chicago: University of Chicago Press; reprinted as ch. 8 in A *Neurocomputational Perspective: The Nature of the Mind and the Structure of Science*, Cambridge, MA: MIT Press.
Churchland, P. (1988). *Matter and consciousness*. Cambridge, MA: MIT Press.
Churchland, P. M. (1995). *The engine of reason, the seat of the soul: A philosophical journey into the brain*. Cambridge, MA: MIT Press.
Clapin, H. (2002). *The philosophy of mental representation*. Oxford: Clarendon Press.
Clapin, H., Staines, P., & Slezak, P. (Eds.). (2004). *Representations in mind: New approaches to mental representation*. Amsterdam: Elsevier Publishing.
Clark, A. (1991). *Microcognition: Philosophy, cognitive science, and parallel distributed processing*. Cambridge, MA: MIT Press.
Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Science, 10*, 370–374.

Clark, A. (2010). Statistical parsing. In A. Clark, C. Fox, & S. Lappin (Eds.), *Handbook of computational linguistics and natural language processing* (pp. 333–363). Wiley-Blackwell: Malden.
Clark, H. H., & Clark, E. V. (1977). *Psychology and language*. New York: Harcourt Brace Jovanovich.
Clark, A., Fox, C., & Lappin, S. (2010). *Handbook of computational linguistics and natural language processing*. Malden: Wiley-Blackwell.
Clifton, C., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language, 49*, 317–334.
Collins, M. (1997). Three generative, lexicalised models for statistical parsing. In *Proceedings of the 35th Meeting of the ACL* (pp. 16–23).
Collins, M. (1999). *Head-driven statistical models for natural language parsing*. PhD thesis, University of Pennsylvania.
Collins, J. (2007). Review of *ignorance of language*. *Mind, 116*(462), 416–423.
Collins, J. (2008a). *Chomsky: A guide for the perplexed*. London: Continuum Press.
Collins, J. (2008b). Knowledge of language redux. *Croatian Journal of Philosophy, 7*, 3–42.
Collins, J. (2009). A question of irresponsibility: Postal, Chomsky, and Gödel. *Biolinguistics, 3*(1), 99–103.
Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences, 3*, 115–120.
Costa, F., Frasconi, P., Lombardo, V., & Soda, G. (2000). Learning incremental syntactic structures with recursive neural networks. In *Proceedings of the fourth international conference on knowledge-based intelligent engineering systems and allied technologies* (Vol. 2, pp. 458–461). IEEE.
Cowie, F. (1997). The logical problem of language acquisition. *Synthese, 111*(1), 17–51.
Cowie, F. (2010). Innateness and language. In E. Zalta (Ed.), *Stanford encyclopedia of philosophy*, http://plato.stanford.edu/archives/sum2010/entries/innateness-language/
Crocker, M. W. (1996). *Computational psycholinguistics*. Dordrecht: Kluwer.
Crocker, M. W. (1999). Mechanisms for sentence processing. In S. Garrod & M. Pickering (Eds.), *Language processing*. Taylor and Francis.
Crocker, M. W. (2010). Computational psycholinguistics. In A. Clark, C. Fox, & S. Lappin (Eds.), *Handbook of computational linguistics and natural language processing* (pp. 482–513). Wiley-Blackwell: Malden.
Crocker, M. W., & Brants, T. (2000a). Probabilistic parsing and psychological plausibility. In *Proceedings of the 18th conference on Computational linguistics – Volume 1*.
Crocker, M. W., & Brants, T. (2000b). Wide-coverage probabilistic sentence processing. *Journal of Psycholinguistic Research, 29*(6), 647–669.
Crocker, M. W., Pickering, M., & Clifton, C. (Eds.). (2000). *Architectures and mechanisms for language processing*. Cambridge: Cambridge University Press.
Cuetos, F., & Mitchell, D. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition, 30*, 73–105.
Culbertson, J., & Gross, S. (2009). Are linguists better subjects? *British Journal for the Philosophy of Science, 60*, 721–736.
Culbertson, J., & Gross, S. (2011). Revisited linguistic intuitions. *British Journal for the Philosophy of Science.*
Cummins, R. (1985). *The nature of psychological explanation*. Cambridge, MA: MIT Press.
Cummins, R. (1991). *Meaning and mental representation*. Cambridge, MA: MIT Press.
Cummins, R. (1996a). *Representations, targets, and attitudes*. Cambridge, MA: MIT Press.
Cummins, R. (1996b). Systematicity. *Journal of Philosophy, 93*(12), 591–614.
Cummins, R. (2010). *The world in the head*. Oxford: Oxford University Press.
Davidson, D., & Hintikka, J. (Eds.). (1975). *Words and objections: Essays on the work of W. V. Quine*. Springer: Dordrecht.
Davies, M. (1987). Tacit knowledge and semantic theory: Can a five per cent difference matter? *Mind, 96*, 441–462.

Davies, M. (1989). Tacit knowledge and subdoxastic states. In A. George (Ed.), *Epistemology of language* (pp. 131–152). Oxford: Oxford University Press.

Davies, M. (1995). Two notions of implicit rules. In J. E. Tomberlin (Ed.), *Philosophical perspectives, 9, AI, connectionism, and philosophical psychology*. Cambridge, MA: Blackwell.

Davis, S., & Gillon, B. (Eds.). (2004). *Semantics: A reader*. New York: Oxford University Press.

Delmonte, R. (2005). Deep & shallow linguistically based parsing: Parameterizing ambiguity in a hybrid parser. In A. M. DiSciullo (Ed.), *UG and external systems, language, brain, and computation*. Amsterdam: John Benjamins Publishing Co.

De Mulder, H., Everaert, M., Nilsen, O., Lentz, T., & Zondervan, A. (Eds.). (2010). *The linguistic enterprise: From knowledge of language to knowledge in linguistics*. Amsterdam: John Benjamins.

Dennett, D. C. (1969). *Content and consciousness*. London: Routledge & K. Paul.

Dennett, D. C. (1978). *Brainstorms: Philosophical essays on mind and psychology*. Cambridge, MA: MIT Press.

Dennett, D. C. (1987). *The intentional stance*. Cambridge, MA: MIT Press.

Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown, and Co.

Dennett, D. C. (1997). *Kinds of minds: Towards an understanding of consciousness*. Basic Books.

Dennett, D. C. (2014). *Intuition pumps and other tools for thinking*. New York: W. W. Norton & Co.

DeVincenzi, M. (1991). Filler-gap dependencies in a null subject language: Referential and non-referential WHs. *Journal of Psycholinguistic Research, 20*(3), 197–213.

Devitt, M. (1996). *Realism and truth* (2nd ed.). Princeton: Princeton University Press.

Devitt, M. (1998). Naturalism and the a priori. *Philosophical Studies, 92*, 45–65. Reprinted in Devitt, 2010b.

Devitt, M. (2002). Moral realism: A naturalistic perspective. *Croatian Journal of Philosophy, 2*, 1–15.

Devitt, M. (2003). Linguistics is not psychology. In A. Barber (Ed.), *The epistemology of language* (pp. 107–139). Oxford: OUP.

Devitt, M. (2005). There is no a priori. In E. Sosa & M. Steup (Eds.), *Contemporary debates in epistemology* (pp. 105–115). Cambridge, MA: Blackwell Publishers.

Devitt, M. (2006a). *Ignorance of language*. Oxford: Oxford University Press.

Devitt, M. (2006b). Defending *ignorance of language*: Responses to the Dubrovnik papers. *Croatian Journal of Philosophy, 6*, 571–606.

Devitt, M. (2008a). Explanation and reality in linguistics. *Croatian Journal of Philosophy, VIII*(23), 203–231.

Devitt, M. (2008b). A response to Collins' note on conventions and unvoiced syntax. *Croatian Journal of Philosophy, 8*, 249–255.

Devitt, M. (2008c). Methodology in the philosophy of linguistics. *Australasian Journal of Philosophy, 86*, 671–684.

Devitt, M. (2009). Psychological conception, psychological reality: A response to Longworth and Slezak. *Croatian Journal of Philosophy, 9*, 37–46.

Devitt, M. (2010a). Linguistic intuitions revisited. *British Journal for the Philosophy of Science, 61*, 833–865.

Devitt, M. (2010b). *Putting metaphysics first: Essays on metaphysics and epistemology*. Oxford: Oxford University Press.

Devitt, M. (2010c). What 'intuitions' are linguistic evidence? *Erkenntnis, 73*, 251–264.

Devitt, M. (2011a). Linguistic knowledge. In J. Bengson & M. Moffett (Eds.), *Knowing how: Essays on knowledge, mind and action* (pp. 314–333). Oxford: Oxford University Press.

Devitt, M. (2011b). *Methodology and the nature of knowing how* (pp. 205–218). CVIII: Journal of Philosophy.

Devitt, M., & Sterelny, K. (1989). What's wrong with 'the right view'. In J. E. Tomberlin (Ed.), *Philosophical perspectives, 3: Philosophy of mind and action theory* (pp. 497–531). Atascadero: Ridgeview Publishing Company.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19*(1), 158–164.

DiSciullo, A. M. (Ed.). (2005). *UG and external systems, language, brain, and computation*. Amsterdam: John Benjamins Publishing.

Dowty, D. R., Kartunnen, L., & Zwicky, A. M. (Eds.). (1985). *Natural language processing: Psychological, computational, and theoretical perspectives*. Cambridge, UK: Cambridge University Press.

Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.

Dretske, F. (1986). *Explaining behavior*. Cambridge, MA: MIT Press.

Dretske, F. (2006). Perception without awareness. In T. S. Gendler & J. Hawthorne (Eds.), *Perceptual experience*. Oxford: Oxford University Press.

Earley, J. (1970). An efficient context-free parsing algorithm. *Communications of the Association for Computing Machinery, 13*(2).

Eddington, D. (2008). 2008 presidential address: Linguistics and the scientific method. *Southwest Journal of Linguistics, 27*, 1–16.

Edelman, S., & Christiansen, M. H. (2003). How seriously should we take minimalist syntax? *Trends in Cognitive Sciences, 7*(2), 60–61.

Egan, F. (1991). Propositional attitudes and the language of thought. *Canadian Journal of Philosophy, 21*(3), 379–388.

Elman, J. L. (1992). Grammatical structures and distributed representations. In S. Davis (Ed.), *Connectionism: Theory and practice* (Vol. 3). Oxford: Oxford University Press.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*, 71–99.

Epstein, S. D., & Hornstein, N. (Eds.). (1999). *Working minimalism*. Cambridge, MA: MIT Press.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*, 429–492.

Featherston, S. (2001). *Empty categories in sentence processing*. Amsterdam: John Benjamins Publishing Company.

Featherstone, C. R., Morrison, C. M., Waterman, M. G., & MacGregor, L. J. (2013). Semantics, syntax or neither? A case for resolution in the interpretation of N500 and P600 responses to harmonic incongruities. *PloS One, 8*(11).

Fernández, E. M., & Cairns, H. S. (Eds.). (2011). *Fundamentals of psycholinguistics*. Malden: Wiley-Blackwell.

Ferreira, F., & Clifton, C., Jr. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348–368.

Ferreira, F., & Henderson, J. M. (1991). Recovery of misanalyses of garden-path sentences. *Journal of Memory and Language, 30*, 725–745.

Ferreira, F., & Henderson, J. M. (1998). Syntactic reanalysis, thematic processing, and sentence comprehension. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis and sentence processing* (pp. 73–100). New York: Kluwer Academic Publishers.

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research, 30*(1), 20.

Field, H. (1978). Mental representation. *Erkenntnis, 13*, 9–61.

Filip, H., Tanenhaus, M. K., Carlson, G. N., Allopenna, P. D., & Blatt, J. (2002). Reduced relatives judged hard require constraint-based analyses. In P. Merlo & S. Stevenson (Eds.), *Sentence processing and the lexicon: Formal, computational, and experimental perspectives* (pp. 255–280). Amsterdam: Benjamins.

Fodor, J. A. (1968a). *Psychological explanation*. New York: Random House.

Fodor, J. A. (1968b). The appeal to tacit knowledge in psychological explanation. *The Journal of Philosophy, 65*(20).

Fodor, J. A. (1970). Three reasons for not deriving 'kill' from 'cause to die'. *Linguistic Inquiry, 1*, 429–438.

Fodor, J. A. (1975). *The language of thought*. Cambridge: Harvard University Press.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry, 9*(3), 427–473.

Fodor, J. D. (1980). Superstrategy. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Studies in psycholinguistics presented to Merrill Garrett*. Hillsdale: Lawrence Erlbaum Associates.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. A. (1987). *Psychosemantics*. Cambridge, MA: MIT Press.

Fodor, J. D. (1989). Empty categories in sentence processing. *Language and Cognitive Processes, 4*, 155–209.

Fodor, J. A. (1990). *A theory of content, and other essays*. Cambridge, MA: MIT Press.

Fodor, J. D. (1995). Comprehending sentence structure. In L. R. Gleitman & M. Liberman (Eds.), *An invitation to cognitive science: Volume 1, language* (pp. 209–246). Cambridge, MA: MIT Press.

Fodor, J. A. (1997). Connectionism and the problem of systematicity (continued): Why Smolensky's solution still does not work. *Cognition, 63*(1), 109–119.

Fodor, J. A. (1998a). *Concepts: Where cognitive science went wrong*. New York: Oxford University Press.

Fodor, J. D. (1998b). Learning to parse? *Journal of Psycholinguistic Research, 27*(2), 285–319.

Fodor, J. D. (1998c). Parsing to learn. *Journal of Psycholinguistic Research, 27*(3), 339–374.

Fodor, J. A. (2002). *The mind doesn't work that way*. Cambridge, MA: MIT Press.

Fodor, J. A. (2003). *Hume variations*. New York: Oxford University Press.

Fodor, J. A. (2008). *LOT 2: The language of thought revisited*. New York: Oxford University Press.

Fodor, J. D., & Ferreira, F. (Eds.). (1998). *Reanalysis in sentence processing*. New York: Kluwer Academic Publishers.

Fodor, J. D., & Frazier, L. (1980). Is the human sentence parsing mechanism an ATN? *Cognition, 8*, 417–459.

Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research, 23*, 407–434.

Fodor, J. D., & Inoue, A. (1998). Attach anyway. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis and sentence processing* (p. xx). New York: Kluwer Academic Publishers.

Fodor, J. D., & Inoue, A. (2000a). Garden path repair: Diagnosis and triage. *Language and Speech, 43*(3), 261–271.

Fodor, J. D., & Inoue, A. (2000b). Syntactic features in reanalysis: Positive and negative features. *Journal of Psycholinguistic Research, 29*(1).

Fodor, J. A., & Katz, J. J. (1964). *The structure of language: Readings in the philosophy of language*. Englewood Cliffs: Prentice-Hall.

Fodor, J. A., & Lepore, E. (1992). *Holism: A shopper's guide*. Cambridge, MA: Blackwell.

Fodor, J., & McLaughlin, B. P. (1990). Connectionism and the problem of systematicity: Why Smolensky's solution does not work. *Cognition, 35*, 183–204.

Fodor, J., & Pylyshyn, P. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28*, 3–71.

Fodor, J. A., Bever, T., & Garrett, M. (1974). *The psychology of language*. New York: McGraw Hill.

Fong, S. (2005). Computation with probes and goals: A parsing perspective. In A. M. DiSciullo (Ed.), *UG and external systems, language, brain, and computation*. Amsterdam: John Benjamins Publishing Co..

Fong, S. (2011). *Minimalist parsing: Simplicity and feature unification*, *Workshop on language and recursion*. Mons: University of Mons. March.

Ford, M., Bresnan, J., & Kaplan, R. (1982). A competence-based theory of syntactic closure. In J. Bresnan (Ed.), *The mental representation of grammatical relations*. Cambridge, MA: MIT Press.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27–85). Hillsdale: Erlbaum.

Forster, K. I., & Olbrei, I. (1973). Semantic heuristics and syntactic analysis. *Cognition, 2*, 319–347.

Frank, S. L. (2008). Self-organizing word representations for fast sentence processing. In French and Thomas (Eds.), *From associations to rules, connectionist models of behavior and cognition*: *Proceedings of the tenth neural computation and psychology workshop* (pp. 78–88). Hackensack: World Scientific Publishing Co.

Frank, R., Mathis, D., & Badecker, W. (2008). The acquisition of anaphora by simple recurrent networks. Unpublished MS, available at the following URL: http://whitney.ling.yale.edu/~rfrank/Bob_Frank/Publications_files/SRN-note-revised.pdf

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. PhD dissertation. available at: http://digitalcommons.uconn.edu/dissertations/AAI7914150/

Frazier, L. (1988). Grammar and language processing. In F. Newmeyer (Ed.), *Linguistics: The Cambridge survey* (Vol. 1, pp. 15–34). Cambridge: Cambridge University Press.

Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes, 4*(2), 93–126.

Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge, MA: MIT Press.

Frazier, L., & Clifton, C. (1998). Sentence reanalysis, and visibility. In F. Ferreira & J. D. Fodor (Eds.), *Reanalysis in sentence processing* (pp. 143–176). New York: Kluwer Academic Publishers.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*, 291–325.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210.

Frenck-Mestre, C., & Pynte, J. (1995). Lexical influences on parsing strategies: Evidence from eye movements. In: J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), *Studies in visual information processing, Vol. 6 Eye movement research: Mechanisms, processes, and applications* (pp. 433–444).

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language, 50*, 259–281.

Friederici, A. D., & Brauer, J. (2009). Syntactic complexity in the brain. In T. Givón & M. Shibatani (Eds.), *Syntactic complexity: Diachrony, acquisition, neuro-cognition, evolution*. Amsterdam: John Benjamins Publishing Company.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological, and syntactic violations. *Cognitive Brain Research, 1*, 183–192.

Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research, 31*(1).

Gallagher, S. (2005). *How the body shapes the mind*. Oxford: Oxford University Press.

Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.

Gardner-Chloros, P. (2009). *Code-switching*. Cambridge, MA: Cambridge University Press.

Garnham, A. (1983). Why psycholinguists don't care about the DTC: A reply to Berwick and Weinberg. *Cognition, 15*, 263–269.

Gazdar, G., Klein, E., Pullum, G., & Sag, I. (1985). *Generalized phrase structure grammar*. Oxford: Basil Blackwell.

George, A. (Ed.). (1989). *Reflections on Chomsky*. Oxford: Basil Blackwell.

Gerth, S., & Beim Graben, P. (2009). Unifying syntactic theory and sentence processing difficulty through a connectionist minimalist parser. *Cognitive Neurodynamics, 3*, 297–316.

Gibbs, R. W., Jr. (2003). Embodied experience and linguistic meaning. *Brain and Language, 84*, 1–15.

Gibson, E. A. F. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown*. Unpublished Ph.D. dissertation, Carnegie Melon University. Available at: tedlab.mit.edu/tedlab_website/researchpapers/Gibson%201991.pdf

Gibson, E. A. F. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1–76.

Gleitman, L., & Liberman, M. (Eds.). (1995). *An invitation to cognitive science: Volume 1, Language*. Cambridge, MA: MIT Press.

Godfrey-Smith, P. (2004). On folk psychology and mental representation, ch. 8. In H. Clapin, P. J. Staines, & P. P. Slezak (Eds.), *Representations in mind: New approaches to mental representation*. Amsterdam: Elsevier Publishing.

Goodman, N. (1949). On likeness of meaning. *Analysis, 10*(1), 1–7.

Goodman, N. (1954/1983). *Fact, fiction, and forecast* (4th ed.). Cambridge, MA: Harvard University Press.

Goodman, N., & Quine, W. V. (1947). Steps toward a constructive nominalism. *Journal of Symbolic Logic, 12*, 105–122. Reprinted in Nelson Goodman, *Problems and Projects*, Bobbs-Merrill, 1972, pp. 173–198.

Gorrell, P. (1991). Subcategorization and sentence processing. In R. C. Berwick, S. P. Abney, & C. Tenny (Eds.), *Principle-based parsing: Computation and psycholinguistics* (pp. 279–300). Dordrecht: Kluwer Academic Publishers.

Greenberg, M., & Harman, G. (2006). Conceptual-role semantics. In *The Oxford handbook of philosophy of language*. Oxford: Oxford University Press.

Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences, 23*, 1–71.

Guasti, M. T. (2002). *Language acquisition: The growth of grammar*. Cambridge, MA: MIT Press.

Gunderson, K. (Ed.). (1975). *Minnesota studies in the philosophy of science, volume VII: Language, mind, and knowledge*. Minneapolis: University of Minnesota Press.

Hadley, R. (1994). Systematicity in connectionist language learning. *Mind and Language, ix*, 247–272.

Haegeman, L. (1994). *Introduction to Government and binding theory*, 2nd ed. Oxford: Blackwell Publishers, 1st ed. 1991.

Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research, 13*, 339–356.

Hale, J. T. (1999). Dynamical parsing and harmonic grammar. Unpublished report, available at the following URL: courses.cit.cornell.edu/jth99/prism.ps

Hale, J. T. (2001). A probabilistic parser as a psycholinguistic model. In *Proceedings of the NAACL*, available at: http://www.aclweb.org/anthology/N/N01/N01-1021.pdf

Hale, J. T. (2003). *Grammar, uncertainty and sentence processing*. Unpublished Ph.D. dissertation, Johns Hopkins University.

Hale, J. T. (2011). What a rational parser would do. *Cognitive Science, 35*(3), 399–344.

Hale, J. T., & Smolensky, P. (2006). Harmonic grammars and harmonic parsers for formal languages, ch. 10. In P. Smolensky & G. Legendre (Eds.), *The harmonic mind* (Vol. 1, pp. 393–415). Cambridge, MA: MIT Press.

Halle, M., Bresnan, J., & Miller, G. A. (Eds.). (1978). *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.

Harkema, H. (2001). *Parsing minimalist languages*. PhD dissertation, UCLA. Available at: http://www.linguistics.ucla.edu/people/stabler/paris08/Harkema01.pdf

Harman, G. (1967). Review: Psychological aspects of the theory of syntax. *The Journal of Philosophy, 64*(2). Some Recent Issues in Linguistics, pp. 75–87.

Harman, G. (1969). Reply to Chomsky's reply to the review. In S. Hook (Ed.), *Language and philosophy: A symposium*. New York: New York University Press.

Harman, G. (1972/1999). Meaning and semantics. In M. Munitz & P. Unger (Eds.), *Semantics and philosophy*. New York: New York University Press. Reprinted in Reasoning, Meaning, and Mind, OUP. (esp. p. 202).

Harman, G. (1973). *Thought*. Princeton: Princeton University Press.

Harman, G. (1974a). Review of language and mind. In Gilbert Harman (Ed.), *On Noam Chomsky: Critical essays* (pp. 201–218). Garden City: Anchor Press, Doubleday.

Harman, G. (Ed.). (1974b). *On Noam Chomsky: Critical essays*. Garden City: Anchor Press, Doubleday.

Harman, G. (1975). Language, thought, and communication. In K. Gunderson (Ed.), *Language, mind, and knowledge*. Minneapolis: University of Minnesota Press.

Harman, G. (1982). Conceptual role semantics. *Notre Dame Journal of Formal Logic, 23*, 242–257.

Harman, G. (1986). *Change in view: Principles of reasoning*. Cambridge, MA: MIT Press. A pdf version is available at the following URL: http://www.princeton.edu/_harman

Harman, G. (1987). (Nonsolipsistic) conceptual role semantics. In E. LePore (Ed.), *New directions in semantics*. London: Academic.

Harman, G. (1999). *Reasoning, meaning, and mind*. Oxford: Oxford University Press.

Harman, G. (2002). Internal critique: A logic is not a theory of reasoning and a theory of reasoning is not a logic. In D. M. Gabbay, R. H. Johnson, H. J. Ohlbach, & J. Woods (Eds.), *Handbook of the logic of argument and inference: The turn towards the practical*, *Studies in logic and practical reasoning* (Vol. 1, pp. 171–186). Amsterdam: Elsevier.

Harnish, R. M. (2002). *Minds, brains, computers: An historical introduction to the foundations of cognitive science*. Malden: Blackwell.

Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. In *IEEE Transactions of systems science and cybernetics*, ssc-4(2) (pp. 100–107).

Henderson, J. B. (2010). Artificial neural networks. In A. Clark, C. Fox, & S. Lappin (Eds.), *Handbook of computational linguistics and natural language processing* (pp. 221–237). Wiley-Blackwell: Malden.

Hodes, H. T. (1984). Logicism and the ontological commitments of arithmetic. *The Journal of Philosophy, 81*(3), 123–149.

Hodes, H. T. (1990). Where do the natural numbers come from? In memory of Geoffrey Joseph. *Synthese, 84*(3), *The Philosophy of Mathematics*, Part II, pp. 347–407.

Hofweber, T. (2005). Number determiners, numbers, and arithmetic. *The Philosophical Review, 114*(2), 179–225.

Hook, S. (Ed.). (1969). *Language and philosophy: A symposium*. New York: New York University Press.

Horgan, T., & Tienson, J. (1999). Rules and representations. In R. A. Wilson & F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 724–726). Cambridge, MA: MIT Press.

Hornsby, J. (2000). Personal and sub-personal; a defence of Dennett's early distinction. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action, 3*(1), 6–24.

Hornsby, J. (2001). *Simple mindedness: In defense of naive naturalism in the philosophy of mind*. Cambridge, MA: Harvard University Press.

Horwich, P. (1998). *Meaning*. Oxford: Clarendon Press.

Horwich, P. (2005). *Reflections on meaning*. Oxford: Clarendon Press.

Houghton, G. (Ed.). (2005). *Connectionist models in cognitive psychology*. New York: Psychology Press.

Hume, D. (1740/2000). In D. F. Norton & J. Mary (Eds.), *A treatise of human nature*. Norton/New York: Oxford University Press.

Hume, D. (1748/1999). In L. Tom (Ed.), *An enquiry concerning human understanding*. Beauchamp/New York: Oxford University Press.

Hutto, D. D., & Myin, E. (2012). *Radicalizing enactivism: Basic minds without content*. MIT Press.

Jackendoff, R. (2000). Fodorian modularity and representational modularity. In Y. Grodzinsky, L. Shapiro, & D. Swinney (Eds.), *Language and the brain: Representation and processing* (pp. 4–30). San Diego: Academic.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar*. Evolution: Oxford University Press.

Johnson, M. (1989). Parsing as deduction: The use of knowledge of language. *Journal of Psycholinguistic Research, 18*(1), 105–128.

Johnson, M. (1991). Deductive parsing: The use of knowledge of language. In R. Berwick, S. Abney, & C. Tenny (Eds.), *Principle-based parsing*. Dordrecht: Kluwer Academic Publishers.

Jurafsky, D, (1993). *A cognitive model of sentence interpretation: A construction grammar approach*. Technical Report TR-93-077. Berkeley: International Computer Science Institute. Available at the following URL: ftp://ftp.icsi.berkeley.edu/pub/techreports/1993/tr-93-077.ps.gz

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic disambiguation. *Cognitive Science, 20*, 137–194.

Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics*. Cambridge, MA: MIT Press.

Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing: An introduction to speech recognition, computational linguistics, and natural language processing* (2nd ed.). Prentice Hall.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kaplan, R. M. (1973). A general syntactic processor. In R. Rustin (Ed.), *Natural language processing* (pp. 193–241). New York: Algorithmics Press.

Kaplan, R. M. (1987). Three seductions of computational psycholinguistics. In P. Whitelock, M. M. Wood, H. L. Somers, R. Johnson, & P. Bennett (Eds.), *Linguistic theory and computer applications* (pp. 149–188). London: Academic Press.

Kartunnen, D. R., & Zwicky, A. M. (1985). Introduction. In D. R. Dowty, L. Kartunnen, & A. M. Zwicky (Eds.), *Natural language processing: Psychological, computational, and theoretical perspectives*. Cambridge: Cambridge University Press.

Katz, J. J. (1981). *Language and other abstract objects*. Totowa: Rowman & Littlefield.

Katz, J. J. (1984). An outline of a platonist grammar. In J. J. Katz (Ed.), *The philosophy of linguistics*. Oxford: Oxford University Press.

Katz, J. J. (Ed.). (1985). *The philosophy of linguistics*. Oxford: Oxford University Press.

Katz, J. J. (1990). *The metaphysics of meaning*. Cambridge, MA: MIT Press.

Katz, J. J. (1996). The unfinished Chomskyan revolution. *Mind and Language, 11*, 270–294.

Katz, J. J. (2000). *Realistic rationalism*. Cambridge, MA: MIT Press.

Katz, J. J., & Postal, P. M. (1991). Realism vs. Conceptualism in linguistics. *Linguistics and Philosophy, 14*, 515–554.

Kelly, S. D. (2001). The non-conceptual content of perceptual experience: Situation dependence and fineness of grain. *Philosophy and Phenomenological Research, 62*(3), 601–608.

Kilbury, J. (1985). Chart parsing and the Earley algorithm. In U. Klenk (Ed.), *Kontextfreie Syntaxen und verwandte Systeme* (pp. 76–89). Tübingen: Niemeyer. Available at the following URL: user.phil-fak.uni-duesseldorf.de/~kilbury/Publ/Kilbury1985-CPEA.pdf

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition, 2*, 15–47.

King, M. (1983). Transformational parsing. In M. King (Ed.), *Parsing natural language*. London: Academic Press.

Kinsch, W. (1984). Approaches to the study of the psychology of language. In T. G. Bever, J. M. Carroll, & L. A. Miller (Eds.), *Talking minds: The study of language in cognitive science* (pp. 111–145). Cambridge, MA: MIT Press.

Kirsh, D. (1990). When is information explicitly represented? In P. Hanson (Ed.), *Information, language and cognition* (pp. 340–365). University of British Columbia Press, re-issued by Oxford University Press (1992).
Knowles, J. (2000). Knowledge of grammar as a propositional attitude. *Philosophical Psychology, 13*(3), 325–353.
Koster, J., van Riemsdijk, H., Vernaud, J.-R. (1977). GLOW manifesto. *GLOW Newsletter*. Available at http://glow.uit.no/index.php?page=manifesto#glow
Kripke, S. (1972/1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.
Kripke, S. (1982). *Wittgenstein on rules and private language*. Oxford: Blackwell.
Kuhn, T. S. (1962/1996). *The structure of scientific revolutions* (3rd ed.). Chicago: University of Chicago Press.
Kuhn, T. S. (1977). *The essential tension: Selected studies in scientific tradition and change* (pp. 320–329). Chicago: Chicago University Press.
Kukla, A., & Walmsley, J. (2006). *Mind: A historical and philosophical introduction to the major theories*. Indianapolis: Hackett Publishing.
Langendoen, D. T., & Postal, P. (1984). *The vastness of natural languages*. Oxford: Wiley-Blackwell.
Larson, R. K. (2010). *Grammar as science*. Cambridge, MA: MIT Press.
Larson, R. K., & Ludlow, P. (1997). Interpreted logical forms. In P. Ludlow (Ed.), *Readings in the philosophy of language* (pp. 993–1039). Cambridge, MA: MIT Press.
Lawrence, S., Giles, L., & Fong, S. (2000). Natural language grammatical inference with recurrent neural networks. *IEEE Transactions on Knowledge and Data Engineering, 12*(1), 126–140.
Levin, M. (1979). Explanation and prediction in grammar (and semantics). *Midwest Studies in Philosophy, II*, 128–137.
Levine, R. D., & Postal, P. (2004). A corrupted linguistics. In Collier and Horowitz (Eds.), *The anti Chomsky reader* (pp. 203–233). San Francisco: Encounter Books.
Lewis, D. K. (1969/2002). *Convention*. Malden: Blackwell Publishing.
Lewis, D. K. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy, 50*, 249–258.
Lin, D. (1993). Principle-based parsing without overgeneration. Available at the following URL: http://www.aclweb.org/anthology/P93-1016.pdf
Lin, D. (1994). PRINCIPAR: An efficient, broad-coverage, principle-based parser. Available at the following URL: http://www.aclweb.org/anthology/C94-1079
Loar, B. (1990). Phenomenal states. *Philosophical Perspectives, 4*, 81–108.
Loewer, B. (2002). A guide to naturalizing semantics. Available at http://www.nyu.edu/gsas/dept/philo/courses/rules/papers/Loewer.pdf
Ludlow, P. (2011). *The philosophy of generative linguistics*. Oxford: Oxford University Press.
Lyons, W. (1986). *The disappearance of introspection*. Cambridge, MA: MIT Press.
MacDonald, M. C. (1989). Priming effects from gaps to antecedents. *Language and Cognitive Processes, 4*, 1–72.
Macdonald, C., & Macdonald, M. (Eds.). (1995). *Connectionism: Debates on psychological explanation*. Cambridge, MA: Blackwell Publishers.
MacDonald, M. E., Pearlmutter, D., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 678–703.
Machery, E. (2009). *Doing without concepts*. Oxford: Oxford University Press.
Maddy, P. (1990). *Realism in mathematics*. Oxford: Oxford University Press.
Magerman, D. M. (1994). *Natural language parsing as statistical pattern recognition*. PhD thesis, Stanford University.
Magerman, D. M. (1995). Statistical decision-tree models for parsing. In *ACL33* (pp. 276–283).
Magerman, D. M., and Marcus, M. P. (1991). Pearl: A probabilistic chart parser. In *EACL* 4.
Magerman, D. M., & Weir, C. (1992). Efficiency, robustness, and accuracy in picky chart parsing. *ACL, 30*, 40–47.
Manning, C. D. (2003). Probabilistic syntax. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic linguistics*. Cambridge, MA: MIT Press.

Manning, C., & Carpenter, B. (1997). Probabilistic parsing using left corner language models. In *Proceedings of the Fifth International Workshop on Parsing Technologies*.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

Marcus, M. P. (1980). *A theory of syntactic recognition for natural language*. Cambridge, MA: MIT Press.

Marcus, M. P. (1984). Some inadequate theories of human language processing. In T. G. Bever, J. M. Carroll, & L. A. Miller (Eds.), *Talking minds: The study of language in cognitive science* (pp. 253–278). Cambridge, MA: MIT Press.

Marr, D. (1980). *Vision: A computational investigation into the human representation and processing of visual information*. Cambridge, MA: MIT Press.

Marslen-Wilson, W. D., & Tyler, L. K. (1998). Rules, representations, and the English past tense. *Trends in Cognitive Sciences, 2*, 428–435.

Marslen-Wilson, W. D., & Tyler, L. K. (2003). Capturing underlying differentiation in the human language system. *Trends in Cognitive Sciences, 7*, 62–63.

Matthews, R. J. (1994). The measure of mind. *Mind, 103*(410, New Series), 131–146.

Matthews, R. J. (2007). *The measure of mind*. Oxford: Oxford University Press.

May, R. (1985). *Logical form: Its structure and derivation*. Cambridge, MA: MIT Press.

May, R. (1997). Logical form as a level of linguistic representation. In P. Ludlow (Ed.), *Readings in the philosophy of language*. Cambridge, MA: MIT Press.

Mayberry, M. R., & Miikkulainen, R. (2000). Combining maps and distributed representations for shift-reduce parsing. In S. Wermter & R. Sun (eds.), *Hybrid neural systems* (pp. 144–157). Berlin: Springer.

McDonald, F. (2009). Linguistics, psychology, and the ontology of language. *Croatian Journal of Philosophy, 9*, 291–301.

McDonald, F. (2012). Why language exists: Stating the obvious. *Croatian Journal of Philosophy, 34*, 1–12.

McElree, B., & Bever, T. G. (1989). The psychological reality of linguistically defined gaps. *Journal of Psycholinguistic Research, 18*, 21–35.

McRoy, S., & Hirst, G. (1990). Race-based parsing and syntactic disambiguation. *Cognitive Science, 14*, 313–353.

Meng, M., & Bader, M. (2000a). The role of case and number features in syntactic ambiguity resolution. In M. de Vincenzi & V. Lombardo (Eds.), *Cross-linguistic perspectives on language processing* (pp. 175–204). Dordrecht: Kluwer Academic Publishers.

Meng, M., & Bader, M. (2000b). Mode of disambiguation and garden-path strength: An investigation of subject-object ambiguities in German. *Language and Speech, 43*(1), 43–74.

Merlo, P. (1995). Modularity and information content classes in principle-based parsing. *Computational Linguistics, 21*(4), 515–541.

Merlo, P., & Stevenson, S. (Eds.). (2002). *The lexical basis of sentence processing: Formal, computational, and experimental issues*. Amsterdam: John Benjamins Publishing Company.

Mill, J. S. (1874). *A system of logic* (8th ed.). New York: Harper & Brothers.

Millikan, R. G. (1984). *Language, thought and other biological categories: New foundations for realism*. Cambridge, MA: MIT Press.

Millikan, R. G. (2004). In defense of public language. In L. M. Antony & N. Hornstein (Eds.), pp. 215–237.

Milner, J. (1979). Reply to GLOW manifesto. *GLOW Newsletter*. Available at http://glow.uit.no/index.php?page=milner#glow

Moore, R. C. (2000). Improved left-corner chart parsing for large context-free grammars. In *Proceedings of the sixth international workshop on parsing technologies*. Available at the following URL: http://research.microsoft.com/apps/pubs/default.aspx?id=68866

Morawietz, F. (2000). Chart parsing and constraint programming. In *Proceedings of the 18th conference on computational linguistics – Volume 1*. Available at: http://portal.acm.org/citation.cfm?id=990900

Miller, G., & Chomsky, N. (1963). Finitary models of language users. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. New York: Wiley.

Neale, S. (1997). Meaning, grammar, and indeterminacy. *Dialectica, 41*(4).

Neale, S. (forthcoming). *Linguistic Pragmatism*. Oxford: Oxford University Press.

Neander, K. (2006). Naturalistic theories of reference. In M. Devitt & R. Hanley (Eds.), *The Blackwell guide to the philosophy of language* (pp. 374–391). Malden: Blackwell Publishing.

Nederhof, M.-J., & Satta, G. (2010). Theory of parsing. In A. Clark, C. Fox, & S. Lappin (Eds.), *Handbook of computational linguistics and natural language processing* (pp. 105–130). Wiley-Blackwell: Malden.

Neeleman, A., & van de Koot, H. (2010). Theoretical validity and psychological reality of the grammatical code. In H. De Mulder, M. Everaert, O. Nilsen, T. Lentz & A. Zondervan (Eds.), *The linguistic enterprise: From knowledge of language to knowledge in linguistics*. Amsterdam: John Benjamins.

Neville, H. J., Nicol, J., Barss, A., Forster, K., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related potentials. *Journal of Cognitive Neuroscience, 6*, 233–255.

Newell, A., & Simon, H. A. (1963). GPS: A program that simulates human thought. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought*. New York: McGraw-Hill.

Ni, W., Fodor, J. D., Crain, S., & Shankweiler, D. (1998). Anomaly detection: Eye movement patterns. *Journal of Psycholinguistic Research, 27*(5), 515–539.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18*, 5–19.

Noë, A. (2004). *Action in perception*. Cambridge, MA: MIT Press.

O'Donnell, M. (1995). Sentence analysis and generation: A systemic perspective. Unpublished Ph.D. dissertation, University of Sydney, available at: http://www.wagsoft.com/Papers/Thesis/

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.

Ogihara, T. (2003). Adverbs of quantification and sequence of tense phenomena. In J. Gutierrez-Rexach (Ed.), *Semantics: Critical concepts* (Vol. 4, pp. 319–337). London: Routledge.

Orenstein, A. (2002). *W. V. Quine*. Princeton: Princeton University Press.

Osterhaut, L., & Swinney, D. A. (1992). On the temporal course of gap filling during comprehension of verbal passives. *Journal of Psycholinguistic Research, 2*(2).

Papineau, D. (2002). *Thinking about consciousness*. New York: Oxford University Press.

Peacocke, C. (1989). When is grammar psychologically real? In A. George (Ed.), *Reflections on Chomsky* (pp. 111–130). Oxford: Basil Blackwell.

Peacocke, C. (1992). *A study of concepts*. Cambridge, MA: MIT Press.

Pereira, F. C. N. (1985). A new characterization of attachment preferences. In D. R. Dowty, L. Kartunnen, & A. M. Zwicky (Eds.), *Natural language processing*. Cambridge: Cambridge University Press.

Pereira, F. C. N., & Warren, D. H. D. (1983). Parsing as deduction. In *Proceedings of the 22nd annual meeting of the association for computational linguistics*, pp. 137–144.

Pereplyotchik, D. (2011). Psychological and computational models of language comprehension: In defense of the psychological reality of syntax. *Croatian Journal of Philosophy, XI*(31), 31–72.

Pereplyotchik, D. (2016). Representations in language processing: Why comprehension is not brute-causal. *Philosophical Psychology, 29*(2), 277–291.

Pereplyotchik, D. (2017). Sellars and psycholinguistics. In D. Pereplyotchik & D. R. Barnbaum (Eds.), *Sellars and contemporary philosophy*. London: Routledge.

Pereplyotchik, D., & Mylopoulos, M. (2013). Is there any evidence for forward models in language production? *Behavioral and Brain Sciences, 36*(4), 368–369.

Peters, P. S., & Ritchie, R. W. (1973). On the generative power of transformational grammars. *Information Sciences, 10*.

Phillips, C. (1994). Order and structure. MIT Dissertation.

Phillips, C. (1996). Summary of *order and structure.* MIT sissertation. In GLOT International. Available at http://www.ling.umd.edu/colin/research

Phillips, C., & Lasnik, H. (2003). Linguistics and empirical evidence: reply to Edelman and Christiansen. *Trends in Cognitive Sciences, 7*(2), 61–62.

Phillips, C., & Lewis, S. (2013). Derivational order in syntax: Evidence and architectural consequences. *Studies in Linguistics, 6*, 11–47.

Pickering, M. J., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes, 6*, 229–259.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin, 134*(3), 427–459.

Pickering, M. J., & van Gompel, R. P. G. (2006). Syntactic parsing. In M. J. Traxler & M. A. Gernsbacher (Eds.), *The handbook of psycholinguistics* (pp. 455–503). San Diego: Elsevier.

Pickering, M. J., Traxler, M. J., & Crocker, M. W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language, 43*(3), 447–475.

Pietroski, P. (2008). Think of the children: Comments on ignorance of language, by Michael Devitt. *Australasian Journal of Philosophy, 86*, 657–669.

Pinker, S. (1994). *The language instinct: How the mind creates language*. New York: William Morrow and Co..

Pinker, S. (1997). *How the mind works*. New York: W. W. Norton and Company.

Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Postal, P. (2009). The incoherence of Chomsky's 'Biolinguistic' ontology. *Biolinguistics, 3*(1).

Price, H. (2011). *Naturalism without mirrors*. Oxford University Press.

Prinz, J. J. (2000). A Neurofunctional theory of visual consciousness. *Consciousness and Cognition, 9*, 243–259.

Prinz, J. J. (2006). Is the mind really modular? In R. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. 22–36). Malden: Blackwell.

Pritchett, B. (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.

Pullum, G. K. (1991). Review of principle-based parsing: Computation and psycholinguistics. In *Studies in linguistics and philosophy* (Vol. 44, pp. 393–396). Dordrecht: Kluwer Academic Publishers. Available at: www.aclweb.org/anthology/J93-2010.

Pullum, G. K., & Scholz, B. C. (2007). Systematicity and natural language syntax. *Croatian Journal of Philosophy, 7*, 375–402.

Pullum, G. K., & Scholz, B. C. (2010). Recursion and the infinitude claim. In H. van der Hulst (Ed.), *Recursion in human language*. Berlin: Mouton de Gruyter.

Putnam, H. (1960). Minds and machines. In S. Hook (Ed.), *Dimensions of mind: A symposium*. New York: New York University Press.

Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. Merrill (eds.), *Art, mind, and religion* (pp. 37–48). University of Pittsburg Press. Reprinted as "The Nature of Mental States," in *The Nature of Mind*, D. Rosenthal, ed. New York: Oxford University Press, 1991, pp. 197–203.

Putnam, H. (1973, November). Meaning and reference. *The Journal of Philosophy,* 70(19). Seventieth annual meeting of the American Philosophical Association Eastern Division (pp. 699–711).

Putnam, H. (1988). *Mental representation and reality*. Cambridge, MA: MIT Press.

Putnam, H. (2004). *The collapse of the fact/value dichotomy, and other essays*. Cambridge, MA: Harvard University Press.

Pylyshyn, Z. (1986). *Computation and cognition: Toward a foundation for cosnitive Science*. MIT Press.

Quine, W. V. (1953). The problem of meaning in linguistics. In *From a logical point of view* (pp. 47–64), Cambridge, MA: Harvard University Press. Reprinted in Katz, ed. (1985).

Quine, W. V. (1960). *Word and object*. Cambridge, MA: MIT Press.
Quine, W. V. (1969). *Ontological relativity and other essays*. New York: Columbia University Press.
Quine, W. V. (1970). Methodological reflections on current linguistic theory. *Synthese, 21*, 386–398. Reprinted in *Semantics of Natural Language* D. Davidson & G. Harman (eds.).
Quine, W. V. (1974). *The roots of reference*. Lasella: Open Court Press.
Quine, W. V. (1975a). Response to Chomsky. In D. Davidson & J. Hintikka (Eds.), *Words and objections*. Dordrecht: Kluwer Academic Publishers.
Quine, W. V. (1975b). Mind and verbal disposition. In S. Guttenplan (Ed.), *Mind and language* (pp. 83–95). Oxford: Oxford University Press. Reprinted in W. V. Quine, *Quintessence*, R. Gobson (Ed.), Cambridge: Harvard University Press, 2004, pp. 313–325; and again in W. V. Quine, *Confessions of a confirmed extensionalist and other essays*, D. Follesdal & D. B. Quine (Eds.), Cambridge: Harvard University Press, 2004, pp. 244–257.
Quine, W. V. (1976). *The ways of paradox, and other essays*, *Revised and enlarged edition*. Cambridge, MA: Harvard University Press.
Quine, W. V. (1981). The pragmatists' place in empiricism. In Mulvancy & Zeltner (Ed.), *Pragmatism its sources and prospects* (pp. 23–39). University of South Carolina Press.
Quine, W. V. (1986). *Philosophy of logic* (2nd ed.). Cambridge, MA: Harvard University Press.
Quine, W. V. (1992). *Pursuit of truth*. Harvard University Press.
Quine, W. V., & Ullian, J. S. (1978). *The web of belief*. McGraw-Hill Press.
Rattan, G. (2002). Tacit knowledge of grammar: A reply to Knowles. *Philosophical Psychology, 15*(2), 135–154.
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.
Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22*, 358–374.
Rayner, K., Garrod, S., & Perfetti, C. A. (1992). Discourse influences during parsing are delayed. *Cognition, 45*, 109–139.
Rey, G. (2003a). Chomsky, Intentionality and a CRTT. In L. M. Antony & N. Hornstein (Eds.), *Chomsky and his critics*. Chichester: Wiley-Blackwell.
Rey, G. (2003b). Representational content and a Chomskyan linguistics. In A. Barber (Ed.), *Epistemology of language* (pp. 140–186). Oxford: Oxford University Press.
Rey, G. (2006a). The intentional inexistence of language—But not cars. In R. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. 237–255). Oxford: Blackwell Publishers.
Rey, G. (2006b). Conventions, intuitions and linguistic Inexistents: A reply to Devitt. *Croatian Journal of Philosophy, 6*, 549–569.
Rey, G. (2008). In defense of Folieism. *Croatian Journal of Philosophy*, 177–202.
Roark, B., & Johnson, M. (1999). Efficient probabilistic top-down and left-corner parsing. In *Proceedings of the 37th annual meeting of the association for computational; linguistics on computational linguistic*s (pp. 421–428). Available at the following URL: www.ldc.upenn.edu/acl/P/P99/P99-1054.pdf
Robbins, P., & Aydede, M. (Eds.). (2008). *Cambridge handbook of situated cognition*. Cambridge, MA: Cambridge University Press.
Rogers, H. (1967). *Theory of recursive functions and effective computability*. Cambridge, MA: MIT Press.
Rohde, D. L. T. (1999). A connectionist model of sentence comprehension and production. Unpublished PhD thesis proposal, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.
Rohde, D. L. T. (2002). *A connectionist model of sentence comprehension and production*. Ph.D. dissertation, Department of Computer Science, Carnegie Mellon University.
Rosenthal, D. M. (1968). *Intentionality: A study of the views of chisholm and sellars*, PhD. dissertation, Princeton University.

Rosenthal, D. M. (1986). Intentionality. In *Midwest studies in philosophy* (pp. 151–184); postscript in Stuart Silvers (Ed.), *Rerepresentation: Readings in the philosophy of mental representation* (pp. 341–344). Dordrecht: Kluwer Academic Publishers, 1989. Reprinted as ch. 3 in Rosenthal (2005).

Rosenthal, D. M. (1990). On being accessible to consciousness. *Behavioral and Brain Sciences, 13*(4), 621–622.

Rosenthal, D. (2005). *Consciousness and mind*. Oxford: Oxford University Press.

Rosenthal, D. (2008). Consciousness and its function. *Neuropsychologia, 46*(3), 829–840.

Rosenthal, D. M. (2012). Higher-order awareness, misrepresentation, and function. *Philosophocal Transactions of the Royal Society, B*(0), 1–15.

Rosenthal, D. M. (2013). Higher-order awareness, misrepresentation and function. *Philosophical Transaction of Royal Society B, 367*, 1424–1438.

Sadeh-Leicht, O. (2007). The psychological reality of grammar, the theta principle in parsing performance. Doctoral dissertation, Universitat Utrecht.

Sag, I. A., & Fodor, J. D. (1994). Extraction without traces. In *Proceedings of the 13th West Coast conference on formal linguistics,* pp. 365–384.

Sahoo, K. (2011). Review of constituent structure, by Andrew Carnie. In *Linguist List*. Available at the following URL: http://linguistlist.org/issues/22/22-2415.html

Schabes, Y., Abeille, A., Joshi, A. K. (1988). Parsing strategies with 'Lexicalized' grammars: Application to tree adjoining grammars. In *Proceedings of the 12th international conference on computational linguistics*. Available at: acl.ldc.upenn.edu/C/C88/C88-2121.pdf

Schank, R. (1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology, 3*, 552–631.

Schank, R. (1975a). *Conceptual information processing*. Amsterdam: North-Holland Publishing.

Schank, R. (1979a). Language and memory. *Cognitive Science, 4*, 243–284.

Schank, R. (1979b). Reminding and memory organization: An introduction to MOPs. In *Research report 170*. New Haven: Yale University, Department of Computer Science.

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals, and understanding*. Hillsdale: Lawrence Erlbaum Associates.

Schank, R., & Birnbaum, L. (1984). Memory, meaning, and syntax. In T. G. Bever, J. M. Carroll, & L. A. Miller (Eds.), *Talking minds: The study of language in cognitive science* (pp. 209–251). Cambridge, MA: MIT Press.

Schank, R., & Riesbeck, C. (1978). Comprehension by computer. In W. Levelt & G. Flores d'Arcais (Eds.), *Studies in the perception of language*. Chichester: Wiley.

Schank, R., & Riesbeck, C. (1981). *Inside computer understanding*. Hillsdale: Lawrence Erlbaum Associates.

Schank, R., & Yale AI Project. (1975b). SAM: A story understander. In *Research report 43*. New Haven: Yale University Dept. of Computer Science.

Schank, R., & Riesbeck, C. (1978). Comprehension by computer. In W. Levelt & G. Flores d'Arcais (Eds.), *Studies in the perception of language*. Chichester: Wiley.

Schank, R., & Riesbeck, C. (1981). *Inside computer understanding*. Hillsdale: Lawrence Erlbaum Associates.

Schank, R., Tesler, L., & Weber, S. (1970). Spinoza II: Conceptual case-based natural language analysis. In *Research report AIM-109*. Stanford: Dept. of Computer Science.

Schank, R., Lebowitz, M., & Birnbaum, L. (1980). An integrated understander. *American Journal of Computational Linguistics, 6*, 13–30.

Schiffer, S. (forthcoming). Aphonic terms and the deep problem with Gricean meaning. In G. Ostertag (Ed.), *Meanings and other things: Essays in honor of Stephen Schiffer*. Oxford: Oxford University Press.

Scholz, B. C., & Pullum, G. K. (2006). Irrational nativist exuberance. In R. Stainton (Ed.), *Contemporary debates in cognitive science*. Malden: Blackwell.

Schwartz, S. (1980). *The search for pronominal referents*, *Research Report 10*. New Haven: Yale University, Cognitive Science Program.

Searle, J. (1990). Consciousness, explanatory inversion, and cognitive science. *The Behavioral and Brain Sciences, 13*(4), 585–696, and Author's Response, "Who is Computing with the Brain?," pp. 632–642.

Sellars, W. S. (1953). Inference and meaning. *Mind, 62*(247), 313–338.

Sellars, W. S. (1957). Counterfactuals, dispositions, and the causal modalities. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science* (Vol. 2). Minneapolis: University of Minnesota Press.

Sellars, W. S. (1963a). Empiricism and the philosophy of mind. In *Science perception and reality*. Atascadero: Ridgeview Publishing.

Sellars, W. S. (1963b). Some reflections on language games. In *Science, perception, and reality* (pp. 321–358). Atascadero: Ridgeview Publishing.

Sellars, W. S. (1963c). Empiricism and abstract entities. In P. Schilpp (Ed.), *The philosophy of Rudolf Carnap*. La Salle: Open Court.

Sellars, W. S. (1974). Meaning as functional classification. *Synthese, 27*, 417–437.

Shapiro, L. A. (2004). *The mind incarnate*. Cambridge, MA: MIT Press.

Shieber, S. M. (1985). Evidence against the context-freeness of natural language. *Linguistics and Philosophy, 8*, 333–343.

Shieber, S. M. (1988). A uniform architecture for parsing and generation. In *Proceedings of the 12th international conference on computational linguistics*, Budapest, Hungary, pp. 614–619. Available at the following URL: http://dash.harvard.edu/handle/1/2265286

Shieber, S. M., & Johnson, M. (1993). Variations on incremental interpretation. *Journal of Psycholinguistic Research, 22*(2).

Shieber, S. M., Schabes, Y., & Pereira, F. C. N. (1993). Principles and implementation of deductive parsing. Technical report CRCT TR-11-94. Cambridge, MA: Computer Science Department, Harvard University. Available at: http://arXiv.org/

Slobin, D. (1966). Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior, 5*, 219–227.

Smolensky, P. (1988/1995). Connectionism, constituency, and the language of thought. In C. Macdonald & G. Macdonald (Eds.), *Connectionism: Debates on psychological explanation*. Oxford: Blackwell Publishers.

Smolensky, Paul. (1991). Connectionism, constituency and the language of thought. In Barry Loewer and Georges Rey (Eds.), *Meaning in mind: Jerry Fodor and his critics* (pp. 201–227). Oxford: Basil Blackwell. Reprinted in Macdonald and Macdonald 1995: 164–98.

Smolensky, P. (1995). Reply: Constituent structure and explanation in an integrated connectionist/symbolic cognitive architecture. In C. Macdonald & G. Macdonald (Eds.), *Connectionism: Debates on psychological explanation* (pp. 222–290). Oxford: Blackwell Publishers.

Smolensky, P. (2006). Harmony optimization and the computational architecture of the mind/brain, ch. 1. In P. Smolensky & G. Legendre (Eds.), *The harmonic mind* (pp. 1–61). Cambridge, MA: MIT Press.

Smolensky, P., & Legendre, G. (Eds.). (2006). *The harmonic mind*. Cambridge, MA: MIT Press.

Soames, S. (1984). Linguistics and psychology. *Linguistics and Philosophy, 7*, 155–179.

Soames, S. (1985). Semantics and psychology. In Katz (Ed.), *Philosophy of linguistics*.

Sprouse, J., & Almeida, D. (2013). The role of experimental syntax in an integrated cognitive science of language. In K. Grohmann & C. Boeckx (Eds.), *The Cambridge handbook of biolinguistics.*

Stabler, E. P. (1983). How are grammars represented? *Behavioral and Brain Sciences, 6*, 391–402.

Stabler, E. P. (1984). Review of Berwick and Weinberg. *Cognition, 17*, 155–179.

Stabler, E. P. (1992). *The logical approach to syntax: Foundations, specifications, and implementations of theories of government and binding*. Cambridge, MA: MIT Press.

Stabler, E. P. (1994). The finite connectivity of linguistic structure. In C. Clifton Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 303–336). Hillsdale: Lawrence Erlbaum Associates.

Stabler, E. P. (1997). Derivational minimalism. In C. Retoré (Ed.), *Logical aspects of computational linguistics*, vol. 1328 of *Lecture Notes in Artificial Intelligence*. Berlin/Heidelberg: Springer.
Stabler, E. P. (2001). Minimalist grammars and recognition. In C. Rohrer, A. Rossdeutscher, & H. Kamp (Eds.), *Linguistic form and its computation*. Stanford: CSLI Publications.
Stainton, R. (Ed.). (2006). *Contemporary debates in cognitive science*. Malden: Blackwell.
Steedman, M. (1985, August). LFG and psychological explanation. *Linguistics and Philosophy, 8*(3), 359–385.
Steedman, M. (1992). Grammars and processors. University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-92-53. Available at: http://repository.upenn.edu/cisreports/475
Steedman, M. (1999). Connectionist sentence processing in perspective. *Cognitive Science, 23*(4), 615–634.
Steedman, M. (2000). *The syntactic process*. Cambridge, MA: MIT Press.
Steedman, M. (2002). Connectionist and symbolist sentence processing. In P. Merlo & S. Stevenson (Eds.), *The lexical basis of sentence processing: Formal, computational, and experimental issues* (pp. 95–108). Amsterdam: John Benjamins Publishing Company.
Sterelny, K. (1991). *The representational theory of mind: An introduction*. Oxford: Wiley-Blackwell.
Stevenson, S. (1993) A competition-based explanation of syntactic attachment preferences and garden path phenomena. In *Proceedings of the 31st annual association for computational linguistics,* pp. 266–273.
Stevenson, S. (1999). Bridging the symbolic-connectionist gap in language comprehension. In E. Lepore (Ed.), *What is cognitive science?* (pp. 336–355). Oxford: Blackwell Publishers.
Stevenson, S., & Merlo, P. (1997). Lexical structure and processing complexity. *Language and Cognitive Processes, 12.2*(*3*), 349–399.
Stevenson, S., & Smolensky, P. (2006). Optimality in sentence processing, ch. 19. In P. Smolensky & G. Legendre (Eds.), *The harmonic mind* (pp. 307–337). Cambridge, MA: MIT Press.
Stich, S. (1971). What every speaker knows. *The Philosophical Review, 80*(4), 476–496.
Stich, S. (1972). Grammar, psychology and indeterminacy. *Journal of Philosophy, LXIX*(22), 799–818.
Stich, S. (1973). What every grammar does: A reply to Prof. Arbini. *Philosophia, 3*(1), 85–96.
Stich, S. (1974). Review of *The underlying reality of language and its philosophical import.* By Jerrold J. Katz. *The Philosophical Review, 83*(2), 259–263.
Stich, S. (1975). Competence and indeterminacy. In D. Cohen & J. R. Wirth (Eds.), *Testing linguistic hypotheses*. Washington: Wiley.
Stich, S. (1978a). Empiricism, innateness, and linguistic universals. *Philosophical Studies, 33*(3), 273–286.
Stich, S. (1978b). Beliefs and subdoxastic states. *Philosophy of Science, 45*(4), 499–518.
Stich, S. (1979). Between Chomskian rationalism and Popperian empiricism. *British Journal for the Philosophy of Science, 30*, 329–347.
Stich, S. (1980). What every speaker cognizes. *Behavioral and Brain Sciences, 3*, 39–40.
Stich, S. (1981). Can Popperians learn to talk? *The British Journal for the Philosophy of Science, 32*(*2*), *157–164*.
Stich, S. (1985). *From folk psychology to cognitive science: The case against belief*. Cambridge, MA: MIT Press.
Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes, 3*, 227–245.
Strawson, G. (1994). *Mental reality*. Cambridge: MIT Press.
Stroud, C., & Phillips, C. (2010). Examining the evidence for an independent semantic analyzer: An ERP study in Spanish
Sturt, P., Pickering, M. J., Scheepers, C., & Crocker, M. W. (2001). The preservation of structure in language comprehension: Is reanalysis the last resort? *Journal of Memory and Language, 45*, 283–307.

Tanenhaus, M. K., Carlson, G. N., & Seidenberg, M. S. (1985). Do listeners compute linguistic representations? In D. R. Dowty, L. Kartunnen, & A. M. Zwicky (Eds.), *Natural language processing: Psychological, computational, and theoretical perspectives*. Cambridge, MA: Cambridge University Press.

Thomas, J. D. (1995). Center-embedding and self-embedding in human language processing. MS thesis, MIT.

Tomasello, M. (2005). Beyond formalities: The case of language acquisition. *The Linguistic Review, 22*, 183–197.

Townsend, D. J., & Bever, T. D. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.

Traxler, M. J., & Pickering, M. J. (1996). Case marking in the parsing of garden-path sentences: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology, 49A*(4), 991–1004.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language, 33*, 285–318.

Tye, M. (1995). *Ten problems of consciousness: A representational theory of the phenomenal mind*. Cambridge, MA: MIT Press.

Tyler, L. (1980). *Serial and interactive parallel theories of syntactic processes* (MIT Center for Cognitive Studies Occasional Paper No. 8).

Tyler, L., & Marslen-Wilson, W. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 683–692.

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., & Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience, 9*, 266–276.

Uriagereka, J. (1999). *Multiple spell-out*. College Park: University of Maryland.

van Gaal, S., Naccache, L., Meuwese, J. D. I., van Loon, A. A. M., Leighton, A. H., Cohen, L., & Dehaene, S. (2014). Can the meaning of multiple words be integrated unconsciously? *Philosophical Transactions of the Royal Society, 369*(1641).

Van Gelder, T. (1990). Compositionality: A connectionist variation on a classical theme. *Cognitive Science, 14*, 355–384.

Van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language, 52*, 284–307.

Vasishth, S., & Lewis, R. (2006). Symbolic models of human sentence processing. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd ed., vol. 5, pp. 410–419). Elsevier. Available at: http://www-personal.umich.edu/~rickl/pubs/vasishth-lewis-2006-ell-article.pdf

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition, 75*, 105–143.

Vosse, T., & Kempen, G. (2009). The unification space implemented as a localist neural net: Predictions and error-tolerance in a constraint-based parser. *Cognitive Neurodynamics, 3*, 331–346.

Walenski, M. (2002). *Relating parsers and grammars: On the structure and real-time comprehension of English infinitival complements*. PhD Dissertation, UC San Diego.

Wanner, E. (1980). The ATN and the sausage machine: Which one is baloney? *Cognition, 8*, 209–225.

Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. Halle, J. W. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality*. Cambridge: MIT Press.

Wasow, T. (1978). On constraining the class of transformational languages. *Synthese, 39*, 81–104.

Weinberg, A. (1999). A minimalist theory of human sentence processing. In S. D. Epstein & N. Hornstein (Eds.), *Working minimalism* (pp. 283–315). Cambridge, MA: MIT Press.

Weiskopf, D. A. (2009). The plurality of concepts. *Synthese, 169*(2009), 145–173.
Wilson, M. A. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*, 625–636.
Woods, W. (1973). An experimental parsing system for transition network grammars. In R. Rustin (Ed.), *Natural language processing* (pp. 111–154). New York: Algorithmics Press.
Yang, C., & Berwick, R. C. (1996). Principle-based parsing for Chinese. In *Language, Information and Computation* (PACLIC 11) (pp. 363–371). Available at: www.aclweb.org/anthology-new/Y/Y96/Y96-1038.pdf
Zorzi, M. (2005). Computational models of reading. In G. Houghton (Ed.), *Connectionist models in cognitive psychology* (pp. 403–444). New York: Psychology Press.