MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. T.R. No. 1260 October 1990

COMPUTATIONAL STRUCTURE OF HUMAN LANGUAGE

Eric Sven Ristad

The central thesis of this report is that human language is NP-complete. That is, the process of comprehending and producing utterances is bounded above by the class $\mathcal{NP}$, and below by NP-hardness. This constructive complexity thesis has two empirical consequences. The first is to predict that a linguistic theory outside $\mathcal{NP}$ is unnaturally powerful. The second is to predict that a linguistic theory easier than NP-hard is descriptively inadequate.

To prove the lower bound, I show that the following three subproblems of language comprehension are all NP-hard: decide whether a given sound is possible sound of a given language; disambiguate a sequence of words; and compute the antecedents of pronouns. The proofs are based directly on the empirical facts of the language user's knowledge, under an appropriate idealization. Therefore, they are invariant across linguistic theories. (For this reason, no knowledge of linguistic theory is needed to understand the proofs, only knowledge of English.)

To illustrate the usefulness of the upper bound, I show that two widely-accepted analyses of the language user's knowledge (of syntactic ellipsis and phonological dependencies) lead to complexity outside of $\mathcal{NP}$ (PSPACE-hard and Undecidable, respectively). Next, guided by the complexity proofs, I construct alternate linguistic analyses that are strictly superior on descriptive grounds, as well as being less complex computationally (in $\mathcal{NP}$).

The report also presents a new framework for linguistic theorizing, that resolves important puzzles in generative linguistics, and guides the mathematical investigation of human language.

Acknowledgements:
My greatest debt is to my teachers, to whom this report is dedicated: to Robert C. Berwick, who got me started in this work, carefully read long drafts on short notice, and sheparded me through MIT into the faculty job of my choice; to Noam Chomsky, who awakened me to the ruthless pursuit of rational thought and empirical truth, was always willing to argue with me, and has single-handedly dragged the study of thought and power out of the slough of idiocy which appears to be its natural resting place; to John Cocke, who was my first mentor; to Morris Halle, who taught me all the phonology I know, including remedial sessions at no extra cost, and always tried to understand what I couldn't always express; and to Jorma Rissanan, who graciously taught me the philosophy and practice of statistical modelling with the minimum description length principle, all in one summer.

The ideas of the report exist only due to the help and intellectual company of Sandiway Fong, James Higginbotham, Howard Lasnik, and Alec Marantz.

A special thank you to my thesis readers Peter Elias and Michael Sipser for their help; to Eric Grimson for being a peerless academic advisor; and to Patrick Winston for clarifying my points and helping me out in times of distress.

Stephen Anderson, Viviane Deprez, Ken Hale, Hilda Koopman, and Peter Sells have shared their invaluable knowledge of the world's pronominal systems with me.

For their assistance, thanks to: Ezra Black, Piroshka Csuri, Faith Frick, Alessandra Giorgi, Maria-Theresa Guasti, Ron Greenberg, Ray Hirschfeld, Norbert Hornstein, Dave Kirsh, Richard Larson, Tomas Lozano-Perez, Scott Meredith, Albert Meyer, Geoff Pullum, Tanya Reinhart, Sally Richter, Giorgio Satta, Barry Schein, Deb Sterling, Chris Tancredi, Esther Torrego, Amy Weinberg, and Ken Wexler.

The author has been supported by a generous four year graduate fellowship from the IBM Corporation, for which he is extremely grateful. This report describes research done in part at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the Laboratory's artificial intelligence research has been provided in part by a grant from the Kapor Family Foundation, in part by NSF Grant DCR-85552543 under a Presidential Young Investigator Award to Professor Robert C. Berwick, and in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-85-K-0124.

This report is a slightly revised version of a thesis submitted to the MIT Department of Electrical Engineering and Computer Science in May 1990, in partial fullfillment of the requirements for the degree of Doctor of Philosophy. The major additions are a new lower-bound proof for the anaphora problem in section 4.1, and the philosphical discussion of appendix A.1.

Preface: Mathematical Analysis in the Natural Sciences

The purpose of this report is to elucidate the structure of human language, in terms of the mathematics of computation and information. Its central thesis is that human language, the process of constructing structural descriptions, is NP-complete. The body of the report is devoted to defending this point.

Any such mathematical investigation of what is fundamentally a topic in the natural sciences must be both relevant and rigorous. It must simultaneously satisfy the seemingly incompatible standards of mathematical rigor and empirical adequacy. Of the two criteria, relevance is the more important, and judging by historical example, also the more difficult to satisfy. If an investigation is not rigorous, it may be called into question and the author perhaps asked to account for its lack of rigor; but if it is not relevant, it will be ignored entirely, properly dismissed as inappropriate mathematics.

In order to ensure relevance, such an investigation must demonstrate a comprehensive understanding of the natural science (linguistics, in this case) and provide two warrants.

The first warrant is a conceptual framework for the investigation, so that the mathematics is used to answer relevant questions. The framework must include a technique for performing the analysis in the domain of the chosen natural science. This technique must ensure that the insights of the science are preserved, and must be sufficiently general so that others can extend the investigation.

The second warrant is a contribution to the natural science itself. Such a contribution might take the form of a simply-stated mathematical thesis that is an independent guide to scientific investigation in the chosen domain. To be useful, this thesis must make strong predictions that are easily falsified in principle, but repeatedly confirmed in practice. Only under these conditions is it possible to develop confidence in such a thesis.

An exemplary mathematical investigation into human language may be found in the work of Noam Chomsky. In *Three Models for the Description of Language*, Chomsky (1956) defined a framework within which to examine the empirical adequacy of formal grammars. He posed the following questions: Do human languages require linguistic descriptions that are outside the range of possible descriptions? Can reasonably simple grammars be constructed for all human languages? Are such grammars revealing, in that they support semantic analysis and provide insights into the use and understanding of language? This is the first warrant.

The article also provided the second warrant, in the form of a simply-stated complexity thesis, namely that human language has a finite—but no finite-state—characterization, and that the simplest and most revealing characterization is given by a class of unrestricted rewriting systems. The essay demonstrates an unqualified commitment to understanding human language. See, by way of contrast, Curry (1961), Lambek (1961), Peters and Ritchie (1973), and Plátek and Sgall (1978).

In this report, I argue that language is the process of constructing linguistic representations from extra-linguistic evidence, and that this process is NP-complete. This brings Chomsky's 1956 complexity thesis up-to-date, with the advantages of a better understanding of language (the result of thirty years of productive research in linguistics) and a more precise theory of structural complexity, based on computational resources rather than on the format of grammars or automata. The resulting thesis is also much stronger, providing tight upper and lower bounds, and therefore is a truly *constructive* complexity thesis for human language.

The complexity thesis is defended with a novel technique, called the direct analysis, that may be contrasted to prior analyses, which have all been indirect. An *indirect analysis* is an analysis of a formal system within which linguistic knowledge may be represented, a kind of programming language for natural language processing. Indirect proofs are based on the ad-hoc particulars of the formal system, and only very tenuously (if at all) on empirical facts. Indirect analyses of human language may be found in Peters and Ritchie (1973) and Barton, Berwick and Ristad (1987). In contrast, a *direct analysis* is a mathematical analysis of linguistic knowledge itself. In it, proofs are based directly on the empirical facts of linguistic knowledge, and therefore are invariant with respect to our scientific ignorance. To the best of my knowledge, this report contains the first direct complexity analyses of human language, which provides the first warrant.

The complexity thesis makes strong predictions, because many proposed linguistic theories violate it, and because the $\mathcal{NP}$ lower bound is in sharp contrast to the prevailing belief that language is efficient, which is held by many linguists psycholinguists, and computational linguists. In the body of the report, I prove the lower bound in three distinct domains, using direct complexity analyses. I demonstrate the utility of the $\mathcal{NP}$ upper bound by using it to guide the revision of the segmental theory of phonology, and of the copy-and-link theory of syntactic ellipsis. This provides the second warrant.

Contents

Chapter 1

Foundation of the Investigation

The chief goal in the computational study of human language is to design a computational model of the language user, that explains the processes of comprehending, producing, and acquiring languages. The central obstacle that we encounter in such an investigation, and in engineering systems for language perception, is our incomplete scientific understanding. In order to overcome this obstacle, we need an independent *a priori* criterion to guide the design of our language models and the revision of our linguistic theories.

The goal of this report is to characterize the complexity of human language, using only empirical facts of linguistic knowledge. Such a complexity thesis, with tight upper and lower bounds, yields a useful design criterion. It objectively measures the significance of a particular design decision, in terms of its effect on the complexity of the language model. It tells us when a language model is too restrictive: language models that do not satisfy the lower bound do not have an adequate account of complex linguistic phenomena. Any particular design decision or empirical generalization may of course violate the lower bound, which is relevant only to the model in its entirety. A complexity thesis also tells us what is a reasonable empirical generalization: that in the absence of overwhelming counterevidence, empirical generalizations must conform to the upper bound.

This introduction establishes the foundation of the research, on which a constructive complexity thesis for human language is built. According to the guidelines set forth in the preface, the chapter must accomplish three things: first, provide a conceptual framework that poses relevant questions for mathematical analyses; second, introduce and motivate the central thesis of the report, that human language is NP-complete; and third, discuss the formal technique of a direct complexity analysis, that will preserve the insights of linguistics in the details of the complexity analysis. Let us consider each in turn.

1.1 The conceptual framework

What is language? According to generative linguistics, language is a cognitive system of knowledge. The generative grammar of a particular human language enumerates all and only the possible (complete, grammatical) structural descriptions of that language. These structural descriptions are acquired on the basis of experience, and put to use in the production and comprehension of expressions.

An important component of linguistic knowledge is knowledge of linguistic dependencies, that is, how the parts of a linguistic form depend on each other. For example, speakers of English know that the subject and main verb of the matrix clause agree with each other: knowing that the subject is plural predicts that the main verb is also marked plural, even though this marking may not be overt. Speakers of some English dialects also know that voiced consonants are immediately preceded by long vowels; that /t/ and /d/ are both pronounced as the voiced flap [D] after a stressed vowel, if another vowel follows; and that this voicing depends on the vowel lengthening process. This knowledge of phonological dependencies explains why pairs of words like *writer,rider* and *latter,ladder*, are distinguished phonetically only by the length of their first vowel, while related pairs such as *write,ride* maintain the underlying voicing distinction between /t/ and /d/. From an information-theoretic perspective, then, a generative grammar provides a constructive characterization of the informational dependencies in the surface forms (ie., the expressions) of a particular language.

The generative framework poses a number of conceptual puzzles, that must be resolved in order to understand what language is.

- The first puzzle is the relation between production and comprehension. Why people can learn to comprehend and speak the same language?

And why can a sound mean the same thing to the person who spoke it and the person who heard it? Generative linguistic theory postulates that these two abilities, production and comprehension, share one component: knowledge of language. But why should Broca's and Wernicke's Areas, the two distinct regions of the brain that seem to perform these tasks, have the same knowledge of language? And how can one 'knowledge component' solve the entirely different problems encountered in production and comprehension? In electronic communication systems, the transmitter and receiver rarely even share hardware because they perform such entirely different functions. They interact successfully only because the human designer intended them to.

- A related puzzle is to explain comprehension independent of production, and vice-versa. That is, how can we explain what it means to comprehend an utterance without direct access to the intentions of the producer? To comprehend an utterance cannot mean to find the exact structural description in the producer's head, because that is never available. It cannot mean to find *some* structural description of the utterance, because this allows the null structural description as a trivial (non)solution. Nor can it mean to find *all* possible structural descriptions for the utterance, because this does not tell us which one is the "intended" structural description.

- A third puzzle is how language can be at once so complex and yet easy to use. Consider the sentence [*Bill expected to see him*]. The pronoun *him* cannot refer to *Bill*. But as a part of another sentence, *I wonder who* [*Bill expected to see him*], it can. Linguistics tells us that complex systems are needed to describe this kind of complex phenomena. Computer science tells us that complex systems do not perform effortlessly. Yet language processing seems to be efficient. How can this be? This is Cordemoy's paradox, named after the Cartesian linguist Géraud de Cordemoy who observed, "We can scarce believe, seeing the facility there is in speaking, that there should need so many parts to be acted for that purpose: But we must accustom ourselves by admiring the structure of our Body, to consider that 'tis made by an incomparable workman, who is inimitable." (1667, pp.84–5)

- The fourth puzzle is how is linguistic knowledge used in actual performance. People produce, comprehend, and acquire languages. These are the empirical phenomena in need of scientific explanation. Generative linguistics attempts to explain these phenomena by postulating theories of linguistic knowledge. But how is this knowledge used? What is the exact relation between a theory of knowledge and a model of the language user? Lacking an answer to this central scientific question, the theory is at best incomplete; at worst, it is incoherent.

These puzzles are best resolved by the conceptual framework within which particular theories are to be proposed. A framework does not itself answer empirical questions. Rather, it explains how these questions will be answered by particular theories.

The conceptual framework of this mathematical investigation is may be summarized as follows. The language faculty is a mental organ that performs a computation, which we may call "language." Language is the process of constructing representations from evidence. The central questions that arise in such an approach to language are: what is evidence, what are representations, and what is the relation between representation and evidence? I adopt an information-theoretic perspective on these issues.

Briefly, the language faculty lies at the interface of several cognitive systems, including the mental lexicon and motor, perceptual, and conceptual systems. The forms continually produced by these cognitive systems are the instantaneous evidence that the language faculty sees. Language, then, is the process of computing the informational dependencies among the codes continually produced by these cognitive systems. A generative grammar is an enumerative code for this instantaneous evidence. Linguistic representations are representations of these dependencies, and the structural description constructed by the language faculty at a particular instant in time is the one that most reduces the apparent information in the instantaneous evidence, that is, the best description of the evidence. An overt expression, whether spoken or written, constitutes only a very small part of the total instantaneous evidence available to the language faculty. The more additional evidence that the language faculty has at the moment of perception, the more complete the structural description of that evidence will be, and the less random the utterance appears to the language faculty.

This *constructive* (as opposed to generative) explication of what language is explains how the preceding puzzles are to be solved:

- Production and comprehension are the same process of representation construction. The central difference between the two is in the direction of data dependencies, and in the distribution of the instantaneous evidence. For example, in production the perceptual system provides less evidence, while in comprehension, it provides more.[1]

- It defines comprehension without reference to the producer's intentions, because comprehension and production always compute the best description of the available evidence. When there is sufficient evidence available (from the cognitive model of the speaker's intentions, the perceptual system, and the priming of lexical entries, for example), then the comprehender constructs the same representation that the producer intended.

- A linguistic representation is the best description of the available evidence. The best description of incomplete or insufficient evidence is an incomplete representation. Therefore, the language faculty need never perform a blind search that can lead to computational intractability. The language faculty does not assign a complete representation to incomplete evidence, even though some complete representation might be consistent with the evidence.[2]

- The constructive framework also suggests a way to understand the

[1]This framework therefore achieves true Cartesian separation between the process of constructing mental representations, and the perceptual evidence that justifies a particular mental representation. Note, however, that the framework fails to provide any understanding of the creative aspects of production that so concerned Cartesians.

[2]On this view, language cannot the computational module that Fodor (1983) argues it is, because it is not informationally encapsulated. Contra Fodor, this is scientifically desirable because it is the only way to explain how language comprehension is possible at all in the face of wild perceptual underdetermination. If the only input to language comprehension was sensation, or even an abstract noise-free expression, then the problem of constructing anything like the intended representation would be ill-posed in a serious way. (The same is true for a statement of the language production problem whose sole input was a representation of "meaning".) As is well-known, a system consisting of computational modules is necessarily inefficient, both computationally and statistically. (Restricting the amount of information available to a module results in a computational inefficiency because that module is unable to prune branches in its computation tree as early as it might otherwise be able to. It results in a statistical inefficiency because a module might need to examine all available evidence in order to determine the optimal estimate, cf. Wax and Kailath, 1985.) The constructive framework suggests an answer to Cordemoy's paradox, namely, language is the process of efficiently constructing the best description of *all* the available evidence. In short, the only modules in human language are abstract *knowledge* modules, not computational modules of any kind.

relationship between generative theory and a computational model of the language user. A generative grammar is a constructive theory of the informational dependencies in the extra-linguistic mental codes. As such, it enumerates the set of complete structural descriptions, and thereby provides a partial, extensional characterization of the relation between extra-linguistic codes and their structural descriptions: "extensional" because a generative grammar only describes the set of possible outputs, and "partial" because this set is limited to complete structural descriptions of complete, noise-free expressions. For these reasons, a generative theory is a necessary first step in the design of an adequate constructive theory of human language.

However, the theory of grammar does *not* specify the function to be computed by the language model. For one, the input to the two computations is not the same. The input to the language model is the set of codes produced by the other cognitive systems; the input to a generative grammar is an underlying form, which is the index of enumeration. Nor are the possible outputs of the two computations the same. The language model assigns a structural description to every input, and therefore the set of possible outputs must include partial descriptions, for inconclusive evidence; the generative grammar only enumerates complete structural descriptions (cf. Jackobson and Halle, 1956). Grammar and language model also specify different relations between a structural description and the expression that is its overt terminal string yield. The generative grammar specifies a relation between complete structural descriptions and complete, noise-free expressions. The language model specifies a relation between structural descriptions and extralinguistic evidence; the relation between structural descriptions and their overt expressions is an almost inconsequential subset of this relation, that also includes structural descriptions for incomplete or noisy expressions.[3]

[3] For these reasons, a parser cannot be a language model, or even part of a language model. A language model is a function from the instantaneous evidence, which looks nothing like an abstract string of terminal symbols, to the best structural description of that evidence. A parser is a function from a string of terminal symbols to that set of structural descriptions whose yield exhausts that symbol-string. For this reason it is plausible to maintain that "the theory of grammar....specifies the function to be computed by the parser" (Berwick and Weinberg, 1984:82). However, it is crucial to realize that parsing has almost no relation to the comprehension, production, or acquisition of languages, and therefore is of little or no scientific interest.

> This is simply to say, generative theory is not a model of human language, at any level of abstraction. Rather, it is a model of linguistic knowledge.[4]

Again, it is important to stress that this is a framework for addressing scientific questions and performing mathematical analysis, not a scientific theory itself. The substantive answers to these puzzles lie in the next generation of constructive linguistic theories.

However, for our present purposes this constructive framework (that language is the process of constructing representations of extralinguistic inputs) poses relevant questions for mathematical analysis. By equating comprehension and production in a fundamental manner, the framework says that the language comprehension problem is a proxy for language as a whole. Therefore, we are assured that the mathematical analysis of subproblems of language comprehension will be relevant to language as a whole.

To view human language from an information-theoretic perspective as we have done here is not to say that language is designed for communication, or even that it's primary use is to communicate. Nor do we claim that language is a general-purpose computer or that language is designed for computation, when we analyze the computational structure of language. The goal of this research is to understand human language in its own terms; computer science and information theory provide useful technical metaphors. That is, important aspects of human language receive an insightful interpretation in terms of computation and information.

[4]These points are subtle, and have confused many people. No less a scientist than David Marr (1980) has confused the competence–performance distinction of generative linguistics with levels of computational abstraction. But, as we have seen, the relationship between competence and performance is not one of abstraction. Competence and performance are simply entirely different classes of computations, both of which may be described at different levels of abstraction. For alternate interpretations of the relation between generative grammar and human language, at odds with the one presented here, see Chomsky (1965), Chomsky (1980), Stabler (1983), Berwick and Weinberg (1984), and Ristad and Berwick (1989).

1.2 The constructive complexity thesis

The central thesis of this work is that human language has the structure of an NP-complete problem.[5,6] An NP-complete problem is hard to solve because the input to the problem is missing some crucial information (the efficient witness), but once the efficient witness (the solution) is found, it is easily verified to be correct. As stressed by the great nineteenth century linguist Wilhelm von Humboldt, every sound uttered as language is assigned a complete meaning and linguistic representation in the mind of the producer.[7] It is the task of comprehension to find the intended representation, given only the utterance. When the utterance is missing crucial disambiguating information, and there are global dependencies in the structural description, then the task of finding the intended representation quickly becomes very difficult. (In effect, we are able to prove the lower bound, that language is NP-hard, because the amount of useful information in the evidence is not a parameter of current linguistic theories; all reductions below take advantage of this fact, which became obvious only in retrospect.) Yet we know comprehension cannot be too difficult, simply because there is always an efficient witness, namely the linguistic representation from which the utter-

[5] This complexity thesis is constructive because the upper and lower bounds are tight enough to tell us exactly where the adequate linguistic theories are, not only where they are *not*. Here "constructive" means "useful." It is in contrast to Chomsky's 1956 complexity thesis, which is not as useful because—as argued in Chomsky (1965)—the upper bound is very loose, the lower bound is weak, and the formal language theory of structural complexity is not sufficiently precise. The conceptual framework presented in the preceding section is constructive because it outlines a model of the language user, that would directly characterize the computations performed in the production, comprehension, and acquisition of human languages. There "constructive" means "given explicitly, by construction." It is in contrast to the generative framework, which provides a partial, extensional characterization of production, comprehension, and acquisition.

[6] As should be clear from the discussion, NP-completeness is not the stigma that many apparently think it is. In fact, exactly the opposite is true. The thesis argues that more efficient language models are fundamentally inadequate, barring of course a revolution in our understanding of language or of nondeterminism.

[7] "The sentence is not to be constructed, is not to be gradually built up of components, but is to be expressed all at once in a form compressed to unity.... Man inwardly relates a complete meaning with every sound emitted as language: that is, for him it is a complete utterance. Man does not intentionally emit merely an isolated word, even though his statement according to our viewpoint may only contain such an entity." (von Humboldt, 1836:110–111) (The upper bound of Chomsky's 1956 complexity thesis, that language has a finite description, also appears to be motivated by an observation due to von Humboldt, that language is the "infinite use of finite means.")

ance was produced. If only the comprehender had the same evidence that the producer did, then he would be able to efficiently compute the intended structural description, because the producer did.

The central empirical consequence of this thesis is that scientifically adequate language models must be NP-complete, under appropriate idealizations (see appendix A.2). If a linguistic system is outside $\mathcal{NP}$, say PSPACE-hard, then the thesis predicts that the system is unnaturally powerful, perhaps because it overgeneralizes from the empirical evidence or misanalyzes some linguistic phenomena. Such a system must be capable of describing unnatural languages. If, however, a complete system is easier than NP-hard, and assuming $\mathcal{P} \neq \mathcal{NP}$, then the system is predicted to be unnaturally weak, most likely because it does not adequately account for some complex linguistic phenomena. Such a system will not be able to describe all human languages. Otherwise the system is NP-complete and is potentially adequate, pending the outcome of more exacting tests of scientific adequacy.

The thesis is weakened if any of the formal arguments presented in this dissertation are refuted. It is falsified if either of its central predictions is falsified. That is, someone must exhibit a comprehensive theory of human language and prove it to not have the structure of an NP-complete problem, or someone must exhibit some complex linguistic phenomena and argue that its complexity is outside $\mathcal{NP}$.

This complexity thesis, then, is an independent guide to the study of language. It is useful because it is a simple decision procedure with which to evaluate linguistic systems, both theoretical and and implemented. The student of language will find it helpful, because, as this dissertation demonstrates, linguistic analyses that have complexity outside of $\mathcal{NP}$ are ripe for reanalysis.

1.3 Direct complexity analysis

The logical next step is to establish this NP-completeness thesis. The central technical obstacle encountered in this work, and in all research on language, is the incomplete nature of our scientific understanding of language. For this reason, it is not clear how to precisely define any computational problem related to human language at all. Our understanding of human language is neither comprehensive, detailed, nor stable. Any formal model of language, obtained perhaps by formalizing some particular linguistic theory, will be based as much on our scientific ignorance as on our understanding. Consequently, no meaningful or comprehensive formalization is possible, and any mathematical analysis of such a formal system would have little or no relevance to language itself.

To overcome this difficulty, we must seek an analysis that is invariant with respect to our ignorance. That way, future work may enrich our analysis, but not falsify it. We may accomplish this objective with a direct analysis.

A *direct analysis* is a mathematical analysis that relies directly on well-understood empirical arguments about the language user's knowledge of language in order to prove mathematical properties of that knowledge. It does not rely on a complete formal model. In a direct analysis, we use the scientific methods of linguistics to construct the simplest theory of a natural, well-understood class of linguistic knowledge, and then analyze the properties of this knowledge. We perform such direct analyses in section 2.2 for the language user's knowledge of suprasegmental phonological dependencies, and in chapter 4 for knowledge of referential dependencies. Slightly less direct analyses may be found in section 2.1 and chapter 3.

Prior mathematical analyses of language have all been highly indirect, proving properties of formal systems within which theories of linguistic knowledge may be represented. This is like analyzing the properties of FORTRAN in order to better understand QUICKSORT.

A related difficulty is that it is not known how to define language comprehension (LC) without reference to the producer's intentions. As mentioned above, the real solution to this difficulty must be provided by the next generation of linguistic theories. The temporary solution adopted here is to select subproblems of LC that may be defined independent of the producer's intentions, and are necessary subproblems of any reasonable constructive theory of comprehension. (Given the "reasonableness" equivocation in the second clause, it is admittedly more a matter of art than science to define such subproblems.)

1.4 Summary of the report

The technical content of the report is apportioned into three chapters and two appendices:

Chapter 2 examines the computational structure of phonological dependencies. It begins by establishing the undecidability of both generation and recognition problems for the segmental model of phonology. Guided by the complexity analysis, we propose a broad range of substantive restrictions, that result in a more natural segmental model whose generation and recognition problems are both in $\mathcal{NP}$. Next, a direct analysis is provided for the complexity of suprasegmental dependencies. The chapter closes by revealing the indispensable role of a complexity thesis in the design of a theory of human knowledge, to guard against gross overgeneralizations.

Chapter 3 examines the LC problem in the domain of morphology and syntax. We demonstrate the art of choosing a subproblem of language comprehension that is relevant despite our inability to define the LC problem itself, and prove the chosen subproblem to be NP-hard. The chapter concludes with a critique of the pursuit of uniform mechanisms in linguistic theory.

Chapter 4 provides a direct analysis of the anaphora problem, which is to determine the intended antecedents of anaphoric elements in a discourse. First we prove in two entirely different ways that this LC subproblem is NP-hard, based on empirical facts of the language user's knowledge of pronominal reference, such as why *John saw him* cannot mean 'John saw John'. Next, we show how a widely-accepted linguistic theory of syntactic ellipsis makes the anaphora problem PSPACE-hard. Finally, guided by the complexity thesis, we falsify this linguistic theory and sketch an empirically superior theory of ellipsis that reduces the complexity of anaphora to inside $\mathcal{NP}$. The conclusion to this chapter critiques an alternate approach to the mathematical investigation of language, based on the complexity analysis of the computational problems posed by linguistic theories.

Appendix A discusses two philosophical issues relevant to the research reported here. First we discuss the implications of complexity classifications for biological computations, and untangle the distinction between competence and performance. Next we examine the *de facto* idealizations to unbounded inputs and unbounded linguistic distinctions, that have been made implicitly in every serious theory of human language, as they must always be.

Appendix B provides another warrant demanded in the preface, that the work make a contribution to the natural science itself. The contribution is an improved generative theory of syntactic ellipsis, with a discussion of a hitherto unnoticed phenomenon of invisible obviation, with important implications for the theory of anaphora. This appendix is an independent, nonmathematical contribution to the field of linguistics.

Chapter 2

Structure of Phonological Knowledge

The goal of this chapter is to elucidate the structure of phonological knowledge, from the perspectives of computer science and information theory.

Linguistic sounds contain predictable information. For example, English speakers invariably aspirate a voiceless stop in the onset of syllable (as illustrated by minimal pairs such as [k^hab]~[gab] and [p^hat]~[bat]) but only when the onset is nonbranching (compare [p^hit]~[spit] and [k^hit]~[skit]). The raised 'h' expresses aspiration, that is, that the segment is pronounced with a slight puff of air. Moreover, vowels are lengthened before voiced consonants (contrast the articulations of *cab*~*cap*, *bag*~*back*, and so forth).

Because these and other phonological dependencies are a part of the language user's unconscious knowledge, they must be represented in an adequate linguistic theory. In generative phonology, dependencies in the surface forms are encoded by a grammar of rewriting rules, using a dictionary of underlying forms. An underlying form represents the true, unpredictable information content of a given surface form. It is constructed by combining (typically, with concatenation or substitution) segmental sequences stored in the dictionary of morphemes. The grammar G derives a surface form $[s]$ from its corresponding underlying form $/u/$ in the dictionary D by repeated application of rules to the intermediate forms $i_1, i_2, \ldots$ of the derivation, as shown in 2.1:

$$\begin{array}{l} \forall\ /u/ \in D \\ \quad G : /u/ \rightarrow /i_1/ \rightarrow /i_2/ \rightarrow \ldots \rightarrow [s] \end{array} \tag{2.1}$$

Each rule represents a natural class of predictable phonological information. Continuing with our example, the generative grammar of an English speaker must contain rules for aspiration and lengthening. The aspiration rule rewrites the underlying form /kab/ as the intermediate form /k^hab/, adding the entirely predictable aspiration information to the voiceless stop k. Next, the lengthening rule applies, rewriting the intermediate form /k^hab/ as the surface form [k^ha:b].

According to generative phonology, the logical problem of language comprehension consists of finding a structural description (that is, an underlying form and a derivation chain) for a given surface form. In effect, comprehension is reduced to the problem of searching for the underlying form that generates a given surface form. When the surface form does not transparently identify its corresponding underlying form, when the space of possible underlying forms is large, or when the grammar is complex, then the logical problem of language comprehension can quickly become very difficult.

The chapter is organized as follows. The next section introduces the segmental model of phonology in some detail, discusses its computational complexity, and proves that even restricted segmental models are extremely powerful (undecidable). Subsequently, we consider various proposed and plausible restrictions on the model, and conclude that plausibly restricted segmental models will be in $\mathcal{NP}$. Section 2.2 introduces the modern autosegmental (nonlinear) model and discusses its computational complexity. We prove that the natural problem of constructing an autosegmental representation of an underspecified surface form is NP-hard.

The central contributions of this chapter are: (i) to analyze the computational complexity of generative phonological theory, as it has developed over the past thirty years, including segmental and autosegmental models; (ii) to suggest a range of restrictions on the segmental model that reduce the complexity of the corresponding language model from undecidable to inside $\mathcal{NP}$; (iii) to resolve some apparent mysteries regarding the SPE evaluation metric and the notion of a linguistically significant generalization; and (iv) to unify the description of suprasegmental processes as establishing the segmental domain within which one 'head' segment in the domain is phonetically distinguished from the nonhead segments in its domain.

2.1 Segmental phonology

We have seen that phonological knowledge includes a grammar of rewriting rules, to derive surface forms from underlying forms. The scientific questions that arise are, what is the class of permissible rewriting rules, and how are they applied in the derivation?

To summarize, the rewriting rules of the segmental phonology are unrestricted and can manipulate morpho-syntactic constituent structure. Rules are ordered into a block and interact with each other. In a derivation the block is repeatedly applied, from the innermost constituent out. Let us examine this system in more detail.[1]

Phonological features are abstract as compared with phonetic representations, although both are given in terms of phonetic features. The set of features includes both phonological features, diacritics, and the distinguished feature `segment` that marks boundaries. Diacritic features are associated with lexical items as a whole; they control the application of rules. An example diacritic is `ablaut`, a feature that marks those stems that must undergo a change in vowel quality, such as tense-conditioned ablaut in the English *sing*, *sang*, *sung* alternation. As noted in SPE, "technically speaking, the number of diacritic features should be at least as large as the number of rules in the phonology. Hence, unless there is a bound on the length of a phonology, the set [of features] should be unlimited."[2] (fn.1, p.390) Features may be specified + or - or by an integral value $1, 2, \ldots, N$ where N is the maximal degree of differentiation permitted for any linguistic feature. The value of N varies from language to language, because languages admit different degrees of differentiation in such features as vowel height, stress, and tone. A set of feature specifications is called a *unit* or sometimes a *segment*. A string of units is called a *matrix* or a *segmental string*.

Suprasegmental relations are relations among segments, rather than properties of individual segments. For example, a syllable is a hierarchical relation between a sequence of segments (the nucleus of the syllable) and the less

[1] The discussion in this chapter is based primarily on the theory presented by Noam Chomsky and Morris Halle (1968) in *The Sound Pattern of English* (SPE). This monumental work defined the field of generative phonology, by formalizing central ideas in the field, including the notions of dependency, process, and linguistically-significant generalization.

[2] Chomsky and Halle continue, "There is no point of principle involved here, and to simplify exposition slightly we shall assume the set to be limited by an a priori condition. A similar comment applied to [the set of specifications]."

sonorous segments that immediately precede and follow it (the onset and coda, respectively).

Rewriting rules formalize the notion of a phonological process, that is, a representation for the dependencies implicit in the surface matrices of a given language. An *elementary rule* is of the form $ZXAYW \rightarrow ZXBYW$ where A and B may be ϕ or any unit, $A \neq B$; X and Y may be matrices (strings of units), and Z and W may be thought of a brackets labeled with syntactic categories such as 'S' or 'N' and so forth.

Some phonological processes, such as the assimilation of voicing across morpheme boundaries, are very common across the world's languages. Other processes, such as the arbitrary insertion of consonants or the substitution of one unit for another entirely distinct unit, are extremely rare or entirely unattested. For this reason, all adequate phonological theories must include an explicit measure of the naturalness of a phonological process. A phonological theory must also define a criterion to decide what constitutes two independent phonological processes and what constitutes a legitimate phonological generalization.

Two central hypotheses of segmental phonology are (i) that the most natural grammars contain the fewest symbols and (ii) a set of rules represent independent phonological processes when they cannot be more compactly combined into a single complex rule (Halle 1961;1962).

A *complex rule*, then, is a finite schema for generating a (potentially infinite) regular set of elementary rules. To a first approximation, the complex rules of SPE are formed by combining units with the operations of union, concatenation, Kleene star, and exponentiation, using variables whose values range over specifications, units, and syntactic categories.[3] Johnson (1972) considers a more powerful class of complex rules (and thereby, a different evaluation metric), as well as alternative modes of rule application.[4]

---

[3]Segmental models belong to the class of nonlocal rewriting systems analyzed by Post, rather than to the local systems due to Thue. This is because complex rules can encode a finite number of nonlocal dependencies, and hence the rewriting activity specified by a complex rule can affect parts of the current derivation string separated by an arbitrary distance.

[4]In Johnson's proposal, the empty string and each unit are schemata; schema may be combined by the operations of union, intersection, negation, Kleene star, and exponentiation over the set of units. Johnson also uses variables and Boolean conditions in his schemata. This "schema language" is a extremely powerful characterization of the class of regular languages over the alphabet of units; it is not used by practicing phonologists.

The complex rules are organized into linear sequence $R_1, R_2, \ldots R_n$; they are applied in order to an underlying matrix to obtain a surface matrix. Ignoring a great many issues that are important for linguistic reasons but irrelevant for our purposes, we may think of the derivational process as follows. The input to the derivation, or "underlying form," is a bracketed string of morphemes, the output of the syntax. The output of the derivation is the "surface form," a string of phonetic units. The derivation consists of a series of cycles. On each cycle, the ordered sequence of rules is applied to every maximal string of units containing no internal brackets, where each R_{i+1} applies (or doesn't apply) to the result of applying the immediately preceding rule R_i, and so forth. Each complex rule R_i itself generates a disjunctive sequence of elementary rules $R_{i,1}, R_{i,2}, \ldots$ in order of increasing generality.[5] That is, R_i generates a sequence of elementary rules where $R_{i,j}$ precedes $R_{i,k}$ in the sequence iff the preconditions of $R_{i,j}$ subsume the preconditions of $R_{i,k}$; the earliest $R_{i,j}$ that can apply to the current derivation matrix is applied to it, to the exclusion of all other $R_{i,k}$. Each elementary rule applies maximally to the current derivation string, that is, simultaneously to all units in the string. For example, if we apply the rule $A \rightarrow B$ to the string AA, the result is the string BB. At the end of the cycle, the last rule R_n erases the innermost brackets, and then the next cycle begins with

---

Because a given complex rule can represent an infinite set of elementary rules, Johnson shows how the iterated, exhaustive application of one complex rule to a given segmental string can "effect virtually any computable mapping," (p.10) that is, *can simulate any TM computation in only one step of the derivation.* Next, he proposes a more restricted "simultaneous" mode of application for complex rules, which is capable of performing at most a finite-state mapping in any single application. The "disjunctive ordering" mode of rule application proposed in SPE is only capable of performing a strictly finite mapping in any single rule application. This mode of application, which is vastly more constrained than either of Johnson's proposals, is also the one used by practicing phonologists. In this chapter we consider the question of what computations can be performed by a finite set of elementary rules, and hence provide very loose lower bounds for Johnson's excessively powerful model. We note in passing, however, that the problem of simply determining whether a given rule is subsumed by one of Johnson's schema is itself wildly intractable, requiring at least exponential space. (The idea of the proof is to construct two complex rules: one generates all possible strings, and the other describes the valid computations of 2^n space-bounded DTM. The latter rule may be constructed by negating the construction due to Hopcroft and Ullman 1979:350ff.)

[5]The interpretation advocated in SPE (p.390ff) is that the true grammar g is the set of elementary rules; the value of this grammar, given by the evaluation metric, is the smallest number of phonological features of any set g' of complex rules that generates the set g. Chapter 9 of Kenstowicz and Kisseberth (1979) contains a less technical summary of the SPE system and a discussion of subsequent modifications and emendations to it.

the rule R_1. The derivation terminates when all brackets are erased.

2.1.1 Complexity of segmental recognition and generation.

Let us say a dictionary D is a finite set of the underlying phonological forms (that is, bracketed matrices) of morphemes. These morphemes may be combined by concatenation and simple substitution (a syntactic category is replaced by a morpheme of that category) to form a possibly infinite set of underlying forms. Then we may characterize the two central computations of phonology as follows.

The *phonological generation problem* (PGP) is: Given a completely specified phonological matrix x and a segmental grammar g, compute the surface form $y = g(x)$ of x.

The *phonological recognition problem* (PRP) is: Given a (partially specified) surface form y, a dictionary D of underlying forms, and a segmental grammar g, decide if the surface form $y = g(x)$ can be derived from some underlying form x according to the grammar g, where x is constructed from the forms in D.

Lemma 2.1.1 *The segmental model can simulate the computation of any DTM M on any input w, using only elementary rules.*

Proof. We sketch the simulation. The underlying form x represents the TM input w, while the surface form y represents the halted state of M on w. The instantaneous description of the machine (tape contents, head position, state symbol) is represented in the string of units. Each unit represents the contents of a tape square. The unit representing the currently scanned tape square will also be specified for two additional features, to represent the state symbol of the machine and the direction in which the head will move. Therefore, three features are needed, with a number of specifications determined by the finite control of the simulated machine M. Each transition of M is simulated by a phonological rule. A few rules are also needed to move the head position around, and to erase the entire derivation string when the simulated machine halts.

There are only two key observations, which do not appear to have been noticed before. The first is that contrary to common misstatement in the linguistics literature, phonological rules are not technically context-sensitive.

Rather, they are unrestricted rewriting rules because they can perform deletions as well as insertions. This is essential to the reduction, because it allows the derivation string to become arbitrarily long. The second observation is that segmental rules can freely manipulate (insert and delete) boundary symbols, and thus it is possible to prolong the derivation indefinitely: we need only employ a rule R_{n-1} at the end of the cycle that adds an extra boundary symbol to each end of the derivation string, unless the simulated machine has halted. The remaining details are straightforward, and are therefore omitted. □The immediate consequences are:

Theorem 1 *PGP is undecidable.*

Proof. By reduction to the undecidable problem $w \in L(M)$? of deciding whether a given TM M accepts an input w. The input to the generation problem consists of an underlying form x that represents w and a segmental grammar g that simulates the computations of M according to lemma 2.1.1. The output is a surface form $y = g(x)$ that represents the halted configuration of the TM, with all but the accepting unit erased. □

Theorem 2 *PRP is undecidable.*

Proof. By reduction to the undecidable problem $L(M) = \phi$? of deciding whether a given TM M accepts any inputs. The input to the recognition problem consists of a surface form y that represents the halted accepting state of the TM, a trivial dictionary capable of generating Σ^*, and a segmental grammar g that simulates the computations of the TM according to lemma 2.1.1. The output is an underlying form x that represents the input that M accepts. The only trick is to construct a (trivial) dictionary capable of generating all possible underlying forms Σ^*. □

Let us now turn to consider the range of plausible formal restrictions on the segmental model.

2.1.2 Restricting the segmental model

We consider ways to bound the length of derivations, limit the number of features, and constrain the form of phonological rewriting rules, as well as their interactions.

The first restriction is to eliminate complex rules. In particular, let us limit complex rules to union and concatenation. This restriction is plausible (for the purposes of this chapter) because complex rules are used to model non-local phonological dependencies, and these dependencies are now modeled by the autosegmental model, which we examine in section 2.2.

Bounding the derivation length

The next restriction is to prevent phonological rules from inserting boundaries. In the SPE formalism, all rules can manipulate boundaries, which are simply those units specified [+**segment**]. However, in the grammars actually postulated by phonologists, only the readjustment rules manipulate boundaries. So let us formally prevent phonological rules from ever inserting or deleting a boundary. Now rules that manipulate boundaries are properly included in the class of readjustment rules.[6]

Boundaries must be manipulated for two reasons. The first is to reduce the number of cycles in a given derivation by deleting boundaries and flattening syntactic structure, for example to prevent the phonology from assigning too many degrees of stress to a highly-embedded structure. The second is to rearrange the boundaries given by the syntax when the intonational phrasing of an utterance does not correspond to its syntactic phrasing (so-called "bracketing paradoxes"). In this case, boundaries are merely moved around, while preserving the total number of boundaries in the string. The only way to accomplish this kind of bracket readjustment in the segmental model is with rules that delete brackets and rules that insert brackets. Therefore, if we wish to exclude rules that insert boundaries, we must provide an alternate mechanism for boundary readjustment. For the sake of argument—and because it is not too hard to construct such a boundary readjustment mechanism—let us henceforth adopt this restriction. Now how powerful is the segmental model?

Although the generation problem is certainly decidable now, the recognition problem remains undecidable, because the dictionary and syntax are both

[6]Not all readjustment rules manipulate boundaries. In general, readjustment rules map the surface forms given by the syntax into the underlying forms of the phonology. For example, they are used to map abstract morphemes, such as inflection or agreement, into phonological matrices, and to modify syntactic categories, as when *Fifth Avenue* is mapped from a noun in the syntax to a noun phrase in the phonology, in order that it be assigned the correct final-stress.

potentially infinite sources of boundaries: the underlying form x needed to generate any given surface form according to the grammar g could be arbitrarily long and contain an arbitrary number of boundaries. Therefore, the complexity of the recognition problem is unaffected by the proposed restriction on boundary readjustments. The obvious restriction then is to additionally limit the depth of embeddings by some fixed constant. (Chomsky and Halle flirt with this restriction for the linguistic reasons mentioned above, but view it as a performance limitation, and hence choose not to adopt it in their theory of linguistic competence.)

Lemma 2.1.2 *Each derivational cycle can directly simulate any polynomial time alternating Turing machine (ATM) computation using only elementary rules.*

Proof. By reduction from a polynomial-depth ATM computation. The input to the reduction is an ATM M with input w. The output is a segmental grammar g and underlying form x s.t. the surface form $y = g(x)$ represents a halted accepting computation iff M accepts w in polynomial time. The major change from lemma 2.1.1 is to encode the entire instantaneous description of the ATM state (that is, tape contents, machine state, head position) in the features of a single unit. To do this requires a polynomial number of features, one for each possible tape square, plus one feature for the machine state and another for the head position. Now each derivation string represents a level of the ATM computation tree. The transitions of the ATM computation are encoded in a block B as follows. An AND-transition is simulated by a triple of rules, one to insert a copy of the current state, and two to implement the two transitions. An OR-transition is simulated by a pair of disjunctively-ordered rules, one for each of the possible successor states. The complete rule sequence consists of a polynomial number of copies of the block B. The last rules in the cycle delete halting states, so that the surface form is the empty string (or reasonably-sized string of 'accepting' units) when the ATM computation halts and accepts. If, on the other hand, the surface form contains any nonhalting or nonaccepting units, then the ATM does not accept its input w in polynomial time. The reduction may clearly be performed in time polynomial in the size of the ATM and its input. $\square$

Because we have restricted the number of embeddings in an underlying form to be no more than a fixed language-universal constant, no derivation can consist of more than a constant number of cycles. Therefore, lemma 2.1.2 establishes the following theorems:

Theorem 3 *PGP with bounded embeddings and elementary rules is PSPACE-hard.*

Proof. The proof is an immediate consequence of lemma 2.1.2 and a corollary to the Chandra-Kozen-Stockmeyer theorem (1981) that equates polynomial time ATM computations and PSPACE DTM computations. □

Theorem 4 *PRP with bounded embeddings and elementary rules is PSPACE-hard.*

Proof. The proof follows from lemma 2.1.2 and the Chandra-Kozen-Stockmeyer result. The dictionary consists of the lone unit that encodes the ATM starting configuration (that is, input w, start state, head on leftmost square). The surface string is either the empty string or a unit that represents the halted accepting ATM configuration. □

There is some evidence that the PGP with bounded embeddings and elementary rules is also inside PSPACE. The requirement that the reduction be polynomial time limits us to specifying a polynomial number of features and a polynomial number of rules. Since each feature corresponds to an ATM tape square and each segment corresponds to an instantaneous description, this kind of simulation seems limited to PSPACE ATM computations. Since each phonological rule corresponds to a next-move relation, that is, one time step of the ATM, the simulation appears further limited to specifying PTIME ATM computations, which correspond to PSPACE DTM computations.[7]

For the PRP, the dictionary (or syntax-interface) provides the additional ability to nondeterministically guess an arbitrarily long, boundary-free underlying form x with which to generate a given surface form $g(x)$. This capacity remains unused in the preceding proof, and it is not too hard to see how it might lead to undecidability.[8]

[7] We must be careful, however, to make our assumptions about the segmental model perfectly explicit. If optional rules are entirely unrestricted, then they can simulate the bounded nondeterminism and the PGP could be as complex as the PRP, which has access to the unbounded nondeterminism of the dictionary. See footnote 8.

[8] As we saw above, the morpheme-dictionary/syntax-interface provides the ability to

Limiting the number of features

Another computational restriction is to limit the number of phonological features.

The number of phonological features has a significant effect on the computational complexity of phonological processes, because each binary feature provides the derivation with a bit of computational space. (This is true even though only a small, fixed number of features were needed to prove undecidability of PRP and PGP; in those reductions, each segment simulated a tape square.) As Chomsky and Halle noted, the SPE formal system is most naturally seen as having a variable (unbounded) set of features and specifications. This is because languages differ in the diacritics they employ, as well as differing in the degrees of vowel height, tone, and stress they allow. Therefore, the set of features must be allowed to vary from language to language, and in principle is limited only by the number of rules in the phonology; the set of specifications must likewise be allowed to vary from language to language.

Yet there is an important distinction to be made between the diacritic features and the phonological features. Diacritics are properties of matrices, while phonological features are properties of units. Diacritics are used only to control the derivation—they are never affected by it. No rule ever rewrites the value of a diacritic. So even though the phonology must have access to an unbounded number of diacritics, it cannot perform unbounded computations on them.

That leaves us with the phonological features, which we can further sepa-

---

nondeterministically guess an arbitrarily long underlying form x with which to generate a given surface form $g(x)$ (in the context of the PRP only). We can harness this power if we can encode the entire TM computation in a single segmental string. Complex rules will ensure that a given underlying form describes a legal computation string. Let units encode tape squares, as in the proof of lemma 2.1.1, requiring a fixed number of features. (If, instead, we let each unit encode a complete instantaneous description, as in the proof of lemma 2.1.2, then we will have proved that the PRP is EXPPOLY time hard, using an unbounded number of features.) As before, the dictionary generates all strings of units, corresponding to all possible computations. The segmental grammar consists of three stages. In the first stage, optional rules nondeterministically specify the units; in the second stage adjacent units are checked to ensure that they obey the next-move relation of the TM, and if they don't they are marked as illegal; in the third stage, the computation string is reduced to a single unit, which is either marked as illegal or as representing a halted accepting state of the machine. The reduction is considerably simpler using the nonlocal Post rewriting specified by complex rules.

rate into two classes: articulatory and suprasegmental. The suprasegmental features, such as degree of stress, is the topic of the next section. So let us examine the articulatory features here.

There are a fixed number of articulatory features, on the order of 10 to 15, determined by the musculature of the vocal apparatus. It is also true that there is an upper limit to the number N of perceivable distinctions that any one phonetic feature is capable of supporting. Therefore, it is an empirical fact that the number of possible phonetic segments is fixed in advance for all human languages. And from a computer science perspective, this number is small, between a thousand and a million.

If we take these empirical upper bounds seriously, then the PSPACE simulation described in lemma 2.1.2 would no longer be allowed, because it requires an unbounded number of features. Although the undecidable TM simulation in lemma 2.1.1 would not be affected by this constraint, because it only requires a (very small) fixed number of features, this simulation is independently excluded by the proposed limit on the number of embeddings. However, using the ideas in footnote 8, we would still be able to prove that the PRP is undecidable.

On the one hand, this fact provides a convenient constraint on the complexity of the PGP, that can be defended solely on empirical grounds. On the other hand, this fact does not lead to any insight or understanding, and it is difficult to defend such a trivial constraint on scientific or mathematical grounds.

To fix the number of segments complicates the linguistic theory needlessly, requiring an extra statement whose only consequence is to block a certain class of complexity proofs, and inhibit the search for more significant constraints. It has no other consequences, and if a new feature was discovered, the bound would only be revised upwards. Therefore, the proper scientific idealization is that the phonology has an arbitrary articulatory base (as seen in sign languages).

The purpose of this report is to illuminate the computational properties of human language, as a function of the natural parameters of variation, using the tools of computer science. We know that the number of phonological distinctions (that is, features and specifications) is a natural parameter of the phonology, because it varies from language to language. We also know that this parameter affects both the computational and informational complexity of the phonology, and for that reason it must be included in an honest mathematical analysis. To exclude it from consideration is only to blind ourselves to the computational structure of the system. Therefore, the idealization to an unbounded number of features is necessary on purely mathematical grounds.

Such a fixed bound is also at odds with a fundamental insight of the SPE system. A central component of SPE is the complex rule formalism, which characterizes the class of linguistically significant generalizations. The central difference between complex and elementary rules is that complex rules naturally describe nonlocal dependencies, whereas elementary rules are limited to describing local dependencies. The central prediction of an evaluation metric defined on the complex rules, then, is that nonlocal phonological dependencies are as natural as local dependencies. In other words, the class of phonological dependencies cannot naturally be encoded with a finite-state automaton. Yet, when we fix the length of derivations and maximal number of features, we limit the phonology to only describing finite-state dependencies.

I return to consider the nature of unbounded idealizations in a more general setting in appendix A.2.

To my mind, the most telling reason not to fix the number of articulatory features is that it is a jejune constraint that does not itself lead to any understanding. Even worse, it distracts us from the truly interesting research questions, such as whether features do in fact correspond to reusable computational space in the phonology, as we have used them in the reductions. When we block the reductions by fixing the number of features, we do not answer this question. All we do is make the question irrelevant, because to fix the features is simply to ignore them. So, for the purpose of increasing our understanding of human language, let us keep our idealization to an unbounded number of features, and proceed with the investigation.

Restricting the rewriting rules

In order to bound the time resources available to the phonology, we have considered limiting the number of derivational cycles directly, and indirectly, by restricting the class of readjustment rules. We have also examined ways to bound the space resources available to the phonology, by limiting the number of phonological features. Although these purely formal restrictions block a class of reductions, and constrain the class of phonological computations (and thereby the class of characterizable phonological dependencies), neither suffices to eliminate the intractability that is inherent in an unrestricted rewriting system. This raises the question, how might we restrict the rules themselves?

Elementary rules are used in at least six ways: (i) to convert binary phonological features to n-ary phonetic features, for example, the `nasal` feature in SPE; (ii) to make a unit agree or disagree with the features of an adjacent unit, that is, to represent assimilation and dissimilation processes; (iii) to insert units that are entirely predictable, as in English epenthetic vowels; (iv) to delete units that violate well-formedness conditions on representations; (v) to swap two adjacent units, which is called metathesis; and (vi) to derive irregular surface forms, as in the English '/*go*/ + <`past`> → [*went*]'.

Abstract words are rewritten as segmental strings in the interface between syntax and phonology (see chapter 3). The derivation of irregular and regular forms are identical from this perspective: both are simply the arbitrary rewriting of abstract morphological constituents into segmental strings. The first restriction on elementary rules, then, is to limit the class of arbitrary rewritings to the interface between phonology and morphology, and to ban the arbitrary rewriting of segmental strings from the phonology proper.

Rules that delete, change, exchange, or insert segments—as well as rules that manipulate boundaries—are crucial to phonological theorizing, and therefore cannot be crudely constrained.[9] More subtle and indirect restrictions are needed for these rules.[10]

[9]One restriction proposed in the literature, is McCarthy's (1981:405) "morpheme rule constraint" (MRC), which requires all morphological rules to be of the form $A \to B/X$ where A is a unit or ϕ, and B and X are (possibly null) strings of units. (X is the immediate context of A, to the right or left.) The MRC does not constrain the computational complexity of segmental phonology, because individual rules can still insert and delete segments, and groups of rules can be coordinated to perform arbitrary rewriting.

[10]That Chomsky and Halle were well aware of these problems is beyond doubt: "A possible direction in which one might look for such an extension of the theory is suggested by certain other facts that are not handled with complete adequacy in the present theory. Consider first the manner in which the process of metathesis was treated in Chapter Eight, Section 5. As will be recalled, we were forced there to take advantage of powerful transformational machinery of the sort that is used in the syntax. This increase in the power of the formal devices of phonology did not seem fully justified since it was made only to handle a marginal type of phenomenon. An alternative way to achieve the same results is to introduce a special device which would be interpreted by the conventions on rule application as having the effect of permuting the sequential order of a pair of segments."

One indirect restriction is to limit the possible interactions among rules. Because segmental grammars do not have a finite state control, all rule interactions must arise via the derivation form (that is, the sequence of segmental strings that is the computation string for the segmental derivation). The computationally significant interactions are ones that use the derivation form to store intermediate results of computations. The segmental model allows one rule to make a change in the derivation form, and a subsequent rule to make a change to this change, and so on. A segment that is inserted can subsequently be deleted; a segment that is switched with another segment can subsequently be switched with another segment, or deleted.

We have every reason to believe that such interactions are not natural. The underlying form of a word must encode all the information needed to pronounce that word, as well as recognize it. This information must be readily accessible, in order to ease the task of speaking, as well as that of acquiring the underlying forms of new words. The underlying form of a given word is that representation that omits all the directly predictable information in the surface form. The methodological directive "omit predictable information" means that a feature or segment of a surface form must be omitted if it is directly predictable from the properties of the phonology as a whole (such as the structure of articulations or the segmental inventory), or from the properties of that particular surface form, such as its morpheme class, adjacent segments, or suprasegmental patterns. To a first approximation, "directly predictable" means "computable by one rule with unbounded context and no intermediate results."

In point of fact, insertions and deletions do not interact in the systems proposed by phonologists. Units are inserted only when they appear in the surface form, and are totally predictable. Such units are never deleted. Because inserted units aren't deleted, and because an underlying form is proportional to the size of its surface form, the derivation can only perform a limited number of deletions, bounded by the size of the underlying form. In general, deletions typically only occur at boundaries, in order to "fix-up" the boundary between two morphemes. Because underlying forms cannot consist solely of boundaries, we would expect the size of an underlying form to be proportional to the size of its surface realization.

The immediate consequence of this "direct prediction" property of segmental rules is that underlying forms cannot contain significantly more segments

(p.427)

or features than their corresponding surface forms. It is also true that the derivation adds predictable information to the underlying form in a nearly monotonic fashion. The next restriction is to severely limit rule interactions in the segmental model: to exclude the storing of intermediate results in derivation forms, to require all derivation paths to accept and to be nearly monotonic. Deletion phenomenon might be modeled using a diacritic that blocks the insertion of the "deleted" segment. The details of such a (nearly or strictly) monotonic model must of course be worked out. But it is promising, and if plausible, as it seems to be, then the simulations in footnote 8 would be excluded. This is one formal way to define the notion of "predictable information," which, based as it is in the fundamental notion of computationally accessible information, seems more coherent and fundamental than the notion of a "linguistically-significant generalization," which has proven elusive.

2.1.3 The SPE evaluation metric

The SPE evaluation metric is a proposal to define the notion of a natural rule and linguistically-significant generalization. At first glance, this proposal seems vacuous. In order to minimize the number of symbols in the grammar, observed surface forms should simply be stored in the dictionary of underlying forms. Then the number of symbols in the grammar is zero, and all the linguistically significant generalizations in the corpus have been discovered, that is, none. Clearly, this is not what Chomsky and Halle intended.

Perhaps the size of the dictionary must be included in the metric as well. Now the most natural phonology is the smallest grammar–dictionary whose output is consistent with the observed corpus. The solution to this problem is also trivial: the optimal grammar–dictionary simply generates Σ^*.

So the simplest coherent revision of the SPE metric states the most natural phonological system is the smallest grammar–dictionary that generates exactly the finite set of observed forms.[11] Ignoring questions of feasibility (that

[11] In this we adopt a standard assumption of the field, that the language acquisition device does not have access to negative evidence. If negative evidence were allowed, then the SPE metric would be revised to state that the most natural system is the smallest grammar-dictionary consistent with the evidence, that is, accepts all positive examples and rejects all negative examples. This approach is promising under the weak definition of "negative evidence" as "absence of confirmation."

is, how to algorithmically find such a system), we run into serious empirical problems because the observed corpus is always finite. The smallest grammar will always take advantage of this finiteness, by discovering patterns not yet falsified by the set of observed surface forms. The underlying forms in such an optimal grammar–dictionary system will in fact look nothing like the true underlying forms, that is, those postulated by phonologists on the basis of scientific evidence that is not available to the language acquisition device (LAD). And even if the set of underlying forms is fixed, the optimal grammar in such a system will still not be natural, failing standard empirical tests, such as those posed by loan words and language change.[12]

This observation is confirmed by the complexity proofs. An important corollary to lemma 2.1.1 is that segmental grammars form a universal basis for computation. It is possible to simulate an arbitrary Post tag system using a very simple set of phonological rules. Or we can simulate the four-symbol seven-state "smallest universal Turing machine" of Minsky (1967) in the segmental model; the resulting grammar contains no more than three features, eight specifications, and 36 trivial rules. These segmental grammars of universal computation contain significantly fewer symbols than a segmental grammar for any natural language. And this is not even the best that can be done. The smallest combined grammar-dictionary for the set of all observed words will be even smaller, because it can take advantage of all computable generalizations among the finite set of observed surface forms, not only the linguistically significant ones. In fact, the attached dictionary would represent the Kolmogorov complexity of the observed surface forms with respect to the optimal segmental grammar, that is, the true information content of the observed surface forms with respect to an arbitrarily powerful encoder. Therefore, this corollary presents severe conceptual and empirical problems for the segmental theory.

In short, even if we ignore questions of feasibility, the smallest segmental grammar–dictionary capable of enumerating the set of observed surface forms cannot be natural because it must discover too many unnatural gen-

[12]In my brief experience as a phonologist, the most natural grammars did not have the smallest number of symbols, even when the proper morphemic decomposition of underlying forms was known in advance. With enough time and mental discipline, it was always possible to construct a smaller grammar than the "correct" one, by taking advantage of "unnatural" patterns in the observed surface forms. Increasing the number of examples does not help, simply because there will never be enough examples to exclude all the computable but unnatural patterns.

eralizations.

How then can we make sense of the SPE evaluation metric? The evaluation metric makes certain sets of disjunctively ordered elementary rules as natural as an elementary rule. The fundamental difference between a complex rule and an elementary rule is that a complex rule is capable of performing nonlocal Post-style rewriting, whereas elementary rules are limited to local Thue-style rewriting. Therefore, the SPE evaluation metric formalizes the observation that nonlocal phonological dependencies can be as natural as local ones. The only difficulty is, the relatively subtle distinction between local and nonlocal rewriting is overwhelmed by the brute power of an unrestricted rewriting system to encode arbitrary r.e. dependencies.

This observation suggests a solution, and a promising line of investigation.

The early evaluation metrics included not only a measure of the number of symbols in the grammar (Kolmogorov complexity), but also a measure of the length of derivations (time complexity). So, in *Morphophonemics of Modern Hebrew*, Chomsky (1951) proposed a mixed evaluation metric:

> "Given the fixed notation, the criteria of simplicity governing the ordering of statements are as follows: that the shorter grammar is the simpler, and that among equally short grammars, the simplest is that in which the average length of derivation of sentences is least." (p.6)
>
> "The criteria for justification of ordering are as given at the conclusion of section 0: simplicity is increased by 1. reduction of the number of symbols in a statement (paired brackets, etc., counting as one symbol); 2. reduction of the length of derivations. with the second requirement subsidiary. Actually, it applies only once, and then in a trivial fashion. I mention it only to indicate explicitly that this consideration, taken as subsidiary, will not materially increase the ordering restrictions." (pp.51–2)

A similar proposal appears in SPE, in the context of how the conventions of conjunctive and disjunctive ordering should be applied:

> A natural principle that suggests itself at once is this: *abbreviatory notations must be selected in such a way as to maximize disjunctive ordering.* .... The principle seems to us a natural one in that maximization of disjunctive ordering will, in general,

> minimize the length of derivations in the grammar. The question of how an internalized grammar is used in performance (speech production or perception) is of course quite open. Nevertheless, it seems reasonable to suppose that the grammar should be selected in a such a way as to minimize the amount of 'computation' that is necessary, and that 'length of derivation' is one factor in determining 'complexity of computation'. Naturally, this principle must be regarded as quite tentative. We will adhere to it where a choice arises, but we have very little evidence for or against it. To find empirical evidence bearing on a principle of this degree of abstractness is not an easy matter, but the issue is important, and one should bear it in mind in a detailed investigation of phonological structure. (SPE, p.63).

As stated the addition of complexity concerns does not make a difference, because the derivation-length is strictly ordered within the symbol-count. The idea of combining Kolmogorov and computational complexities is attractive, however, for all the reasons mentioned. Let us therefore replace the SPE/Morphophonemics metric with a metric inspired by the time-bounded Kolmogorov complexity of Levin (1973). Let $\mathsf{time}(g, u)$ be the number of steps taken by grammar g to produce surface form $g(u)$ on input underlying form u. Then the complexity $\mathcal{K}(g, D)$ of a phonological grammar g and dictionary of underlying forms D is:

$$\mathcal{K}(g, D) = |g| + \sum_{u \in D} |u| + \mathsf{time}(g, u)$$

A grammar-dictionary (g_1, D_1) is preferred over another grammar-dictionary (g_2, D_2) when (i) it is less complex ($\mathcal{K}(g_1, D_1) < \mathcal{K}(g_2, D_2)$); and (ii) both are extensionally equivalent ($\{g_1(u) : u \in D_1\} = \{g_2(u) : u \in D_2\}$).

Now it is possible, at least in principle, to find the optimal grammar-dictionary grammar, by a simple sequential search. And the phonology is only able to discover efficiently computable and simple patterns, a potential improvement over the SPE proposal.

Although this evaluation metric solves the technical problems of earlier proposals, it is not clear that it would result in natural grammars. In my opinion, the notion of a "linguistically significant generalization" is best formalized by postulating a weak encoder, that can *only* discover linguistically significant generalizations. The evaluation metric, rather than the symbol count of the grammar, is the minimum description length criterion applied to the set of observed surface forms, which are encoded relative to the model class of restricted direct-prediction phonologies (cf., Rissanan 1978). That way, in order to minimize the encoding of the set of observed surface forms, the grammar *must* discover the linguistically significant generalizations after having seen a sufficiently large (finite) set of surface forms. The grammar is also forced to describe phrase-level phonological processes, because no finite word/morpheme dictionary is capable of doing so. The research challenge, then, is to design such a probabilistic model class for phonology. The modern autosegmental model seems to be a good place to start, because suprasegmental processes are a large class of the phonologically significant generalizations.

2.1.4 Modeling phonological dependencies

In SPE, knowledge of phonological dependencies is modeled with an unrestricted Post-style rewriting system. Such a system is capable of encoding arbitrary r.e. dependencies in the derivation of surface forms from underlying forms. It forms a universal basis for computable knowledge.

We know that phonological dependencies can be complex, exceeding the capacity of a finite-state encoder. However, not every segmental grammar generates a natural set of sound patterns. So why should we have any faith or interest in the formal system? The only justification for these formal systems then would seem to be that they are good programming languages for phonological processes, that clearly capture our intuitions about human phonology. But segmental theories are not always appropriate. Their notation is constrained, which limits their expressive power. Interactions among phonological processes are hidden in rule ordering, disjunctive ordering, blocks, and cyclicity. Yet, despite these attempts to formalize the notion of a natural phonological dependency, it is possible to write a segmental grammar for any recursively enumerable set.

Natural phonological processes seem to avoid complexity and simplify interactions. It is hard to find a phonological constraint that is absolute and inviolable. There are always exceptions, exceptions to the exceptions, and so forth. Deletion processes like apocope, syncopy, cluster simplification and stray erasure, as well as insertions, seem to be motivated by the necessity of modifying a representation to satisfy a phonological constraint, not to exclude representations, to hide vast computations, or to generate complex sets, as we have used them here.

The next step in the research program initiated here is to design an appropriate formal phonological model, along the lines discussed above, in order to answer the fundamental questions of naturalness, appropriate generalization, and what seems to be the lynchpin of the phonology, the omission of directly predictable information. It is also now possible to discuss the notion of computationally accessible information, which has played such an important role in modern cryptography, and to consider a more sophisticated—yet still constructive—complexity thesis for human language, based on the fundamental ideas of entropy and computation. We might hypothesize that knowledge of phonology, and of all linguistic dependencies, is computationally accessible in the sense of Yao (1988).

2.2 Autosegmental phonology

In the past decade, generative phonology has seen a revolution in the linguistic treatment of suprasegmental processes such as tone, harmony, infixation/interleaving, and stress assignment. Although these autosegmental models have yet to be formalized, they may be briefly described as follows. Rather than one-dimensional strings of segments, representations may be thought of as "a three-dimensional object that for concreteness one might picture as a spiral-bound notebook," whose spine is the segmental string and whose pages contain simple constituent structures that are independent of the spine (Halle 1985). One page represents the sequence of tones associated with a given articulation. By decoupling the representation of tonal sequences from the articulation sequence, it is possible for segmental sequences of different lengths to nonetheless be associated to the same tone sequence. For example, the tonal sequence Low-High-High, which is used by English speakers to express surprise when answering a question, might be associated to a word containing any number of syllables, from two (*Brazil*) to twelve (*floccinaucinihilipilification*) and beyond. Other pages (called "planes") represent morphemes, syllable structure, vowels and consonants, and the tree of articulatory (that is, phonetic) features.

2.2.1 Complexity of autosegmental recognition

Now we prove that the PRP for autosegmental models is NP-hard, a significant reduction in complexity from the undecidable and PSPACE-hard computations of segmental theories. (Note however that autosegmental representations have augmented—but not replaced—portions of the segmental model, and therefore, unless something can be done to simplify segmental derivations, modern phonology inherits the intractability of purely segmental approaches.)

Let us begin by thinking of the NP-complete 3-Satisfiability problem (3SAT) as a set of interacting constraints. In particular, every satisfiable Boolean formula in 3-CNF is a string of clauses $C_1, C_2, \ldots, C_p$ in the variables $x_1, x_2, \ldots, x_n$ that satisfies the following three constraints: (i) negation: a variable x_j and its negation $\overline{x}_j$ have opposite truth values; (ii) clausal satisfaction: every clause $C_i = (a_i \vee b_i \vee c_i)$ contains a true literal (a literal is a variable or its negation); (iii) consistency of truth assignments: every positive literal of a given variable is assigned the same truth value, either 1 or 0.

Lemma 2.2.1 *Autosegmental representations can enforce the 3SAT constraints.*

Proof. The idea of the proof is to encode negation and the truth values of variables in features; to enforce clausal satisfaction with a local autosegmental process, such as syllable structure; and to ensure consistency of truth assignments with a nonlocal autosegmental process, such as a nonconcatenative morphology or long-distance assimilation (harmony). To implement these ideas we must examine morphology, harmony, and syllable structure.

Morpheme Interleaving. In the more familiar languages of the world, such as Romance languages, words are formed primarily by the concatenation of morphemes. In other languages, such as the Semitic languages, words are formed by interleaving the segments of different morphemes. For example, the Classical Arabic word *kataba*, meaning 'he wrote', is formed by interleaving (with repetition) the 1 segment of the active perfective morpheme *a* with the 3 segments of the *ktb* morpheme (cf., McCarthy 1981). (Constraints on syllable structure, discussed below, explain why the 1 underlying vocalic segment /a/ appears 3 times in the surface form.) In the autosegmental model, each morpheme is assigned its own plane. We can use this system of representation to ensure consistency of truth assigments. Each Boolean variable x_i is represented by a separate morpheme μ_i, and every literal of x_i in the string of formula literals is associated to the one underlying morpheme μ_i.

Harmony. Assimilation is the common phonological process whereby some segment comes to share properties of an adjacent segment. In English, consonant nasality assimilates to immediately preceding vowels. Assimilation also occurs across morpheme boundaries, as the varied surface forms of the prefix *in-* demonstrate: *in+tolerable* → *intolerable*, but *in+logical* → *illogical* and *in+probable* → *improbable*. In other languages, assimilation is unbounded and can affect nonadjacent segments: these assimilation processes are called harmony systems. In the Turkic languages all suffix vowels assimilate the backnesss feature of the last stem vowel; in Capanahua, vowels and glides that precede a word-final deleted nasal (an underlying nasal segment absent from the surface form) are all nasalized. In the autosegmental model, each harmonic feature is assigned its own plane. As with morpheme-interleaving, we can represent each Boolean variable by a harmonic feature, and thereby ensure consistency of truth assignments.

Syllable structure. Words are partitioned into syllables. Syllables are the fundamental unit of segmental organization (Clements and Keyser, 1983). Each syllable contains one or more vowels 'V' (its nucleus) that may be preceded or followed by consonants 'C'. For example, the Arabic word *ka.ta.ba* consists of three two-segment syllables, each of the form CV. Every segment is assigned a sonority value, which (intuitively) is proportional to the openness of the vocal cavity. For example, vowels are the most sonorous segments, while stops such as /p/ or /b/ are the least sonorous. Syllables obey a language-universal sonority sequencing constraint (SSC), which states that the nucleus is the sonority peak of a syllable, and that the sonority of adjacent segments swiftly and monotonically decreases. We can use the SSC to ensure that every clause C_i contains a true literal as follows. The central idea is to make literal truth correspond to the stricture feature, so that a true literal (represented as a vowel) is more sonorous than a false literal (represented as a consonant). Each clause $C_i = (a_i \vee b_i \vee c_i)$ is encoded as a segmental string $x - x_a - x_b - x_c$, where x is a consonant of sonority 1. Segment x_a has sonority 10 when literal a_i is true, 2 otherwise; segment x_b has sonority 9 when literal b_i is true, 5 otherwise; and segment x_c has sonority 8 when literal c_i is true, 2 otherwise. Of the eight possible truth values of the three literals and the corresponding syllabifications, only the syllabification corresponding to three false literals is excluded by the SSC.

In that case, the corresponding string of four consonants C-C-C-C has the sonority sequence 1-2-5-2. No immediately preceding or following segment of any sonority can result in a syllabification that obeys the SSC. Therefore, all Boolean clauses must contain a true literal. $\square$

The only fact needed to obtain an NP-hardness result from this lemma 2.2.1 is the fundamentally elliptical nature of speech, as described by Jakobson and Halle (1956):

> Usually ... the context and the situation permit us to disregard a high percentage of the features, phonemes and sequences in the incoming message without jeopardizing its comprehension. The probability of occurrence in the spoken chain varies for different features and likewise for each feature in different texts. For this reason it is possible, from a part of the sequence, to predict with greater or lesser accuracy the succeeding features, to reconstruct the preceding ones, and finally to infer from some features in a bundle the other concurrent features.
>
> Since in various circumstances the distinctive load of the phonemes is actually reduced for the listener, the speaker, in his turn is relieved of executing all the sound distinctions in his message: the number of effaced features, omitted phonemes and simplified sequences may be considerable in a blurred and rapid style of speaking. The sound shape of speech may be no less elliptic than its syntactic composition. Even such specimens as the slovenly /tem mins sem/ for 'ten minutes to seven', quoted by D. Jones, are not the highest degree of omission and fragmentariness encountered in familiar talk. (pp.5–6)

The direct consequence of lemma 2.2.1, and the fact that not all sound distinctions are executed, and those that are may be corrupted, is:

Theorem 5 *PRP for the autosegmental model is NP-hard.*

Proof. By reduction to 3SAT. The idea is to construct a surface form that completely identifies the variables and their negation or lack of it, but does not specify the truth values of those variables. That is, the stricture feature has been ellipsed. The dictionary will generate all possible underlying forms (interleaved morphemes or harmonic strings), one for each possible truth assignment, and the autosegmental representation of lemma 2.2.1 will ensure that generated formulas are in fact satisfiable. □

2.2.2 Suprasegmental dependencies

It is informative to reexamine these suprasegmental processes, from an information-theoretic perspective. The relationship between the sound of a word and its meaning is inherently arbitrary. A given sequence of articulations could in principle mean anything; a given meaning could in principle have any articulation. And it seems that the storage capacity of the human brain has, for all practical purposes, no limit (cf., Luria 1968).

Yet there appear to be two primary sources of phonological systematicity, that is, of sound patterns both among and inside surface forms.

The phonetic systematicity in the mental lexicon arises from the fact that words consist of morpheme combinations. Although words that share morphemes need not in principle share phonological patterns, most in fact do. This makes it easier to acquire words, and to invent them, because the meaning of a word is given by its relation to other words, as well as by its intrinsic content. A regular mapping from morphology to phonology simplifies the acquisition and invention of new words.

The phonetic systematicity in a surface form is due to suprasegmental processes. A suprasegmental process p establishes the domains within which one element in each domain, the head, is distinguished phonetically from the other nonhead elements in that domain. There are three parameters of variation:[13]

1. The phonological representation r on which p operates, including syllables and any node in the tree of articulatory features. (The elements of r are called the "p-bearing units.")

2. A restricted class of trees, whose leaves are attached to the p-bearing units. Each nonterminal in such a tree immediately dominates one head and a (possibly empty) set of nonheads, thereby representing a suprasegmental domain. The domains defined by these trees are contiguous and exhaustive.

[13]This incomplete proposal is inspired by the Halle and Vergnaud (1987) treatment of phonological stress, and by conversations with Morris Halle.

3. The entirely local process that realizes the abstract distinction between heads and nonheads in the phonetics.

Suprasegmental processes maintain systematic distinctions between adjacent segments of each surface form, as well as ensuring that segmental strings have global properties, and thereby contribute to efficient production and error-correcting comprehension (cf. Jakobson and Halle, 1956).

Syllables organize the phonological segments of a given language. A string of segments is a possible sound if and only if it can be partitioned into a sequence of substrings, each of which corresponds to a permissible syllable. Syllables represent, in part, the sonority hierarchy between the nucleus vowel and its consonental onset and coda. Syllabic domains are given by the language-universal sonority hierarchy as well as by language-particular constraints, that may be represented with a small set of syllable templates, such as 'CV' and 'CVC'.

```
1.Segments are the syllable-bearing units.
   a.local sonority maxima are inherent heads.
2.Compute syllabic domains.
3.Perform segmental adjustments.
   a.insert or delete units to satisfy constraints.
```

The stress-bearing units of a language are either entire syllables, or their moras.[14] Stress domains are defined by a class of finite-depth trees, each of whose levels are characterized by four language-universal parameters: feet are bounded or unbounded; the head of a foot is left-terminal, medial, or right-terminal; and feet are constructed left-to-right or right-to-left (Halle and Vergnaud, 1987). For example, the English suprasegmental stress process would be described as:

```
1.Syllable heads are the stress-bearing units.
   a.unmarked: heavy ultima is inherent head.
   b.marked: heavy syllable is inherent head.
2.Compute stress domains.
3.Perform phonetic adjustments.
```

[14] Moras are the units of syllable weight: a heavy syllable has more moras than a light syllable. Technically, a mora is a segment dominated by the syllable nucleus, Clements and Keyser (1983). .

```
   a.shorten head of open word-medial nonhead syllables.
   b.destress head adjacent to more prominent head.
   c.reduce stressless vowels to shwa.
```

The assimilation-bearing units of a language are those nodes of the articulatory tree that interact with assimilation processes, including harmonic and blocking features. Long-distance harmony corresponds to an unbounded domain of assimilation. A prototypical assimilation process might look like:

```
1.Articulatory node n is the assimilation-bearing unit.
   a.assimilation and blocking features are inherent heads.
2.Compute assimilation domains.
3.Perform phonetic adjustments.
   a.spread node n from head segment to its domain.
```

The remaining suprasegmental processes are also straightforward. For example, the melody plane represents the segmental domain of equi-tonality. This is, of course, an informal proposal intended to illuminate the relationship between the autosegmental processes and the segmental string, and to suggest a formalization of autosegmental phonology.

Chapter 3

Syntactic Agreement and Lexical Ambiguity

In this chapter, we consider the computational process of computing the language user's knowledge of syntax. Knowledge of syntax includes knowledge of syntactic dependencies, such as agreement or selection, and knowledge of purely syntactic distinctions, such as noun/verb or singular/plural. Syntactic dependencies are defined with respect to the syntactic distinctions. That is, we say "a verb selects a noun phrase," or "the noun and verb agree on number." An adequate linguistic theory must represent these dependencies and list the set of possible syntactic distinctions.

By any account, syntactic dependencies are complex, involving the interaction of local and nonlocal relations. Seemingly local decisions, such as the disambiguation of a particular word, can have global consequences. This suggests that it may be difficult to assign a structural description to a sequence of ambiguous words. In order to translate this informal observation into a formal proof, we must define the problem of assigning a structural description to a string of words.

We immediately encounter two difficulties. The first is that no one understands what it means to successfully comprehend an utterance. As discussed in the introduction, it cannot mean to find exactly the structural description in the head of the speaker, because this may not be possible. Nor can it mean to find *some* structural description for the utterance, because this is the trivial language *mis*comprehension problem. In short, it is not possible to define the LC problem for syntax without a simple characterization of the class of appropriate structural descriptions. In order to overcome this obstacle, we must define our problem so that it is a natural subproblem of any reasonable statement of the language comprehension problem.

The second difficulty is that, unlike the anaphora problem considered in chapter 4, the class of structural descriptions for syntax does not have a simple theory-invariant (that is, *direct*) characterization. There are a wide range of competing syntactic theories, and they differ significantly. It is possible to broadly distinguish two classes of syntactic theories.

- In unification-based theories, such as lexical-functional grammar or generalized phrase structure grammar, atomic features represent possible distinctions. Syntactic dependencies are all stated in terms of one simple mechanism: the unification of uniform sets of features between a phrase structure node and its immediate ancestor, children, or siblings. For example, subject–verb agreement is implemented by a chain of local unification: the subject noun is unified with the subject NP (its ancestor); the subject, with the matrix VP (its sibling); and the matrix VP with the main verb (its child).

- In current transformational theories, possible distinctions are represented by features and morphemes. Syntactic dependencies consist of particular linguistic relations, such as predication, selection, and theta-role assignment. They are defined primarily in terms of local phrase structure configurations at different levels of representation. The mapping between different levels of representations is performed by transformations. For example, subject–verb agreement results because the subject specifies the features of an agreement morpheme; this morpheme is subsequently combined with the verb root morpheme at a different level of representation.

In order to overcome this obstacle of competing theories and achieve an invariant analysis, we must first define the language comprehension problem for syntax relative to the linguistic theory, and then analyze its complexity for both classes of linguistic theories.

Our subproblem of choice is the *lexical resolution problem* (LRP) for a given syntactic theory: Given a partial syntactic representation R that yields a string of ambiguous or underspecified words, and a lexicon L containing ambiguous words, can the words in R be found in the lexicon L? This problem statement overcomes the two difficulties. It is defined relative to the syntactic theory, and the language user must solve the LRP in order to find an appropriate structural description.

Unification-based theories are very similar from a formal perspective. And because all syntactic dependencies are stated in terms of a simple, uniform mechanism (feature unification), it has been straightforward to prove that the LRP for these theories is NP-hard (Ristad and Berwick, 1989). In this chapter, we prove that the LRP for modern transformational theories is also NP-hard. Both proofs rely on the particular details of some linguistic theory. But the combined effect of these two results for the LRP is to argue for the NP-hardness of the "true" language comprehension problem for syntax.

The chapter is organized into four sections. The first section introduces the structural descriptions of current transformational theories, with motivation. In section 3.2, we prove that the LRP is NP-hard for these theories, under a very abstract formulation. Next, section 3.3 shows exactly how this abstract formulation applies to the "Barriers" theory of Chomsky (1986). The conclusion discusses the central role that locality plays in transformational theories, and the consequences of allowing uniform mechanisms in the linguistic theory.

3.1 Morpho-syntactic dependencies

The syntactic structure underlying even extremely simple constructions can be quite intricate. Consider the simple passive expression (1).

(1) John was seen.

What is it that English speakers know about this expression? For one, they know that the expression describes an event that occurred in the past. This information is contained in the verb *was*, which is overtly inflected for the past tense as well as overtly agreeing with the surface subject on number. One way to represent this knowledge is to say that the overt form *was* underlyingly consists of the three morphemes [be], [past], and [singular].

English speakers also know that the expression is in the passive voice—that the overt subject *John* is subjected to the "seeing action," and therefore

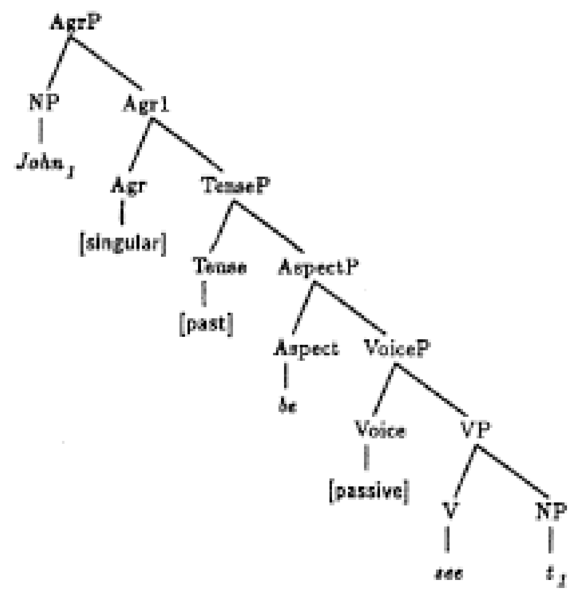


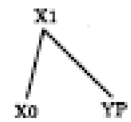
Figure 3.1: The underlying (partial) structural description of *John was seen*.

stands in the same relation to the verb *see* in (1) as it does in the corresponding active expression *Someone saw John*. That is, (i) the verbal form *seen* consists of the the verb root [see] and the voice morpheme [passive], which happens to be realized as the *+en* suffix here, and (ii) *John* is the underlying direct object of the verb *see* in both active and passive variants of the expression.

In order to represent this knowledge that language users in fact have about such expressions, we assign the partial structural description depicted in figure 3.1 to the utterance (1), where each surface word has been exploded into its underlying morphemes.

The morphemes are organized hierarchically according to X-bar theory. X-bar theory states that morphemes of type X project into syntactic constituents of type X. That is simply to say, for example, that a verb phrase must contain a verb and a noun phrase must contain a noun. The relation between a morpheme X and its complement (a phrase YP) is represented by the sisterhood configuration inside the first projection of X (2).

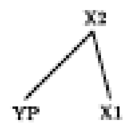
(2)



Selection is an instance of complementation. For example, the aspect morpheme [be] selects the voice morpheme [passive], which accounts for the deviancy of the minimally different expression *John was see*.

The relation between the first projection of X and its specifier (a phrase YP) is represented by sisterhood in the second projection of X (3).

(3)



The second projection X2 of the morpheme X is a phrase of type X, also written XP. Agreement is an instance of specification. For example, the proper noun *John* specifies the agreement morpheme [singular] in the underlying structural description in figure 3.1.

Finally, the underlying thematic relation between the verb *see* and *John* is represented indirectly, by postulating a trace t_i that is selected by the verb *see* (a trace is a phonologically silent place-holder) and is assigned the same index as element whose place it is holding (*John*$_i$).

In assigning this representation to the utterance (1), we were guided by the principle of universal explanation (UE). UE states that there is only one underlying language, from which particular languages differ in trivial ways.[1] One consequence of UE is that if any language makes an overt distinction, then all languages must make that distinction in their underlyingly representations. For example, in languages such as Hindi, verbs agree with their direct object and their subjects (Mahajan 1989); therefore object and subject must both appear in the specifier position of agreement phrases in all

[1]The principle of universal explanation is a particular theory of what constitutes universal grammar, that is, a theory of the innate endowment of the language user. It is fundamental to the study of language. For historical example, see the influential work of James Beattie (1788), especially his analysis of tense. And in his award-winning 1827 essay on the origin of language, Johann Gottfried von Herder says, "Who can (whatever he may have to say upon the subject) entirely deny the fundamental connection, existing between most languages? There is but one human race upon the earth, and but one language." (p.112) .

languages. A second consequence of UE is that all clauses have the same underlying structure. At the very least, a clause must contain a subject, a tense, and a verb. And because it contains a subject, it must also contain an agreement morpheme.

We were also guided by the goal of representing linguistic relations uniformly, via local phrase structure configurations. So the selection relation between a verb V and its underlying direct object XP is always represented by the complement configuration "[$_{V1}$ V XP]." And when the direct object appears as the surface subject, as in the passive, a trace is used as a place-holder.

Now consider the expression (4).

(4) Tom saw Mary yesterday.

If we examined certain cross-linguistic facts, and obeyed the principle of universal explanation, we would assign the underlying structural description in figure 3.2 to this expression.

The verb *see* selects the proper noun *Mary* as its direct object, resulting in a V1 projection. This V1 predicate is specified by its subject, the proper noun *Tom*. The relation of modification between the resulting VP and the adverb *yesterday* is represented configurationally as adjunction to VP. The remaining morphemes—object agreement, verbal tense, and subject agreement—appear in this structural description, but have not yet been specified. This is indicated by the empty categories "[*e*]" in their specifier positions. "C" is the complementizer morpheme, which is phonologically null in declarative expressions.

The underlying representation in figure 3.2 undergoes certain movement transformations, resulting in the surface form in figure 3.3.

First, the underlying object *Mary* moves to the specifier position of the object agreement phrase, so that the agreement morpheme will be specified [singular] and so that the object *Mary* will be assigned objective case. Next, *Tom*, the underlying subject of the verbal predicate [*see Mary*], moves to the specifier position of the subject agreement phrase, in order to specify the agreement morpheme as [singular] and be assigned nominative case. Finally, the verb *see* combines first with the object agreement morpheme, then with the tense morpheme, and finally with the subject agreement morpheme. It is spelled out as *saw*. Each movement transformation leaves behind an indexed trace.

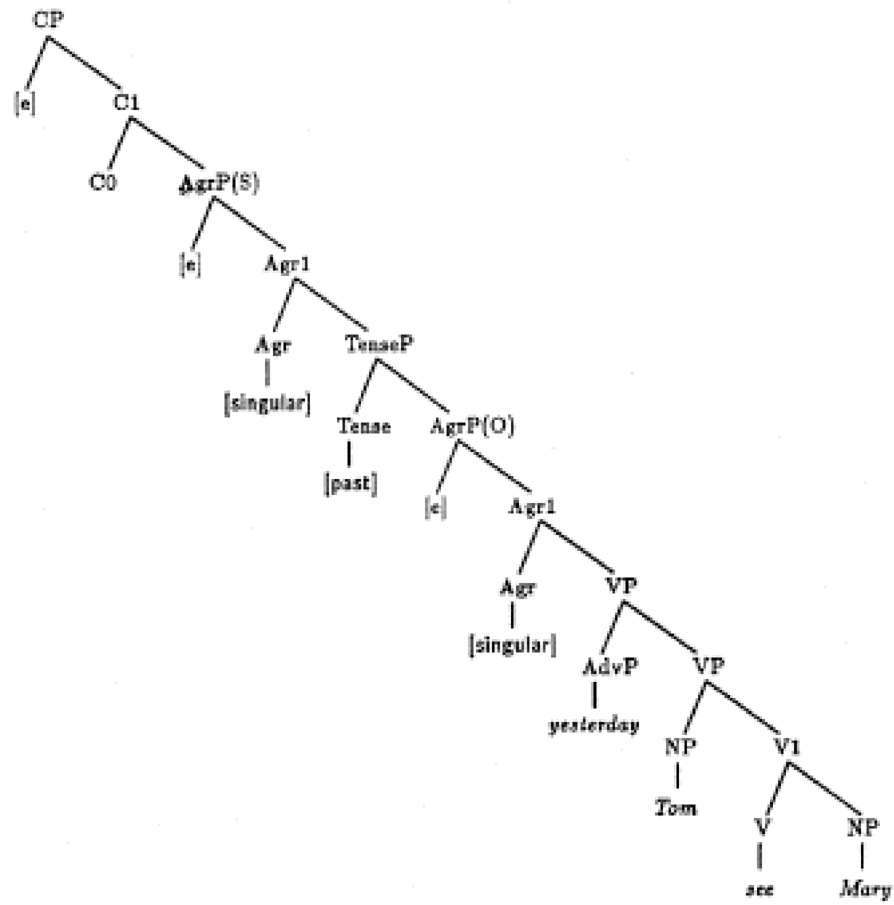


Figure 3.2: The underlying structural description of *Tom saw Mary yesterday*, which is a partial representation of the language user's knowledge of morphological dependencies.

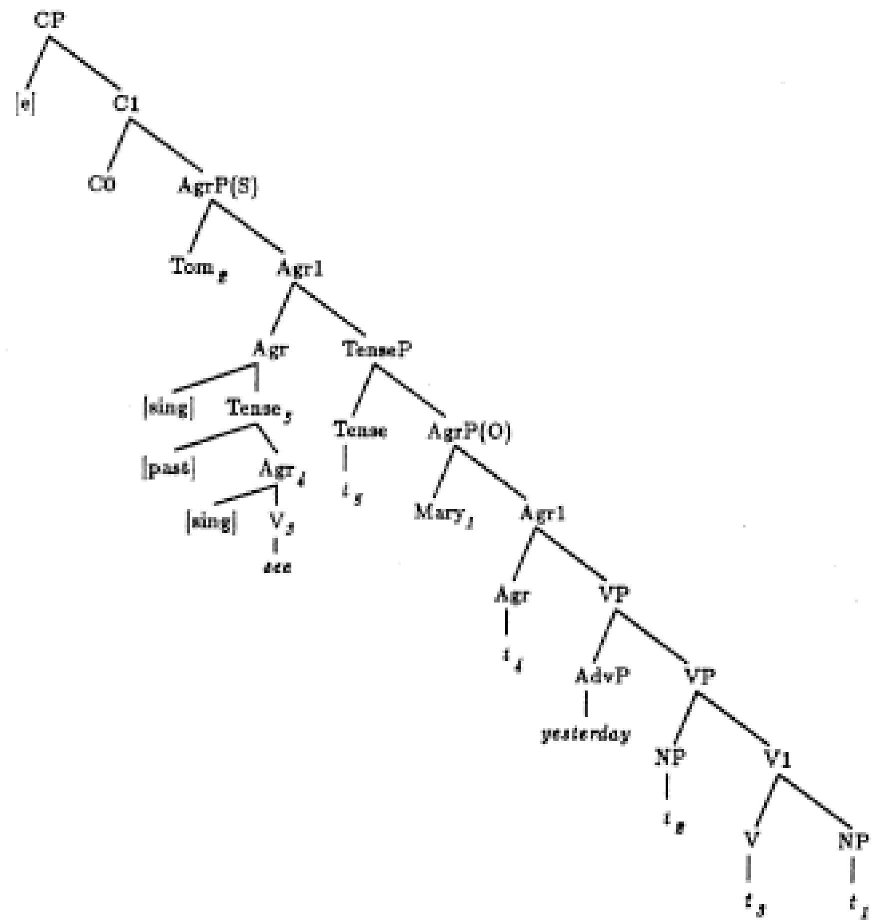


Figure 3.3: The surface form of *Tom saw Mary yesterday*, the result of repeatedly applying movement transformations to the underlying form in figure 3.2.

This analysis is motivated by Chomsky (1986;1989) and Pollack (1989). The movement transformations proposed in that work are considerably more complex than those shown here.

By explicitly representing the dependencies between the morphemes in this fashion, a number of things become clear. For one, each morpheme typically interacts with only a few other morphemes, such as its specifier and its complement. However, because each word consists of several morphemes, every word interacts with every other word in a clause.

Reconsider our first example (1). In that example, the passive verb form *see+en* selects the underlying object t_1 and assigns it a 'patient' thematic role. The underlying object t_1 appears as the surface subject *John*; the subject agrees with the inflected aspect *be+past+singular*, which assigns it nominative case; and, to complete the circle of interactions, the inflected aspect *was* selects the passive verb form. These properties of words, such as case-marking, thematic role assignment, selection and agreement, are all independent, not directly deducible from the phonological form of the words, and potentially correlated in the lexicon.

It is easy to see that interactions among the words in a sentence can become extremely complex. Imagine that the lexicon contained three homophonous verbs—see_1, see_2, and see_3—with the same phonological form but different selectional restrictions. Then verb phrases could encode satisfied 3-CNF clauses: see_1 would be false and select a true subject; see_2 would be false and select a true object; and see_3 would be true and select a subject and object of any truth value. The consequence is that any verb phrase headed by *see* must contain a word representing a true literal. We could even get two literals of the same variable to agree on their truth values by moving one to the subject position of the other, where they must agree, exactly as in the passive construction: the underlying object moves to the subject position, where it must agree with the auxiliary verb. Then if words were Boolean literals, it would be possible to encode 3SAT instances in sentences. The proof in the next section formalizes this intuitive argument.

So far our discussion of the language user's knowledge of syntax has concentrated on knowledge of syntactic dependencies. Let us therefore conclude this section with a brief discussion of the range of possible syntactic distinctions. There are two broad classes of distinctions with syntactic effects, the purely syntactic and the semantic or pragmatic.

Many semantic and pragmatic distinctions—such as animate/inanimate or abstract/concrete—have syntactic effects. As Chomsky (1968:75ff) has observed, this can be used to account for the deviancy of certain expressions. For example, the contrast between *sincerity may frighten the boy* and *the boy may frighten sincerity* is accounted for by the fact that *frighten* selects animate objects.

Purely syntactic distinctions are those distinctions that are independent of meaningful extra-linguistic distinctions.[2] For example, the syntactic distinction among masculine, feminine, and neuter does not correspond to biological sex; nor does singular/plural correspond to physical or perceptual indivisibility. Nor do nouns denote things, or verbs, actions, as has been very wittily argued by Thomas Gunter Browne in 1795. This class of purely syntactic distinctions includes morpheme class (noun, verb, adjective, and so forth), so-called agreement features (gender, number, person, kinship class, etc.), case, grammatical function, thematic role, and so on. These distinctions vary from language to language, and are all treated uniformly by the linguistic theory. Syntactic distinctions appear to originate from semantic distinctions, but soon lose their connection to meaning. If this is so, as linguists have argued it is, then we have good reason to believe that the number of syntactic distinctions is unbounded in principle, limited only by the amount of time it takes the language user to acquire a new distinction.

3.2 Complexity of linguistic transforms

In this section we prove that the LRP for modern transformational theories is NP-hard. The idea of the proof is quite similar to the proofs of lemma 2.2.1 and theorem 5 above. As in those proofs, we will construct a structural description that enforces the 3SAT constraints. Ambiguous words will play the role of elliptical speech. The words in the structural description will be ambiguous with respect to a syntactic distinction that corresponds to truth value.

Recall that transformation theories postulate at least two levels of syntactic representation—the underlying and surface representations—and a mapping from underlying form to surface form. The underlying form is called the D-

[2] The definition of what is and isn't a syntactic distinction is of course entirely theory-internal. However much as there is any agreement, syntax includes that portion of linguistic form that is logically independent of real-world meaning or phonology.

structure (DS) and the surface form is called the S-structure (SS). DS is a representation of thematic role assignment, selection, and grammatical function (subject, object, etc.). SS is a the syntactic representation closest to the surface form of a sentence. In current transformational theories, DS is mapped onto SS by the generalized move-α transformation.

The idea of the proof is to simulate the 3SAT constraints with a complex syntactic representation, that we will build using one simple building block, called a "stair."

Definition. A *stair* is an underlying form U_i with the following structure:

1. *Recursive structure.* U_i contains another stair U_{i+1}.

2. *Selection and agreement are correlated.* U_i contains a morpheme μ_i that selects U_{i+1}. Local affixation rules will morphologically merge the head of U_i with the morpheme μ_i, thereby correlating selectional properties of μ_i with the agreement features of U_i in the lexicon.

3. *Undergoes obligatory movement.* U_i selects and assigns a theta-role to U_{i+1}, but does not assign it case. Therefore U_{i+1} is a properly governed argument that undergoes obligatory movement in order to satisfy the case filter. (The same will be true for U_i.)

4. *Transparent to extraction.* U_i allows nodes that can be moved out of U_{i+1} to also be moved out of U_i. (This kind of long movement is typically done by successive cyclic movement between bounding nodes in order to satisfy the subjaceny condition of bounding theory.)

5. *Contains a landing site.* U_i contains a specifier position that is assigned case. The head of U_i will agree with its specifier; therefore only stairs that agree with the head of U_i can be moved to this specifier position. (Correspondingly, this means that U_i can only move to the specifier position of a stair U_j, $j < i$, that agrees with it.)

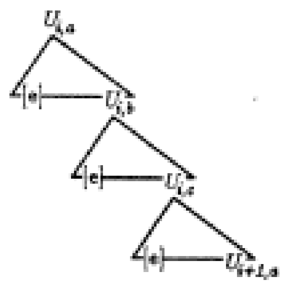
Recall the 3SAT constraints on page 35: (i) negation: a variable x_j and its negation $\overline{x}_j$ have opposite truth values; (ii) clausal satisfaction: every clause $C_i = (a_i \vee b_i \vee c_i)$ contains a true literal (a literal is a variable or its negation); (iii) consistency of truth assignments: every positive literal of a given variable is assigned the same truth value, either 1 or 0.

Lemma 3.2.1 *Stairs can enforce the 3SAT constraints.*

Proof. The idea of the proof is to represent negation as a morpheme; to encode the truth values of variables in syntactic features; to enforce clausal satisfaction in the underlying representation (DS), using selectional constraints; and to ensure consistency of truth assignments in the surface representation (SS), using long distance movement and specifier-head agreement.

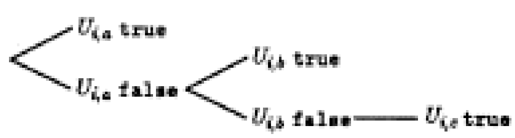
The DS consists of one stair per formula literal, which is three stairs per formula clause. Let the clause $C_i = (a_i \vee b_i \vee c_i)$ be represented by the three stairs $U_{i,a}$, $U_{i,b}$, and $U_{i,c}$:

(5)



The selectional constraints of the three stairs ensure that each 3-clause contains at least one true literal, although lexical ambiguity will prevent us from knowing which literals in the 3-clause are true. To do this, the first stair $U_{i,a}$ must promise to make C_i true, either by being true itself or by selecting a stair $U_{i,b}$ that promises to make the 3-clause true; to fulfill its promise, the second stair $U_{i,b}$ must either be true or select a true stair $U_{i,c}$. (If $U_{i,a}$ is true, it selects the next stair $U_{i,b}$ with either truth value.) This chain of selectional dependencies is shown in (6).

(6)



Affixes listed in the lexicon will negate or preserve variable truth values, according to whether the corresponding formula literal is negative or positive.

Then, scanning from right to left, each stair is moved to the specifier position of the closest stair of the same variable, either by long movement or by successive cyclic movement (see figures 3.5, 3.6).

In the resulting SS, the specifier position of the stair that corresponds to ith occurrence of a given variable contains the stair that corresponds to the

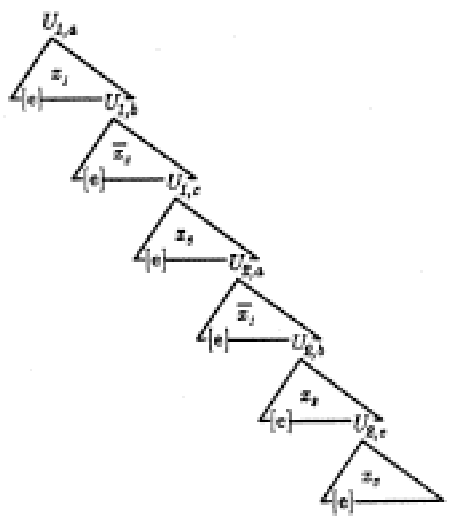


Figure 3.4: On the input 3SAT instance $f = (x_1, \overline{x_2}, x_3), (\overline{x_1}, x_2, x_3)$, the DS in the figure is created to represent f. Each literal in f is represented by a stair construction. For example, the first literal of the first clause, x_1, is represented by the outermost stair construction, $U_{1,a}$. Selectional constraints are enforced at DS. They ensure that every Boolean clause contains a true literal.

$i + 1$th occurrence of the same variable. These two stairs agree with each other by specifier-head agreement. Now all the stairs the correspond to literals of a given variable are contained in the specifier position of the stair that corresponds to first occurrence of that variable (see figure 3.6).

Now all variables have consistent truth assignments, by specifier-head agreement at SS. All clauses contain a true literal by DS selection. Negation is performed by affixes. The formula is satisfiable if and only if the corresponding DS and SS are well-formed. $\square$

Using the construction in lemma 3.2.1, and the fact that words may be ambiguous, we can now prove the following theorem about the lexical resolution problem:

Theorem 6 *The LRP is NP-hard in models that permit a stair.*

Proof. By reduction to 3SAT. The input is a Boolean formula f in 3-CNF;

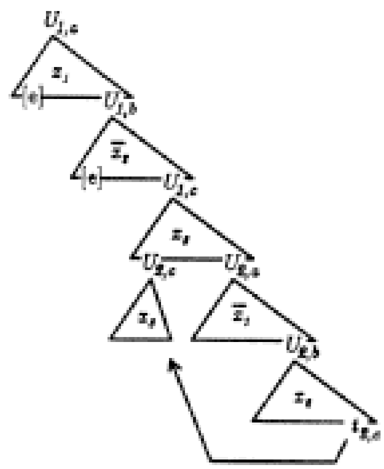


Figure 3.5: This figure depicts the first movement transformation that is applied in the mapping of the DS in figure 3.4 to the SS in figure 3.6. The innermost stair ($U_{2,c}$, representing the last literal in f) moves to the specifier position of the third stair ($U_{1,c}$), leaving behind a trace $t_{2,c}$. This movement transformation relates the x_3 literal of the second Boolean clause to the x_3 literal of the first Boolean clause. Now both stairs agree, by specifier-head agreement; therefore, the corresponding literals of the formula variable x_3 will be assigned the same truth value, even though they appear in different clauses.

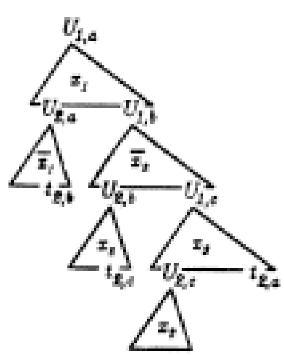


Figure 3.6: This figure shows the SS that results from repeatedly applying movement transformations to the DS depicted in figure 3.4. Specifier-head agreement is enforced at SS. It ensures that all instances of a variable are assigned the same truth value.

the output is a lexicon L and a structure S containing underspecified words such that the words in S can be found in L if and only if f is satisfiable. The structure S will be built from f according to the stair construction in lemma 3.2.1. Two stairs will agree if and only if they correspond to literals of the same variable and have been assigned the same truth value. The words in the syntactic structure will be ambiguous only in the syntactic distinction that corresponds to truth value. One agreement feature encodes variable truth assignments, and another identifies Boolean variables. One non-agreement feature encodes literal truth values, and a second one keeps track of the promise in the chain of selectional dependencies shown in (6). The stair construction ensures that the 3SAT constraints are satisfied by all permissible lexical choices for the words. $\square$

3.3 Complexity of agreement interactions

A central question for current transformational theories of syntax, such as Chomsky (1986) and Lasnik and Saito (1984), is what are the consequences of interacting agreement relations, such as specifier-head agreement, head-head agreement, head-projection agreement, and the various forms of chain agreement (link, extension, composition)?

In this section, we reduce this broad question to the narrow question: can these transformational theories simulate the stair? If yes, then we have proved that the LRP for those theories is NP-hard. This, in turn, will give us reason to believe that the interaction of agreement relations can be quite complex in these models.

Lemma 3.3.1 *Barriers allows a stair.*

Proof. The noun complement structure depicted in figure 3.7 is a stair according to the Barriers model of Chomsky (1986). (The definition of a stair appears on page 51.)

1. *Recursive structure.* NP_i contains NP_{i+1}, the next stair.
2. *Selection and agreement are correlated.* NP_i contains a verbal morpheme V0 that selects NP_{i+1}. V0 undergoes obligatory head movement to the inflectional element I0, creating an inflected verb in the

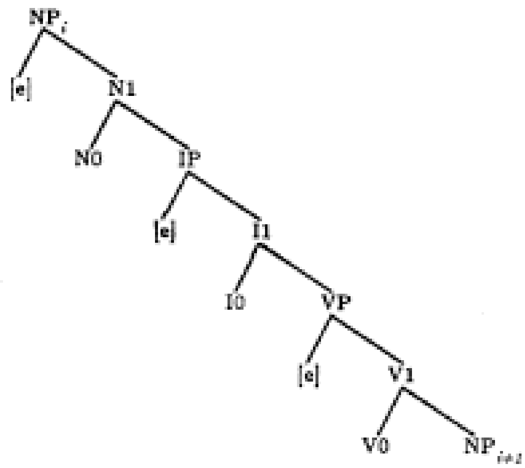


Figure 3.7: A stair construction for the Barriers model of Chomsky (1986). This is the phrase structure that would be assigned to noun complement constructions, such as *desire to visit places*.

head of IP. The φ-features will appear on the inflected verb by specifier-head agreement, where they may be systematically correlated with the verb's selectional properties in the lexicon.

3. *Undergoes obligatory movement.* V0 selects and assigns a theta-role to NP_{i+1}, but does not assign it case. Therefore NP_{i+1} must move. This is possible if V0 has lost its ability to assign case (passive morphology) or if NP_{i+1} is the underlying subject of VP_i, as in currently popular VP-internal subject analyses.

4. *Transparent to extraction.* In Barriers, blocking categories (BCs) stop unbounded application of move-α. Informally, a BC is a category not theta-marked by a lexical X0. For example, matrix verb phrases are BCs because they are selected by the nonlexical category I0 (inflection) without being assigned a theta-role. Unbounded A-movement becomes possible when a category is moved local steps, adjoining to intermediate nonargument positions before moving on (adjunction is typically to BCs).

In our noun complement construction (figure 3.7), NP_{i+1} can be moved out of NP_i. VP is a BC and a barrier for NP_{i+1} because it is not L-marked, but NP_{i+1} can adjoin to the nonargument VP and void its barrierhood because nonarguments may be freely adjoined to. Both NP_i and IP_i are L-marked, and therefore are neither BCs nor barriers for further NP_{i+1} raising. Thus, NP_{i+1} can be A-moved to any c-commanding specifier-of-IP position $[e]$ without violating the ECP because all traces are properly governed (both theta-governed by the verb V that selects NP_{i+1}, *and* γ-marked (antecedent-governed) by the deleted trace adjoined to VP).

Reinhart (pc) suggests a similar, albeit marginal, natural example where an NP containing an argument trace is topicalized to CP specifier from an L-marked position:

(7)
a. ?[What burning t_{i+1}]$_i$ did John say [of what book]$_{i+1}$ [t_i would be magnificent]
b. [what burning]$_i$ did John say [[t_i of what book] would be magnificent]

Chomsky (pc) suggests that the proper analysis of (7) is (7b), and that a better topicalization example is (8).

(8) What burning did John say (that) of that book, Mary thought would be magnificent.

5. *Contains a landing site.* The internal IP_i contains a specifier position (landing site) that will agree with I0 by specifier-head agreement in nonlexical categories; the specifier position will also agree with N0 (the head of NP_i), by predication. Alternately, head movement from V0 to I0 to N0 can create an inflected noun "[[V I] N]" in the X0 position of NP_i that will agree with the landing site. Although it is difficult to find a natural example of such an inflected noun, no arguments or analyses exclude it in principle. A close natural example is noun incorporation in Mohawk verbs (Baker 1985:139).

This establishes that the noun complement construction in figure 3.7 is a stair in the Barriers model. □

Theorem 7 *The LRP is NP-hard in Barriers model.*

Proof. The proof follows from lemma 3.3.1 and theorem 6 above. The lexicon contains ambiguous inflected nouns "[[V I] N]" that have undergone verbal incorporation. $\square$

Ristad (1988:22–28) contains a direct proof of this theorem 7, with all the details explicitly worked out.

Theorem 8 *The LRP is NP-hard in the Lasnik-Saito model.*

Proof. The preceding proof proceeds without alteration in the Lasnik-Saito (1984) model because in that model, theta-government suffices for proper government, and traces may be deleted after γ-marking. $\square$

How might we change the Barriers model in order to block the preceding reduction?

The preceding proof relies on long movement of the NP complement of a verb (in a noun complement construction), which is precisely what Barriers strives to prevent by reducing proper government to antecedent government, using the Lasnik-Saito γ-marking mechanism. (The commitment to eliminate theta-government from the definition of proper government is tentative at best. The strongest position taken is "Possibly, a verb does not properly govern its θ-marked complement," p.79.) In the Barriers stair construction, an argument undergoes long movement by adjoining the argument NP to VP, γ-marking its trace, and then deleting the intermediate A'-trace at LF.

This is the exact derivational sequence (adjoin, γ-mark, delete adjoined trace) used in Barriers (pp.21-22) to move a wh-subject from a theta-marked CP complement to a specifier of CP, provided the wh-phrase is licensed at LF. Barriers attempts to exclude similar long movement of an NP from a similar (but caseless) subject position by appeal to Binding Theory condition C at S-structure: the NP trace in subject position would be an A'-bound R-expression A-bound in the domain of the head its chain (p.93, fn.20). (Barriers differentiates the two constructions solely by the nature of their traces: wh-traces are not R-expressions, while NP-traces are.)

Crucially, condition C will exclude long movement of NPs only if trace deletion is restricted to LF. Otherwise, adjoined traces could be deleted before causing an S-structure binding violation. But trace deletion cannot be restricted solely to LF. If it were, then any ECP violation created by LF-movement may be avoided, simply by deleting offending intermediate traces after they have done their γ-marking duty. This can be done because adjoined A'-traces are not required by the extended projection principle at LF. This is why neither Barriers nor Lasnik-Saito in fact restrict trace-deletion to LF. Therefore, long movement of an NP is not excluded in these models.

There is another conundrum. The long movement used in the proof is applied cyclically, so that the trace of the argument NP is no longer c-commanded by the argument NP once all movement has applied, and hence is not A-bound by the head of its chain at S-structure. This violates the c-command condition on chain links, but such violations are standardly ignored in the literature, and therefore do not raise any special problems here. Structures where such violations are ignored include the topicalization example (7) above, antecedent-contained ellipsis (9a) and passive VP topicalization in English (9b).

(9) a. [Everyone that Max wants to e_2]$_1$ John will [kiss e_1]$_2$
b. [[$_{VP}$ Arrested t_i by the FBI]$_j$ John$_i$ has never been t_j]

Finally, even if trace deletion were disallowed entirely, long movement would still be possible from theta-marked noun complements, and the proof of lemma 3.3.1 would proceed, because theta-government cannot be eliminated without negative consequences in the rest of the theory.

Proper government can be reduced to antecedent government only if antecedent government suffices for NP-movement (eg., passive and raising) in accordance with the chain extension operation. This fails because only the terminus of an (extended) A-chain may theta-mark or case-mark, in order to obtain the CED effect (Condition on Extraction Domains, see Barriers, p.72). Therefore, in passive constructions, where the A-chain headed by the subject NP must be extended to include the verb and inflection and thereby achieve antecedent government of the NP-trace at S-structure, the inflection will simultaneously loose its ability to case-mark the subject position. The direct consequence is that both passives in (10) violate the case filter and are ungrammatical in Barriers without theta-government, although only (10a) should be excluded.

(10) a. *[e] was killed John
b. John was killed t

In short, the chain extension required to satisfy the ECP without theta-government will prevent the subject NP from receiving case, and thereby violate the case filter. This open problem may be remedied by abandoning either (i) the case filter, which would without question be disastrous for the theory, (ii) the Barriers analysis of CED effects, which would reduce empirical coverage, or (iii) the coindexing/chain extension analysis of NP-movement, which will have the direct consequence that proper government cannot be reduced to antecedent government.

The possibility of long distance argument movement by adjunction to intermediate positions remains in more recent proposals based on the Barriers model. One such proposal, due to Chomsky (1989), is that derivations be subject to a "least effort principle," with the following provisos. LF permits only the following five elements: arguments, adjuncts, lexical elements, predicates, and operator–variable constructions. Affect-alpha must apply at LF to each illegitimate object to yield one of these five legitimate elements. Chomsky (1989:63) urges us to "consider successive-cyclic A-bar movement from an argument position. This will yield a chain that is not a legitimate object; it is a 'heterogeneous chain,' consisting of an adjunct chain and an (A-bar, A) pair (an operator-variable construction, where the A-bar position is occupied by a trace). This heterogeneous chain can become a legitimate object, namely a genuine operator-variable construction, only by eliminating intermediate A-bar traces. We conclude, then, that these must be deleted at the point where we reach LF representation."

A direct consequence of this theory, then, is that successive-cyclic A-bar movement from a theta-marked argument position to a case-marked argument position will also yield an illegitimate object, that can become a legitimate object, namely an A-chain, only by eliminating intermediate A-bar traces at the point where we reach LF (that is, before LF chain conditions apply). We conclude, then, that these intermediate traces must be deleted at that point, and that long distance NP movement is permitted by the theory.

3.4 Conclusions

3.4.1 Locality in linguistic theory

The guiding idea behind transformational theories is to map each linguistic relation $R(x, y)$ into a local phrase structure configuration at some level of representation. When this is not possible, because x and y are not proximate at any level of representation, then the relation $R(x, y)$ must be broken into a chain of local relations $R(x,t_1), R(t_1,t_2), \ldots, R(t_n, y)$ using intermediate elements t_i.

Locality in linguistic representations is a way to describe complex interactions with intervening elements. When a nonlocal relation $R(x, y)$ is broken into a chain of local relations, then an element z that is on the path between x and y can affect the nonlocal relation, by interacting with one of the intermediate positions in the chain of local relations. For example, move-α is an operation that induces a "movement" relation between the moved element and its trace. This operation is constrained by subjacency and by the ECP. As a consequence, unbounded chain dependencies can arise only from successive cyclic movement, which constructs intermediate traces. If some step of the movement is blocked, as when an intermediate landing site is filled or there is a barrier to movement, then the nonlocal movement dependency is blocked.

Locality, then, is only a constraint on the way relations are described. Any nonlocal relation can always be described as a chain of local relations. In fact, all recursively enumerable sets have local descriptions by definition, because the idea of an "effective procedure" is one that consists entirely of simple local operations, too elementary to be further decomposed (Turing, 1936; Minsky, 1969). In short, "locality" has no empirical consequences and cannot be falsified.

Of course, a particular linguistic constraint, that happens to be stated in terms of local configurations, can always be called a "locality constraint." And particular constraints may in fact be falsifiable. However, no empirical evidence can distinguish the way in which a constraint is stated, that is, in terms of local or nonlocal configurations, and therefore "locality" is not itself a source of constraint.

A case in point is the intermediate traces of Barriers-type theories, whose only apparent purpose is to allow the iterated local description of nonlocal movement relations. Intermediate traces result from a particular conception of the ECP as a local bound on move-α. There is no direct empirical evidence for the existence of intermediate traces. Nor is there indirect evidence, simply because they do not interact with other components of the grammar. For example, they might have binding or phonological effects. Adjunct traces may satisfy the ECP only via antecedent government at LF; as a consequence, adjunct extraction results in intermediate traces that may not be deleted at SS. Thus, the only intermediate traces required at SS are the traces of adjunct extraction, but these non-case-marked traces do not block *want+to* → *wanna* contraction, which is only blocked by case-marked elements (Chomsky 1986:162). For example:

(11) How do you wanna solve the problem?

As expected, the intermediate traces in specifier of CP and adjoined to VP do not block phonological contraction. Neither do these intermediate A'-traces affect binding relations, whose domain is NPs in A-positions:

(12) [which woman]$_i$ did John [$_{VP'}$ dream [$_{CP}$ t_i^2 [$_{IP}$ Bill [I0 [$_{VP}$ t_i^1 [$_{VP}$ saw $\left\{ \begin{array}{c} \text{Joan} \\ {*}\ \text{herself} \end{array} \right\}$ [$_{PP}$ with t_i]]]]]]]

The governing category of the direct object is IP (the complete functional complex), and therefore the c-commanding trace t_i^1 adjoined to VP could bind the anaphor in object position within its governing category, if the trace were in an A-position. But, as expected, *herself* is in fact unbound, which strongly suggests that t_i^1 is only relevant to the computation of nonlocal A'-movement as constrained by the ECP. The precise formulation of the ECP, and the existence of the intermediate traces it requires, is the topic of much active research and debate. But the fact that these intermediate traces do not enter into other syntactic relations casts doubt on their explicit syntactic representation, at least in my mind.

Finally, locality has no logical relation to explanatory adequacy. The linguistic theory that best satisfies the locality requirement is generalized phrase structure grammar. In GPSG, all linguistic relations are reduced to maximally local relations between mother and daughter, or among sisters. Relations may extend beyond immediate domination only by iteration, whether from lexical head to its projections, or from gap to filler in a unbounded dependency relation. Because all relations are uniformly represented in syntactic categories, many formal devices may interact in constraining the distribution of a given linguistic relation. This, when coupled with the iteration of local relations to achieve nonlocal effects, can lead to severe computational intractability: the universal recognition problem for GPSGs can take more than exponential time (Ristad, 1986). More importantly, there are infinitely many unnatural GPSG languages, including finite, regular, and context-free languages. Thus, the linguistic theory that most closely satisfies the locality requirement lacks both computational constraint and explanatory adequacy.

The mapping of linguistic relations onto local configurations must therefore be justified in the same way that descriptions are, by criteria such as elegance, perspecuity, and expressive power. Locality does not always result in elegant linguistic representations; nor can all interactions be naturally modeled in this manner.

Nonlocal relations are broken into a chain of local relations in order to explain potential interactions with intervening elements. When there are no actual interactions, the resulting representations are inelegant, containing superfluous intermediate elements. A uniform bound on the domain of relations, as in Koster (1987), will allow too many potential interactions that won't occur. More seriously, as shown by the constructions in this chapter, it is difficult to prevent undesirable interactions from occurring in such a system of representation. (The alternative is a relativized bound on the domain, that models all and only the actual interactions with intervening elements. But this is identical in conception to a nonlocal relation, antithetical to locality.)

Not all interactions can be naturally described in terms of local configurations. For example, a linguistic relation that depends on elements outside its domain cannot modeled via local interaction. The transitive relations of obviation, arising from the binding theory, have nonlocal effects on relations of antecedence, and these interactions are not naturally modeled in terms of local configurations.

3.4.2 The search for uniform mechanisms

The pursuit of general mechanisms for linguistic theory—such as feature unification, the uniform local decomposition of linguistic relations, or the coindexing mechanism of Barriers—-has repeatedly proven treacherous in the study of language. It distracts the attention and efforts of the field from the particular details of human language, that is, what are the true representations, constraints, and processes of human language.

General mechanisms have also invariably resulted in unnatural intractability, that is, intractability due to the general mechanisms of the theory rather than the particular structure of human language. This is because no one mechanism has been able model all the particular properties of human language unless it is the unrestricted mechanism. However, the unrestricted mechanism can also model unnatural properties, including computationally complex ones.

In segmental phonology, rules are needed to insert, delete, and transpose segments. Rules are also needed to perform arbitrary substitutions, as in the case of irregular forms. Therefore, we conclude that phonological rewriting rules must be completely unrestricted. However, this is a false conclusion, because we have entirely ignored the possible interactions between rules. In an unrestricted rewriting system, each rule applies to the derivation string in a Markovian fashion, entirely oblivious to the previous rules that have applied. But in a phonological system this is not the case. A phonological rule cannot delete a segment that was inserted by another rule: inserted segments are never rewritten and then deleted. Nor can arbitrary segmental strings be arbitrarily rewritten: irregular forms may only be rewritten at the interface between morphology and phonology, where all morphemes are rewritten as segmental strings.

In current syntactic theories, many different types of agreement are used, including specifier-head agreement, head-complement agreement (selection), head-head agreement, head-projection agreement, and the various forms of chain agreement (link, extension, composition). Therefore, we conclude, all agreement may be subsumed under the most general agreement mechanism, either feature unification (as in GPSG/LFG) or the coindexing operation (as in Barriers). However this conclusion invariably leads to unnatural analyses. Specifier-head agreement includes the relations of morphological specification and of predication, which is the saturation of an external theta-role. Head-complement agreement includes selection relation, which is sensitive to an entirely different set of syntactic distinctions than morphological specification is. So, for example, the agreement morpheme represents certain distinctions—such as person, gender, or number—that selection is not sensitive to, at least in English. English verbs do not select plural objects, although they are morphologically marked for the plurality of their subjects. The assignment of theta-roles is likewise insensitive to number, person, or gender. When all these particular forms of agreement are subsumed under one general mechanism, whether it be unification or coindexing, unnatural forms of agreement invariably result from interactions. (The unification mechanism, however, is considerably more general and powerful than the coindexing mechanism of Barriers.) The complexity investigations in this chapter have exploited this flaw in current transformational theories, by simulating the unnatural stair construction.

In a way, these overgeneralizations reflect the mindset of formal language theory, which is to crudely equate structural complexity with syntactic form. By choosing the least general rule format that includes all the natural rules, we need not allow any unnatural rules. However, as we have seen, we do allow unnatural computations, because the resulting rule interactions are almost surely unnatural.

The remedy is, we must adopt the mindset of computational complexity theory, which is to equate structural complexity with computational resources. In order to limit resources, we must limit the number of possible rule interactions. The only way to satisfy these limits is to look for a more powerful class of linguistic constraints, that limit interactions among linguistic processes.

Chapter 4

Complexity of Anaphora

This chapter defends the thesis that human language is NP-complete. In order to defend such a thesis, we must defend both the upper bound, that language comprehension is in $\mathcal{NP}$, and the lower bound, that language comprehension is NP-hard.

The chief obstacle that we face in defending either bound is the incompleteness of our understanding of human language. Because our understanding is incomplete, it would be premature to formalize the linguistic theory. It would also be meaningless, because in order to be precise, a formal model of human language would make statements that could not be justified scientifically. Lacking a comprehensive formal model, it is not possible to prove the upper bound. Nor can we prove the lower bound without a precise statement of the language comprehension (LC) problem.

We overcome the impossibility of defining the complete LC problem as follows. First we select a natural class of utterances, and use the scientific methods of linguistics to determine what knowledge language users in fact have about those utterances. Next, we construct the simplest theory of that knowledge, under an appropriate idealization to unbounded utterances. Finally, we pose the abstract problem of computing that knowledge for a given utterance. This problem is a natural subproblem of language comprehension, because in order to comprehend an utterance in that class, the language user must compute that knowledge. Therefore, the complexity of such a subproblem is a lower bound on the complexity of the complete LC problem, by the principle of sufficient reason.

Although we cannot prove the upper bound, we can still accumulate empirical evidence for it. One way to confirm a thesis is to confirm its predictions. The upper bound makes the following prediction: if an analysis of a linguistic phenomena leads to complexity outside $\mathcal{NP}$, then the analysis is in error, and the phenomena has an empirically superior analysis whose complexity is inside $\mathcal{NP}$. Therefore, every time that we improve an analysis while reducing its complexity from outside $\mathcal{NP}$ to inside $\mathcal{NP}$, we accumulate additional empirical evidence for the upper bound.[1]

In this chapter, we illustrate both upper and lower bound techniques with respect to the language user's knowledge of anaphora. *Anaphora* is the process of interpreting anaphoric elements. An *anaphoric* or *referentially-dependent element* is a word or morpheme that lacks intrinsic reference, such as a pronoun or an anaphor. An *anaphor* is a reflexive, such as is marked in English by the [+*self*] suffix, or a reciprocal, such as the English *each other*. A *pronoun* is an element that does not have its own reference, such as the English *they*, and is not an anaphor.

In order to completely understand an utterance, the language user must determine the intended antecedent of every anaphoric element in the utterance; otherwise, he has failed to understand the utterance. However, there is no known satisfactory characterization of what is "the intended antecedent of an anaphoric element." It cannot be *any* antecedent, because this results in the trivial language *mis*comprehension problem. Such a problem statement would allow the trivial (non)solution where every pronoun in the utterance is assigned a distinct antecedent, none of which were previously mentioned in the discourse.

In order to overcome this difficulty without making any unjustified or un-

[1] One piece of evidence for this upper bound may be found in Ristad (1986), which proves that the universal recognition problem (URP) for generalized phrase structure grammars (GPSG) is EXPPOLY-hard, and shows how to construct an empirically superior "Revised GPSG," whose URP is NP-complete. Another piece of evidence is the analysis of phonological theories in chapter 2, whose complexity is reduced from undecidable to inside $\mathcal{NP}$. To my knowledge, no other linguistic theory has been proved to have a complexity outside of $\mathcal{NP}$. The work of Peters and Ritchie (1973), who proved that a formal model of their own design was undecidable, and Rounds (1975), who proved that a restricted version of the Peters-Ritchie model was exponential time, is not relevant because their formal model was not independently proposed by linguists, or ever defended as a remotely plausible linguistic theory. It amounts to little more than a very loose interpretation of the statement "a grammar maps deep structures to surface structures by the repeated application of rewriting rules."

necessarily strong assumptions, we require that solutions to the anaphora problem introduce no new information. That is to say, the antecedent of each anaphoric element must be drawn from the set of available antecedents, in the current utterance and in previous utterances, produced earlier in the discourse. The anaphora problem must also be defined in terms of structural descriptions, and not expressions or utterances, in order to prevent the trivial solution where the expression is assigned the null structural description, and no anaphoric elements are understood to be present.

For these reasons the *anaphora problem* is: Given a structural description S lacking only relations of referential dependency, and a set A of available antecedents, decide if all the anaphoric elements in S can find their antecedents in A. The set of available antecedents models the context in which the utterance is produced.

The chapter is organized into three sections. Section 4.1 reviews the language user's knowledge of referential dependencies, and proves in two entirely different ways that the anaphora problem is NP-hard, thereby establishing the NP-hardness of language comprehension as a whole and demonstrating the power of a direct complexity analysis. Section 4.2 shows how a widely-accepted analysis of the linguistic phenomenon of ellipsis leads to a proof that the anaphora problem is PSPACE-hard. Next, I falsify this analysis empirically, and sketch an improved analysis of ellipsis that reduces the complexity of anaphora to $\mathcal{NP}$. This illustrates the utility of the upper bound. The conclusion evaluates an alternate approach to the mathematical investigation of human language, based on the complexity analysis of linguistic theories.

4.1 Two proofs of the $\mathcal{NP}$ lower bound

In this section, we use basic facts about the language user's knowledge of referential dependencies to prove in two different ways that the anaphora problem is NP-hard. What is it that language users know about referential dependencies?

For one, language users know that an anaphoric element α may inherit the reference of an argument β, as in *Todd*$_1$ *hurt himself*$_1$, where the anaphor *himself* is understood as referring to the proper noun *Todd*, or *Todd*$_1$ *said Mary liked him*$_1$, which could mean that 'Todd said Mary liked Todd'. The judgement of coreference between α and its antecedent β is depicted here by assigning α and β the same integral subscript. (Careful attention must be paid to the the intended interpretation of the anaphoric elements in the examples below, as indicated by the subscripts.)

Language users also know that an anaphoric element must agree with its antecedent in certain respects. Examples (13) are possible only if *Chris* is masculine, whereas (14) are possible only if *Chris* is feminine.

(13) a. Chris$_1$ liked himself$_1$.
b. Chris$_1$ thought Bill liked him$_1$.

(14) a. Chris$_1$ liked herself$_1$.
b. Chris$_1$ thought Bill liked her$_1$.

This condition on agreement is transitive, as illustrated by the paradigm in (15), where *the student* can be masculine (15a) or feminine (15b), but not both simultaneously (15c).

(15) a. The student$_1$ prepared her$_1$ breakfast
b. The student$_1$ did his$_1$ homework.
c. *The student$_1$ prepared her$_1$ breakfast after doing his$_1$ homework.

The asterix '*' is used in (15c) to indicate the impossibility of the depicted interpretation.

These facts of interpretation are widely described as the standard agreement condition (SAC): all anaphoric elements that share an antecedent must be nondistinct from it and from each other. Two elements are *nondistinct* if and only if they do not disagree on any common feature (that is, they may be unified).

Pronouns in different languages are marked for a wide range of distinctions in person, gender, number, animacy, case, social class, kinship, reference, antecedent noun class, grammatical function, thematic role, and so on (cf., Wiesemann 1986; Sells 1987). It is true that every particular language contains a fixed number of agreement features. However, linguistic theory idealizes to an unbounded number of agreement features because these features and the range of their possible values varies considerably from language to language and does not seem to be restricted in principle (see appendix A.2).

The fact that all anaphoric elements that share an antecedent must agree with it and with each other, in combination with the fact that anaphoric elements must have antecedents, suffices to establish the following lemma.

Lemma 4.1.1 *Anaphoric agreement can simulate graph k-coloring.*

Proof. On input k colors and a graph $G = (V, E)$ with vertices $V = \{v_1, v_2, \ldots, v_n\}$ and edges E, we construct a linguistic expression S containing $|V|$ pronouns and k available antecedents such that G is k-colorable if and only if the pronouns in S can refer to the k available antecedents. Available antecedents correspond to colors; pronouns in S correspond to vertices in G; and disagreement between the pronouns in S corresponds to edges in G.

To do this we need the n binary agreement features $\varphi_1, \varphi_2, \ldots, \varphi_n$; the pronouns $p_1, p_2, \ldots, p_n$; and the available antecedents $R_1, R_2, \ldots, R_k$. Each R_i is an argument, such as a noun phrase. Pronoun p_i represents vertex v_i: for each edge $(v_i, v_j) \in E$ attached to v_i, pronoun p_i has $\varphi_i = 0$ and pronoun p_j has $\varphi_i = 1$. It does not matter how the pronouns and arguments are arranged in the expression S, provided that every argument is a possible antecedent for each pronoun, and that no other linguistic constraints interfere with the disagreement relations we are constructing. It is always trivial to quickly construct such a sentence, as we did in example (15).

In order to be understood, every pronoun must refer to one of the k available antecedents. If there is an edge between two vertices in the input graph G, then those two corresponding pronouns cannot share an antecedent in the expression S without disagreeing on some agreement feature. Therefore each permissible interpretation of S exactly corresponds to a k-coloring of the input graph G. $\square$

The reduction uses n binary agreement features, one for each vertex in the graph. The feature system is used to represent subsets of the n vertices, and therefore must be capable of making an exponential number of distinctions. (In terms of the input length $m = |G|$, this feature system is capable of making $2^{m^{1/2}}$ distinctions.)

4.1.1 Agreement reconsidered

Agreement is always stated in terms of nondistinctness of features (cf. SAC). The nondistinctness relation may be broken into three mutually-exclusive collectively-exhaustive subcases as follows. Two nonidentical elements are nondistinct if and only if (i) they are *orthogonal* (have no common feature, eg., [-plu] and [+masc]); (ii) one *subsumes* the other (is strictly more general than, eg., [+plu] subsumes [+plu,+masc]); or (iii) they *partially overlap* (have some but not all features in common, eg., [-plu,person 1] and [-plu,+masc]).

The standard agreement condition makes two broad empirical predictions. However, neither appears to be true.

The first prediction is that there are languages with nonidentical nondistinct pronouns. Otherwise, the SAC cannot be falsified, and should be abandoned because it is unnecessarily powerful. The strongest confirmation of the SAC, then, would be to find languages for each of the three subcases of nondistinctness: languages with orthogonal pronouns, languages with pronouns that subsume each other, and languages with pronouns that partially overlap. Such a language would have a pronoun system that could not be written down in the standard textbook format (a chart that partitions the space of possible feature combinations).

Although I am out of my depth at this point, I observe that English is not such a language, nor could I find such a language among those spoken in Europe, or in the fourty nine non-European languages discussed in Wiesemann (1986). The English system of personal pronouns exactly partitions the space of possible feature combinations, as shown in the following table (suppressing case marking):

<table>
<tr><th></th><th colspan="2">singular</th><th>plural</th></tr>
<tr><td>1st</td><td colspan="2">I</td><td>we</td></tr>
<tr><td>2nd</td><td colspan="3">you</td></tr>
<tr><td rowspan="3">3rd</td><td>masc</td><td>he</td><td rowspan="3">they</td></tr>
<tr><td>fem</td><td>she</td></tr>
<tr><td>neut</td><td>it</td></tr>
</table>

(The pronoun *one* is not a counterexample, because it does not have a linguistic antecedent. Rather, it has an arbitrary interpretation and may only "accidentally" corefer.)

The second prediction is that nonidentical nondistinct pronouns can share an antecedent. Given the difficulty of obtaining truly nondistinct pronouns, this prediction is not easily tested. However, we may imagine the following scenario. You are talking to a new friend about his recent health checkup, and, not wishing to make any assumptions about the gender of the examining doctor, you use the third person plural pronoun (16a) instead of either third person singular pronoun (16b).

(16) a. About your doctor_1, did they_1 seem competent?
b. About your doctor_1, did he_1/she_1 examine you thoroughly?

In this case, *they* is used as a pure third person pronoun, unmarked for number or gender, and hence subsumes both *he* and *she*. However, it is still not possible to use *they* in combination with *he* or *she*, contra the standard agreement condition:

(17) *About your doctor_1, did they_1 seem competent when he_1/she_1 examined you?

These and related facts suggest that the agreement condition must be strengthened, to state that every argument in a sentence must have a unique anaphoric root morpheme in each of the language's anaphoric systems. That is, we assume that the systems of personal, relative, demonstrative, and interrogative pronouns are all independent. This version of the agreement condition, which partitions the set of anaphoric elements into equivalence classes, is the *anaphoric equivalence condition.*[2]

The anaphoric equivalence condition helps to explain the uncertain nature of obviation violations between pronouns with distinct anaphoric roots. The split antecedence examples in (18) demonstrate that—unlike obviation violations between pronouns that share the same anaphoric root morpheme—

[2]This condition may underlyingly be a fact about language acquisition. It seems that the language acquisition device first chooses a set of relevant semantic, syntactic and phonological distinctions, next partitions the resulting feature space into natural classes, and finally assigns a phonologically distinctive anaphoric root morpheme to each class (cf., Chiat 1986). One consequence of such an acquisition procedure would be the anaphoric equivalence condition. Ultimately, the anaphoric equivalence condition is a linguistic process that simplifies the relation between sound and meaning: once a pronoun is syntactically linked to a given antecedent, then the sound of that pronoun comes to stand for that antecedent within the scope of the utterance.

obviation is not always enforced between pronouns with different anaphoric root morphemes.

(18) a. John_1 suggested to Bill_2 that [he_2 shoot $\text{them}_{\{1,2\}}$].
  b. Bill_1 reminded Sue_2 that [he_1 introduced $\text{them}_{\{1,2\}}$ to the pope].

Rather, lexical factors, such as the choice of verb, play a significant role.

(19) a. Navarre_1 suggested to Benedict_2 that [he_2 persuade $\text{them}_{i=\{1,2\}}$ [PRO_i to perjure themselves_i]].
  b. *Navarre_1 suggested to Benedict_2 that [he_2 tell $\text{them}_{i=\{1,2\}}$ [PRO_i to perjure themselves_i]].

4.1.2 Relations of referential dependence

We have seen that anaphoric elements must have antecedents, subject to an agreement condition, perhaps the anaphoric equivalence condition. A second component of linguistic knowledge is that that pronouns must be disjoint in reference from certain arguments. For example, every English speaker knows that *Todd_1 hurt him_{*1}* cannot mean that 'Todd hurt Todd'. This judgement of disjoint reference, that α cannot refer to β, is depicted here by assigning α the integral subscript of β preceded by an asterisk.

Knowledge of coreference and disjoint reference must be represented in the brain of the language user, and by every scientifically adequate linguistic theory. The simplest, and least controversial, representation is to postulate two abstract linguistic relations: an asymmetric $\mathsf{link}(\alpha, \beta)$ relation that holds between an anaphoric element α and its immediate antecedent β, subject to the agreement condition, and a symmetric $\mathsf{obviate}(\alpha, \beta)$ relation that holds between two arguments α and β that cannot share any referential values (Higginbotham, 1985).

Every structural description, then, includes a directed graph of link relations, as well as an undirected graph of obviate relations whose vertices are the arguments in the structural description and whose undirected edges represent the obligatory nonoverlapping reference of two arguments. Let this graph of referential dependencies, which consists of the obviate and link relations of a given structural description, be called the *RDG*.

Our goal is to prove the NP-hardness of the anaphora problem without using the agreement condition. The idea of the proof will be to construct RDGs that simulate some NP-complete problem. Let us therefore examine a range of syntactic configurations, in order to better understand the distribution of link and obviate relations in linguistic representations.

The first, most important, syntactic configuration is "local c-command": an anaphor must link to some locally c-commanding β and a pronoun must obviate all such β.[3] To a first approximation, two elements are local if they are co-arguments, that is, arguments of the same verb or noun. (The exact definition of locality does not matter for our purposes; all that matters here is the fact that antecedence and disjoint reference are possible or necessary in some configurations, and not in others.) We say β *c-commands* α in a phrase structure tree if and only if all branching nodes that dominate β in the tree also dominate α. In particular, the direct object c-commands the indirect object, and the subject of a clause c-commands both direct and indirect objects.

An anaphor must be linked to a unique c-commanding argument, and this argument must be local. This is illustrated by the paradigm in (20), where the domain of locality is indicated by square brackets.

(20) a. [John$_1$ shot himself$_1$]
b. [John$_1$ introduced Bill$_2$ to himself$_{1/2}$]
c. John$_1$ thought Bill$_2$ said [Mary liked himself$_{*1/*2}$]

Example (20b) shows that an anaphor can take any c-commanding antecedent inside the local domain; example (20c) shows that an anaphor must have some antecedent inside the local domain. *Mary* is not a possible antecedent for *himself* in (20c) because they disagree on gender.

Pronouns are locally obviative: a pronoun cannot share referential values with any argument that c-commands it in its local domain. (The domain of locality is roughly the same as for anaphors; again, all is needed for the proofs below is that there exist configurations that result in obviation.) This is illustrated by the paradigm in (21).

[3] The requirement that anaphors have local antecedents is called "condition A" and the requirement that pronouns be locally obviative is called "condition B" in the linguistics literature.

(21) a. [John$_1$ shot him$_{*1}$]
b. [John$_1$ introduced Bill$_2$ to him$_{*1/*2}$]
c. John$_1$ thought Bill$_2$ said [Mary liked him$_{1/2}$]

Example (21b) shows that a pronoun is disjoint from all locally c-commanding arguments; example (21c) shows that a pronoun can link to any argument outside its local domain.

Like the agreement condition, the prohibition against sharing referential values is a transitive condition that is enforced globally, as shown by the paradigm in (22).

(22) a. John$_1$ said that [Bill$_2$ liked him$_{1/*2}$].
b. John$_1$ said that [he$_{1/*2}$ liked Bill$_2$].
c. *John$_1$ said that [he$_1$ liked him$_1$].

Him can refer to *John* in (22a); *he* can refer to *John* in (22b); but *he* and *him* cannot both refer to *John* in (22c), because *he* locally c-commands *him* and hence they are obviative.

Obviation applies equally to all linguistic coreference, including the intra- and extra-sentential linking of pronouns, because obviation cannot be violated, even if a pronoun and its antecedent are in different sentences. Without loss of generality then, all linkings in this chapter will be intrasentential for expository convenience.

The other local c-command configuration is "exceptional case-marking" (ECM), where the subject of an ECM verb locally c-commands the subject of its infinitival complement. This is illustrated by the paradigm in (23), with the ECM verbs *want* and *expect*.[4]

(23) a. John$_1$ wants [himself$_1$ to shoot Bill]
b. John$_1$ expects [him$_{*1}$ to shoot Bill]

Examples (23) demonstrate that the subject *John* of an ECM verb locally c-commands the subject α of the infinitival complement [α *to shoot Bill*], for both anaphors and pronouns.

[4]These verbs are called "exceptional case-marking" because, unlike other verbs that take a finite clausal complement, ECM verbs take an infinitival clausal complement and assign abstract case to its subject. Other ECM verbs in English include *believe*, *prefer*, *like*, and related verbs.

Our goal is to prove that the anaphora problem is NP-hard, without using any agreement features. Let us therefore pause to consider how such a proof might work.

Imagine that we must color the following four-vertex graph G_4 with three colors:

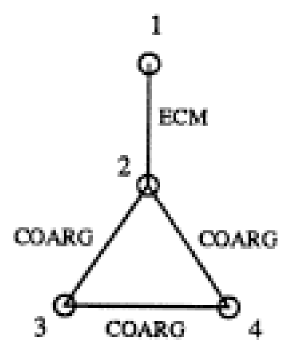
$$\{(1,2),(2,3),(3,4),(4,2)\}$$

Then our reduction might construct a sentence containing three available antecedents and four pronouns. The first part of the sentence, *Before Bill$_a$, Tom$_b$, and Jack$_c$ were friends....*, would represent the three colors, where each proper noun corresponds to a different color. The second part of the sentence would have an obviation graph equivalent to G_4, where the pronoun p_i in the sentence corresponds to vertex i in G_4. As expected, it is difficult to understand the resulting sentence:

(24) Before Bill$_a$, Tom$_b$, and Jack$_c$ were friends,
[he$_1$ wanted him$_2$ to introduce him$_3$ to him$_4$].

The corresponding obviation graph appears in (25).

(25)



Each vertex is labeled with the numerical index of its corresponding pronoun, and each edge is labeled with the syntactic configuration responsible for the corresponding obviate relation. (Recall "coarg" means "argument of same verb or noun," and "ecm" means "exceptional case marking" configuration.)

By carefully grounding the reference of each pronoun in turn, we can confirm that the obviation graph for (24) exactly corresponds to the four-vertex graph G_4. Let *he$_1$* link to *Bill$_a$* in the sentence—this corresponds to coloring vertex 1 in G_4 with the color a. Then in the simplified sentence [*Bill wanted* *him*$_2$ *to introduce him*$_3$ *to him*$_4$] we can clearly see that *Bill* can be the antecedent of any pronoun but *him*$_2$—this corresponds to G_4, where coloring vertex 1 with a given color only prevents us from coloring vertex 2 the same color. Continuing in this fashion, one can convince oneself that the pronouns in such a sentence can find their antecedents in the sentence iff the corresponding graph G_4 is 3-colorable.

The local c-command configurations used in (24) only give rise to simple obviation graphs, and therefore the second proof of NP-hardness will employ three additional syntactic configurations: adjunct control, strong crossover, and invisible obviation.

In the expression *Sue screamed before jumping*, all English speakers know that *Sue* is the understood subject of the gerund *jumping*, that is, everyone knows that Sue did the jumping. In order to represent this linguistic knowledge, as we must, we may postulate a silent pronoun 'PRO' in the subject position of the adjunct [*before jumping*], and obligatorily link PRO to *Sue*.

(26) Sue$_1$ screamed [before PRO$_1$ jumping]

This is called "subject control" because the reference of PRO is controlled by the subject of the main clause. Another example of subject control appears in (27a) with the verb *promise*. Contrast this to the verb *persuade* in (27b), which is an object control verb.

(27) a. Tom$_1$ promised Mary$_2$ [PRO$_{1/*2}$ to attend school]
b. Tom$_1$ persuaded Mary$_2$ [PRO$_{*1/2}$ to attend school]

Further evidence for the existence of this silent pronoun comes from its interaction with overt anaphora. Observe that *himself* must refer to *Bill* in (28a), and *him* must be disjoint from *Bill* in (28b).

(28) a. Mark$_1$ vomited [after PRO$_1$ getting himself$_1$ plastered]
b. Mark$_1$ vomited [after PRO$_1$ getting him$_{*1}$ plastered]

Without PRO, such facts are a complete mystery. But once the understood subject of the gerund is explicitly represented using PRO, as we have done in (28), the facts are trivially accounted for as canonical configurations of local c-command between silent PRO and an overt anaphoric element. (In any event, the complexity reduction proceeds whether PRO exists or not; all that matters for the reduction is the empirical fact that *himself* must refer to *Bill* and *him* must be disjoint from *Bill* in expressions like (28).)

"Wh-movement" is the configuration that holds between a wh-phrase, such as [*who*] or [*what person*], that appears displaced from its underlying argument position. For example, in *Who$_k$ did Mary see t_k*, the underlying position of the wh-phrase *who$_k$* as the direct object of the verb *see* has been marked with a trace t_k coindexed with it. This represents the fact that *who$_k$* stands in the same relation to the verb *see* as it does in the related expression *Mary saw who.*

"Strong crossover" occurs when an anaphoric element α intervenes between a wh-phrase and its trace, and c-commands the trace. In such a configuration, α obviates the subject of the wh-phrase. This is shown in (29a), where the pronoun *he* c-commands the trace t_k of the wh-phrase [*which person*], and for this reason must be understood as disjoint from the person who Mary kissed.

(29) a. [Which person]$_k$ did he$_{*k}$ say t_k kissed Mary.
b. [Which person]$_k$ t_k said he$_k$ kissed Mary.

In (29b), however, there is no strong crossover configuration, and no obviation. That is, (29b) has an interpretation of the form, "for which person x, did x say x kissed Mary." Such facts are difficult to explain without an explicit trace, because the wh-phrase [*which person*] and the pronoun *he* stand in the same relation in both sentences.

In the expression *The man who$_k$ Mary saw t_k*, we say that *who* heads the relative clause [*who Mary saw*], and that it predicates its subject, [*the man*]. When the relative clause contains a pronoun in a strong crossover configuration, then the pronoun obviates the subject of the relative clause, as in (30).

(30) [the man]$_1$ [who$_k$ he$_{*1}$ likes t_k].

"Ellipsis" is the syntactic phenomenon where a phrase is understood but not expressed in words, as in *The men ate dinner and the women did too*, which can only be understood to mean that 'the women did <u>eat dinner</u> too'.

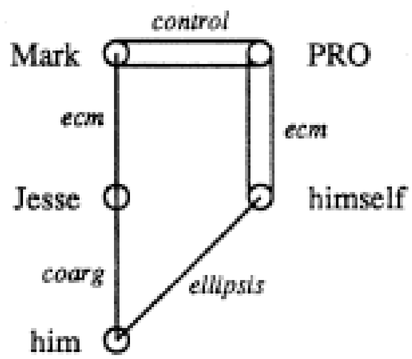
For this example, we would say that the verb phrase [*eat dinner*] has been ellipsed in the second conjunct; this is called VP-ellipsis.

A configuration of "invisible obviation" arises between the subject of an ellipsed verb phrase and the direct and indirect objects of the overt verb phrase, because both subjects in effect locally c-command the other arguments of the verb. Observe that *him* can refer to *Romeo* in (31a), but not in (31b).

(31) a. Mark_1 wanted Jesse to love him_1
b. *Mark_1 wanted Jesse to [love him_1] before PRO_1 allowing himself_1 to [e]

In this example, the anaphor *himself* is obligatorily linked to *PRO*, and *PRO* to the matrix subject *Mark*. The pronoun *him* invisibly obviates *himself*, the subject of the ellipsed VP, because they are in an invisible configuration of local c-command. The RDG for (31b) appears in (32):

(32)



Single lines depict relations of obviation, while double lines depict relations of coreference. Vertices are labelled with the corresponding noun phrase arguments, and the edges are labelled with the configurations to which they are attributable.

It must again be emphasized that the complexity classifications of this section in no way depend on the existence of traces, PRO, or on any other details of the linguistic analysis. The reduction relies only on the empirical facts of obligatory disjoint reference in configurations such as (28b), (29a), (30), and (31). The linguistic analysis is included for pedagogy, and to organize the presentation.

This concludes our survey of the language user's knowledge of referential dependence, which has been studied extensively. The next step is a direct complexity proof.

4.1.3 From satisfiability to referential dependence

The conceptually natural reduction is from graph coloring to the anaphora problem. However, the transformation of arbitrary graphs into linguistic representations is cumbersome. To overcome this difficulty, we might reduce from 3SAT, by way of the graph 3-coloring problem. That is, on input a Boolean formula f in 3-CNF, we would first use the classic reduction of Lawler (1976) to construct a corresponding instance g of the graph 3-coloring problem. Next, from g we would construct an instance (s, a) of the anaphora problem, such that the anaphoric elements in s can find their antecedents in a iff g is 3-colorable (and f is satisfiable.) By restricting our attention to this class of "3SAT graph colorings," we would simplify the reduction into the task of transforming a simple class of "difficult" graphs into linguistic representations. Of course, the intermediate "graph 3-coloring" stage of the reduction won't really be used in the proof.

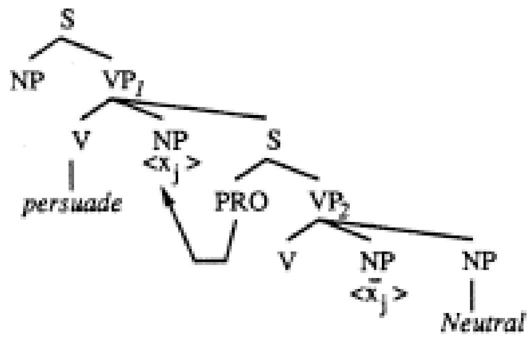
Lemma 4.1.2 *Referential dependencies can simulate 3SAT.*

Proof. On input a Boolean formula f consisting of the clauses $C_1, C_2, \ldots, C_p$ in the variables $x_1, x_2, \ldots, x_n$, we construct a linguistic representation s and a set of available antecedents a such that the anaphoric elements in s can find their antecedents in a iff f is satisfiable.

The set a will contain exactly three distinct antecedents: *True*, *False*, and *Neutral*. These noun phrases represent, respectively, the three possible truth values 'true', 'false', and 'unassigned.'

For every variable x_j in f, s will contain two pronouns, one to represent x_j and the other to represent its negation $\overline{x}_j$. In order to preserve the semantics of negation, these pronouns must obviate each other. Both will also obviate the proper noun *Neutral*, and therefore can only link to *True* or *False*. To be precise, for every Boolean variable we build an object control construction (33) that contains two possible targets for ellipsis, VP_1 and VP_2.

(33)



This is the phrase structure that would be assigned to expressions such as *He persuaded him to introduce him to Hector.*

In this construction (33), configurations of local c-command in the lower clause cause mutual obviation among PRO, the pronoun for $\overline{x}_j$, and *Neutral.* The object control verb *persuade* obligatorily links PRO to the pronoun for x_j, as depicted by the arrow. The subject position is filled with a dummy pronoun so as not to increase then number of available antecedents in the construction. The resulting RDG ensures that the pronoun for x_j must refer to either *True* or *False*; that the pronoun for $\overline{x}_j$ must refer to the other antecedent; and that neither pronoun can refer to *Neutral.* In short, the construction (33) correctly ensures consistency of truth assignments, as well as correctly representing the semantics of Boolean negation.

Both pronouns are in the direct object position of a separate verb phrase, and hence possible targets for VP-ellipsis and the resulting invisible obviation. We will take advantage of this below, by representing positive literals of x_j as ellipsis of the higher VP_1, and negative literals of $\overline{x_j}$ as ellipsis of the lower VP_2.

For each clause $C_i = (a_i \vee b_i \vee c_i)$, we construct the rather intricate syntactic structure shown in figure 4.1, whose graph of referential dependencies appears in figure 4.2. The effect of this obviation graph in combination with the limited set of available antecedents is to ensure that one of the three ellipsed verb phrases in figure 4.1 must contain a pronoun linked to the antecedent *True.* This corresponds exactly the requirement that each clause contain a true literal, concluding our simulation. □

It is not to be expected that the linguistic expressions constructed by the preceding reduction are easy to comprehend, any more than we expect to actually build the physical devices used to prove lower bounds on the com-

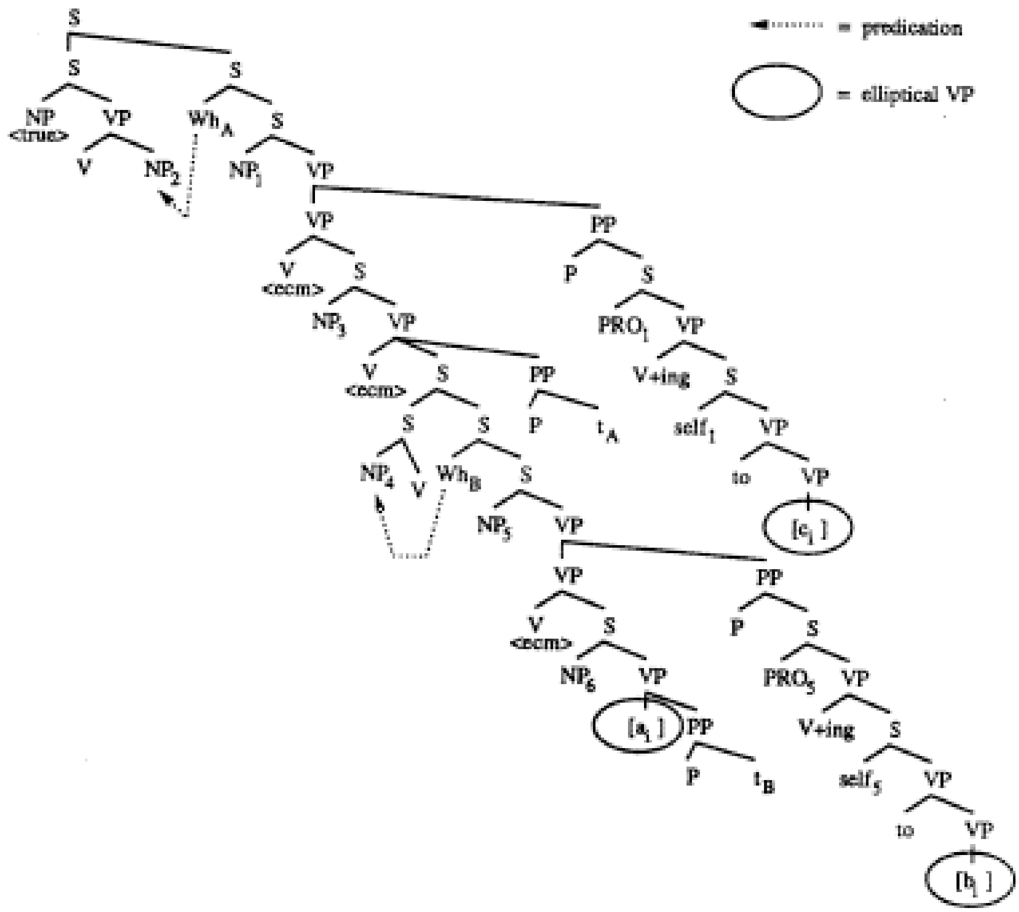


Figure 4.1: This phrase structure simulates the ith Boolean clause $C_i = (a_i \vee b_i \vee c_i)$, with irrelevant details suppressed. Dashed arrows depict the predication of a noun phrase by an extraposed relative clause. "S" is a clause, "NP" is a noun phrase, "VP" is a verb phrase, and "PP" a prepositional phrase. All NPs dominate pronouns. The structure contains configurations of local c-command, strong crossover, adjunct control, and invisible obviation, yielding the graph of referential dependencies shown in figure (4.2). The targets of the ellipsed VPs are the VPs in the literal constructions (33), depending on the polarity of the corresponding literal. Consequently, NP_6, NP_5, and NP_1 invisibly obviate the pronouns representing a_i, b_i, and c_i, respectively. This is the phrase structure that would be assigned to expressions such as *Hector met him*$_2$*, who*$_A$ *he*$_1$ *expected him*$_3$ *to want him*$_4$ *framed, who*$_B$ *he*$_5$ *believed he*$_6$ *did [e] with* t_B *after exposing himself*$_5$ *to [e] for* t_A*, before telling himself*$_1$ *to [e].* (Indices, traces of wh-movement, and elliptical VPs are included here solely to help the reader align the expression with its phrase structure.)

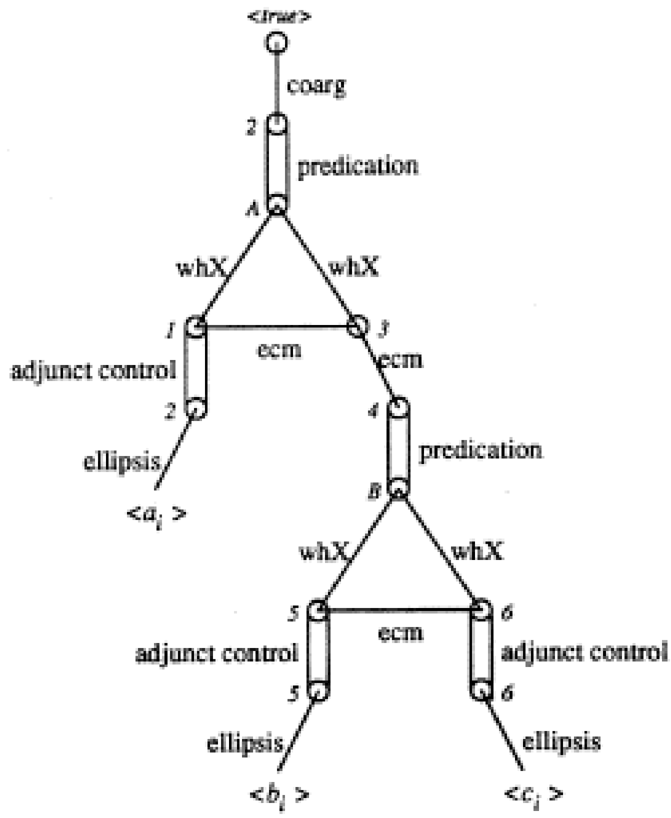


Figure 4.2: This is the graph of referential dependencies for the phrase structure of figure 4.1. Single lines depict obviation relations; double lines depict relations of coreference, that is, links and predications. Vertices in the graph correspond to noun phrase arguments, and are labelled with the identifying indices from figure 4.1. The edges in the graphs are likewise labelled with the configurations to which they are attributable. Recall that "coarg" and "ecm" are configurations of local c-command; "whX" is for a strong crossover involving a wh-phrase; and "ellipsis" for the invisible obviation arising from the ellipsis of a verb phrase representing the relevant Boolean literal from (33). The obviation graph that results when coreferential vertices are coalesced, in combination with the three available antecedents, ensures that at least one of pronouns representing a_i, b_i, or c_i must link to the proper noun *True*. This corresponds exactly to the constraint that a true Boolean 3-clause contain at least one true literal.

plexity of problems in robot motion planning (Reif 1979; Reif and Sharir 1985; Canny 1988). Certainly, it is not possible to build physical devices of such intricacy, any more than it is possible for a language user to comprehend the utterances we have just constructed. Yet the practical questions of what physical devices can and cannot be built, or of what linguistic expressions can and cannot be easily understood do not concern us here. We want to understand the theoretical structure of abstract computational problems, and use complexity analysis to better reveal this structure.

All that remains is to state the main theorem of this section.

Theorem 9 *The anaphora problem is NP-hard.*

Proof. By lemma 4.1.1, or by lemma 4.1.2. □

We now have two complexity lower-bounds for human language comprehension that rely only on the empirical facts of referential dependency. *It does not matter exactly how the conditions on coreference and disjoint reference are stated, only that there are such conditions, as there are in all known human languages.* For this reason, we may believe with confidence that the NP-hardness result applies to all adequate linguistic theories. Moreover, the directness of the reductions suggests that the anaphora problem is one of the more difficult subproblems of language comprehension, because graph k-coloring is one of the most difficult NP-complete problems, a trap-door function with no known approximation scheme, and known to be average-case NP-complete.

4.2 Evidence for an $\mathcal{NP}$ upper bound

Lacking a complete, scientifically plausible linguistic theory, it is not possible to prove an upper bound on the complexity of human language. It is however possible to provide empirical evidence for an upper bound, and that is the task of this section. The argument goes as follows. First, we examine the linguistic phenomenon of ellipsis, and present empirical arguments for a simple theory of the knowledge that language users have about such phenomenon. Next, we use this theory to prove that the anaphora problem is PSPACE-hard. Using the insights of the complexity proof, we reexamine the phenomenon of ellipsis, falsify the copy-and-link theory, and suggest an empirically superior predicate-sharing theory. Finally, we prove that the anaphora problem is in $\mathcal{NP}$ according to the predicate-sharing theory. By reducing the complexity of ellipsis to inside $\mathcal{NP}$, while strictly improving the empirical adequacy of the theory of ellipsis, we provide evidence for an $\mathcal{NP}$ upper bound.

In developing our simple linguistic theories, we will briefly introduce the relevant phenomenon, state the theory, and conclude with a concise enumeration of the empirical arguments in support of the theory.

4.2.1 Simple theory of ellipsis

A central goal of linguistics is to explicitly represent the knowledge that language users have about utterances. Let us *purely as a matter of convenience* distinguish the representation of how an utterance expressed in words and phrases, from a representation of the logical aspects of its meaning, such as referential dependencies and predication. Let us call the former representation the *surface form* of the structural description, and the latter representation, the *logical form*. (The number of levels of representation does not in and of itself affect the computational complexity.)

Theory 4.1 *The logical form of ellipsis is constructed by (recursively) copying the overt structure into the position of the corresponding ellipsed structure; an anaphoric element α may link to its antecedent either before or after copying; when the antecedent of α is a quantified NP β, then α must link to β after copying.*

Evidence. First, the elliptical structure is understood as though it were really there, as shown in (34).

(34) a. The men [ate dinner] and the women did [e] too.
b. 'the women did *eat dinner* too'.

This fact about our linguistic knowledge must be represented somehow in the logical form, perhaps by copying the overt structure into the position of the null structure, as first suggested by Chomsky (1955).

Second, an elliptical structure may itself be understood as containing an elliptical structure, as in (35a), which is understood to mean (35b).

(35)
a. Jack [[corrected his spelling mistakes]$_1$ before the teacher did $[e]_1]_2$ and Ted did $[e]_2$ too.
b. Jack corrected his spelling mistakes before the teacher did *correct his spelling mistakes* and Ted did *correct his spelling mistakes before the teacher did correct his spelling mistakes.*

This suggests that copying is a recursive process. The depth of recursion does not appear to be constrained by the principles of grammar, as shown in (36).

(36) Harry [claims that Jack [[corrected his spelling mistakes]$_1$ before the teacher did $[e]_1]_2$ and that Ted did $[e]_2$ too]$_3$, but Bob doesn't $[e]_3$.

Third, the invisible structure behaves as though it was really there. In particular, it can induce a violation of obviation, as in the discourse (37).

(37) Ann: Romeo$_1$ wants Rosaline to [love him$_i$] ($i = 1$)
Ben: Not any more—now Rosaline wants Romeo$_1$ to $[e]$
([*love him$_i$*], $i \neq 1$)

In this example, Ann's use of the pronoun *him* is most naturally understood as referring to *Romeo*. Yet when Ben replies, the coreferential interpretation ($i = 1$) is no longer possible in Ann's statement. These facts of invisible obviation are difficult to explain unless the overt structure is in fact copied in the syntax, as illustrated in (38), where the obviation violation between *him*$_1$ and *Romeo*$_1$ has been made explicit by copying the overt VP *love him* into the position of the null VP.

(38) Rosaline wants [Romeo$_1$ to *love him*$_{*1}$]

The invisible structure is not merely an invisible VP-pronoun, simply because the obviation violation in (39a) vanishes when an overt pronoun is used instead in (39b).

(39)
a. Juliet$_1$ thought the Friar$_2$ [poisoned her$_1$] without realizing that she$_{*1}$ did $[e]$.
b. Juliet$_1$ thought the Friar$_2$ [poisoned her$_1$]$_3$ without realizing that she$_1$ did it$_3$.

Fourth, corresponding anaphoric elements in the overt and invisible structures may be understood as having different antecedents, as in (40), where the invisible pronoun *his* is ambiguous, referring either to *Felix* ('invariant' interpretation) or *Max* ('covariant' interpretation).[5]

(40) Felix$_1$ [hates his$_1$ neighbors] and so does Max$_2$ [*e*].
([*hates his*$_{1/2}$ *neighbors*])

This suggests that an anaphoric element may be linked to (that is, related to) its antecedent either <u>before</u> the overt structure is copied, resulting in the invariant interpretation (41), or <u>after</u>, resulting in the covariant interpretation (42).

(41) a. Felix$_1$ [hates his$_1$ neighbors] and so does Max$_2$ [*e*].
b. Felix$_1$ [hates his$_1$ neighbors] and so does Max$_2$ [hate his$_1$ neighbors].

(42) a. Felix$_1$ [hates his neighbors] and so does Max$_2$ [*e*].
b. Felix$_1$ [hates his$_1$ neighbors] and so does Max$_2$ [hate his$_2$ neighbors].

Fifth, the invisible pronoun must agree with its antecedent, which excludes the covariant interpretations in (43) that are possible in the minimally different examples in (44).

(43) a. Barbara$_1$ read her$_1$ book and Eric$_2$ did [*e*] too.
([*read her*$_{1/*2}$ *book*])
b. We$_1$ ate our$_1$ vegetables and so did Bob$_2$ [*e*].
([*ate our*$_{1/*2}$ *vegetables*])

(44) a. Barbara$_1$ read her$_1$ book and Kate$_2$ did [*e*] too.
([*read her*$_{1/2}$ *book*])
b. We$_1$ ate our$_1$ vegetables and so did they$_2$ [*e*].
([*ate our*$_{1/2}$ *vegetables*])

---

[5]In each example, careful attention must be paid to the relevant construal of the null structure, indicated with brackets, and the intended reference of anaphoric elements, as indicated in the italicized parenthetical following the example.

Sixth, the covariant interpretation is forced when the antecedent of the anaphoric element is a quantified noun phrase (QNP), as shown in (45). (That is, (45) must mean that every boy ate his own dinner; it cannot mean that every boy ate every man's dinner.)

(45) Every man_1 [ate his_1 dinner] and so did every boy_2 [e]
([*eat* $his_{*1/2}$ *dinner*])

Therefore, an anaphoric element must be linked to its antecedent β after copying when β is a quantified noun phrase.

To summarize, we have seen evidence that the overt structure must be copied to the position of null structure in the syntax, that copying is a recursive process, and that anaphoric elements may be linked to their antecedents either before or after the copying, and that they must be linked after copying when their antecedent is a quantified noun phrase. $\square$

The earliest account of invariant and covariant interpretations in VPE, due to Ross (1967), is equivalent to this theory 4.1, because deletion in Ross's deep-structure to surface-structure derivation is identical to copying in the surface form to logical form mapping. This model has also been proposed in recent government-binding literature. See for example, Pesetsky (1982), May (1985), Koster (1987), and Kitagawa (1989).

More generally, any linguistic theory that represents the meaning of an elliptical utterance using devices that can achieve the effect of copy and link operations will inherit the complexity of the copy theory 4.1. This is true regardless of how that linguistic theory is defined, how many levels of representation it has, or what they are called.

4.2.2 Complexity outside $\mathcal{NP}$

In this report, we only consider the problem of assigning structural descriptions to utterances, which is a trivial subproblem of the much more intractable and less well understood problem of determining the semantic 'truth value' of a given utterance. The following proof shows that assigning a complete structural description to a given class of utterances can be as difficult as determining the truth of quantified Boolean formulas; the proof does *not* make the entirely unnourishing argument that determining the 'truth' of human language utterances can be as difficult as determining the truth of quantified Boolean formulas.

Lemma 4.2.1 *The anaphora problem is PSPACE-hard.*

Proof. By reduction from QUANTIFIED 3SAT. The input Ω is a quantified Boolean formula in prenex 3-CNF, consisting of alternating quantifiers $\forall x_1 \exists x_2 \ldots \forall x_{n-1} \exists x_n$ preceding (and quantifying the literals in) the clauses $C_1, C_2, \ldots, C_p$ in the Boolean variables $x_1, x_2, \ldots, x_n$. Each clause contains exactly three distinct literals labeled by $C_i = (a_i \vee b_i \vee c_i)$.

The output is a surface form S and a set A of available antecedents, such that all the anaphoric elements in S have antecedents in A if and only if Ω is true. In order to verify that all anaphoric elements in S have antecedents, we must construct the logical form of S. The reduction uses one binary agreement feature to represent literal negation, and one n-valued agreement feature (or equivalently, $\log_2 n$ binary agreement features) to identify the n distinct Boolean variables.

The idea of the proof is to mimic the structure of Ω with linguistic constructions, by reducing the quantification of variables in Ω to the linking of pronouns in S. Each quantifier Qx in Ω will correspond to a pair of available antecedents in S, one to represent $x = 0$ and the other to represent $x = 1$. Boolean literals in Ω will correspond to pronouns in S. As shown in figure 4.3, the surface form S is built from three distinct components: universal quantifiers, existential quantifiers, and Boolean clauses. We will now motivate each of these parts in turn, using intricate yet still natural English sentences.

The first step is to simulate a universal quantifier. Recall that a universally quantified predicate $[\forall x_i P(x_i)]$ is true if and only if $[P(x_i = 0) \wedge P(x_i = 1)]$. The latter Boolean formula can be expressed in a VP-ellipsis construction whose surface form is abstracted in (46).

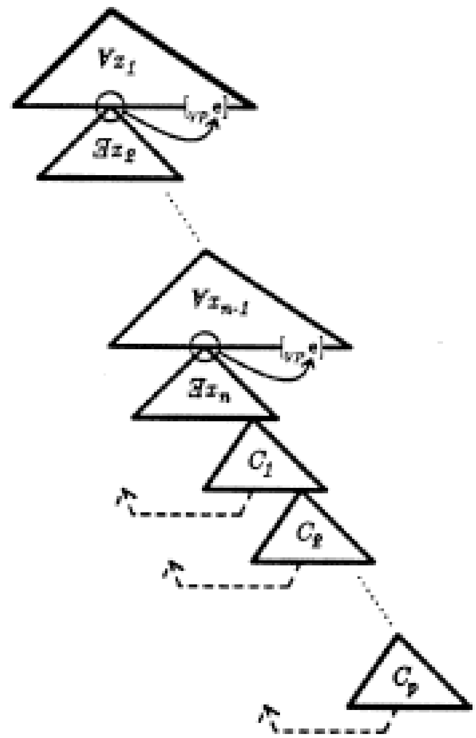
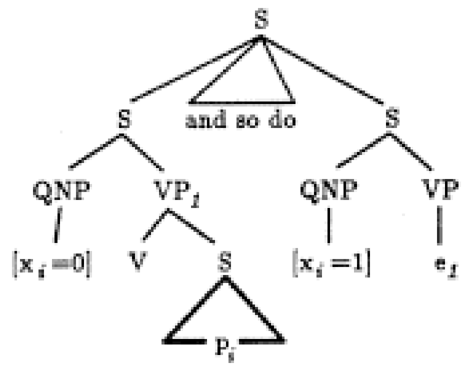


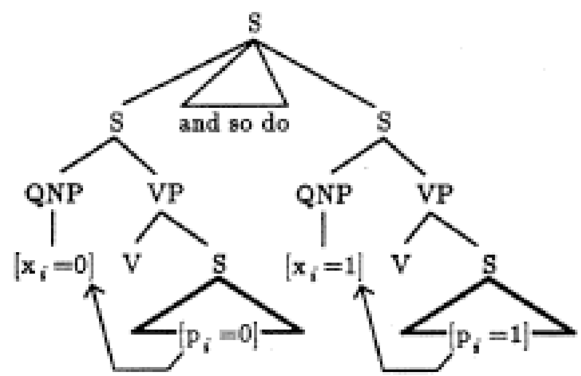
Figure 4.3: The surface form S that corresponds to the input instance $\Omega = \forall x_1 \exists x_2 \ldots \forall x_{n-1} \exists x_n [C_1, C_2, \ldots C_p]$. The quantifier constructions contain two antecedents to represent the two possible truth assignments to the quantified variable. Each universal quantifier $\forall x_i$ is represented by a VP-ellipsis template. In the logical form that corresponds to S, each of the $n/2$ circled overt VPs is copied to its corresponding ellipsed VP position $[_{VP}e]$, according to the copy-and-link theory 4.1. Each existential quantifier $\exists x_{i+1}$ is represented by an extraposed strong crossover template, as discussed in the text. Each clause C_j is represented by a pigeonhole construction that contains three pronouns, one for each literal in C_j; one of these pronouns (the *selected* pronoun) must link to an antecedent outside that construction, in some dominating quantifier construction. These obligatory long distance links are drawn with dashed arrows. The selected pronouns represent the literals that satisfy the clauses.

(46)



According to the copy-and-link theory 4.1, the language user's knowledge of the construction (46) is represented in the abstracted logical form (47). First, the overt VP is copied to the position of the null VP. Next, pronouns inside the original and copied VPs link to their antecedents independently.

(47)

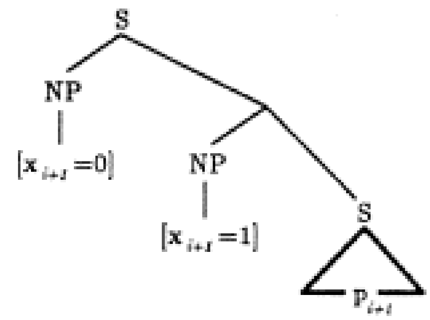


The VP is used by the reduction to represent the Boolean predicate $P(x_i)$; the embedded pronoun p_i represents a true literal of x_i inside P; the two QNP subjects represent the truth values $x_i = 0$ and $x_i = 1$, respectively. Each p_i must link to the subject of its own conjunct in the logical form, because the subjects are quantified noun phrases. Therefore the pronoun p_i in the first VP may only link to the first subject $[_{\text{QNP}}\ x_i = 0]$, which represents the conjunct $P(x_i = 0)$, and the pronoun p_i in the second (copied) VP may only link to the second subject $[_{\text{QNP}}\ x_i = 1]$, which represents the conjunct $P(x_i = 1)$. As shown in figure 4.3 above, the verb phrase will also contain the construction (48) that represents the next quantifier $\exists x_{i+1}$.

The second step is to simulate an existential quantifier. An existentially quantified predicate $[\exists x_{i+1} P(x_{i+1})]$ is true if and only if $[P(x_{i+1} = 0) \vee$

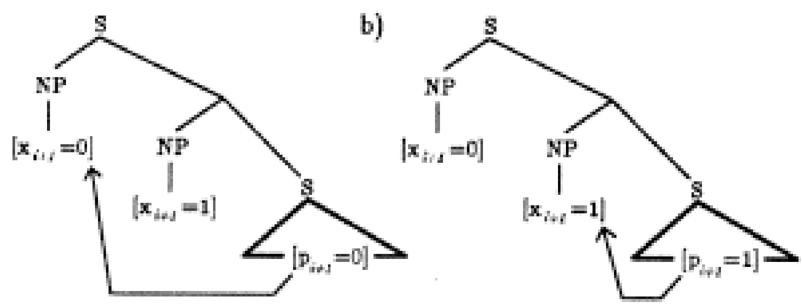
$P(x_{i+1} = 1)]$. The latter Boolean formula can be expressed in a construction whose surface form is (48).

(48)



This structure will have two possible meanings, as represented by the two logical forms in (49):

(49) a)



The embedded sentence represents the predicate $P(x_{i+1})$; the embedded pronoun p_{i+1} represent a true literal of x_{i+1} inside the predicate P; the two NPs represent the truth values $x_{i+1} = 0$ and $x_{i+1} = 1$, respectively. Linguistic constraints to be disclosed below ensure that p_{i+1} can only be linked to one of the two noun phrases, and that p_{i+1} can be linked to the first NP $[_{\text{NP}}\ x_{i+1} = 0]$ if and only if $P(x_{i+1} = 0)$ is true, as shown in (49a); and that p_{i+1} can be linked to the second NP $[_{\text{NP}}\ x_{i+1} = 1]$ if and only if $P(x_{i+1} = 1)$ is true, as shown in (49b). The embedded clause will also contain the construction (46) that represents the next quantifier $\forall x_{i+2}$, as shown in figure 4.3.

In order to ensure consistency of truth assignments, all embedded pronouns representing true literals of x_{i+1} must link to the same antecedent. This constraint may be enforced using the powerful strong crossover configuration introduced in the previous section. The details of how this might be done arose from discussion with Alec Marantz, who suggested all the examples.

Recall that strong crossover is the configuration where an anaphoric element α intervenes between a displaced wh-phrase and its trace, and c-commands the trace. In such a configuration, α obviates the subject of the wh-phrase.

(50) a. Who$_k$ did he$_{*k}$ say Mary kissed t_k.
b. [the man]$_1$ [who$_k$ he$_{*1}$ likes t_k].

The noun phrase in (50b) contains a relative clause [*who he likes t*] that predicates [*the man*]; the pronoun *he* is in a strong crossover configuration, and therefore cannot refer to [*the man*], which is the subject of the relative clause.

Now consider the effect of extraposing a relative clause containing a strong crossover configuration in (51).

(51) a. At the airport, a man$_1$ met Jane$_2$, who$_{k=1/*2}$ she$_2$ likes t_k.
b. At the airport, a man$_1$ met Jane$_2$, who$_{k=*1/2}$ he$_1$ likes t_k.

In (51a), if we understand *she* as referring to *Jane*, then we must understand *who* as predicating *a man*. Conversely, if we understand *he* as referring to *a man* in (51b), then *who* must predicate *Jane*. This simple example establishes the ambiguity of predication when the predicate is an extraposed relative clause containing a strong crossover configuration.

When the extraposed relative clause contains two obviative pronouns, as in (52), then the sentences cannot have the intended interpretation because the relative clause must predicate a subject, yet cannot without violating strong crossover.

(52) a. *At the airport, [a man]$_1$ met Jane$_2$, who$_k$ she$_2$ thinks he$_1$ likes t_k.
b. *At the airport, [a man]$_1$ met Jane$_2$, who$_k$ he$_1$ thinks she$_2$ likes t_k.

This example establishes that the strong crossover configuration gives rise to inviolable obviation between the wh-phrase and all embedded pronouns that c-command its trace.

Now we have our construction:

(53) At the airport, NP_0 met NP_1, [who$_k$ $\ldots \alpha_{*k} \ldots t_k$].

As before, two possible antecedents NP_0 and NP_1 represent the truth assignments $x_{i+1} = 0$ and $x_{i+1} = 1$, respectively. Pronouns in the embedded clause that represent true negative literals of x_{i+1} can only link to the 'false' noun phrase NP_0; pronouns that represent true positive literals of x_{i+1} can only link to the 'true' noun phrase NP_1. Observe that the relative pronoun *who$_k$* may predicate either NP_0 or NP_1 in the example (53). The strong crossover configuration ensures that all anaphoric elements α in the extraposed relative clause obviate the subject of the wh-phrase *who$_k$*. Therefore, once the ambiguous predication relation is determined, pronouns representing literals of x_{i+1} must all be linked to the same antecedent because (i) the pronouns must all obviate the predicated noun phrase by strong crossover and (ii) there is only one other permissible antecedent by construction. This exactly corresponds to assigning a consistent truth value to x_{i+1} everywhere.

The third and final step of the reduction is to simulate a Boolean 3-clause $C_j = (a_j \vee b_j \vee c_j)$ using the pigeonhole principle. A Boolean clause C_j is true if and only if one of its literals it true: let us call the literal that satisfies the clause the *selected literal.* Only selected literals need be assigned consistent truth values: nonselected literals simply don't matter, and can receive any arbitrary inconsistent value, or none at all. We have been reducing the quantification of variables to the binding of pronouns, and so must now represent each literal in C_j with a pronoun. For each 3-clause, the reduction builds a sentence that contains three disjoint pronouns and only two possible antecedents. At least one of the pronouns must be bound by an antecedent outside the sentence—this pronoun represents the selected literal. The following English sentence shows how this works:

(54) [$_S$ [the student] thought [the teacher] said that
[he$_a$ introduced her$_b$ to him$_c$]]

Only two neutral antecedents [*the student*] and [*the teacher*] are locally available to the three obviative pronouns *he$_a$*, *her$_b$*, and *him$_c$* in this construction. Therefore at least one of these three pronouns must be bound outside the sentence, by one of the noun phrases in some dominating quantifier construction (either (46) or (48)). This selected pronoun corresponds to a true literal that satisfies the clause C_j. Agreement features on pronouns and their antecedents ensure that a pronoun representing a literal of x_i can only link to an antecedent representing the quantifier of x_i.

Note that this construction is contained inside $n/2$ VP-deletion constructions in the surface form of the entire sentence S, and that therefore the corresponding logical form will contain $2^{n/2}$ copies of each such construction, each copy with its own selected pronoun. (This corresponds to the fact that different literals may satisfy a given quantified clause, under different quantifier-determined truth assignments.) The verb phrase that appears in our English example (54) as [*he introduced her to him*] will immediately contain the construction representing the next Boolean clause C_{j+1}, as shown in figure 4.3.

The pigeonhole construction representing C_j is permissible iff all of its logical form copies are all permissible, which is only possible when the Boolean clause C_j contains a true literal for any possible quantifier-determined truth assignment to its literals, as represented by the dominating quantifier constructions (either (46) or (48)). Therefore, the logical form for the complete surface form S is permissible iff the quantified Boolean formula Ω is true. $\square$

Note that the constructions used in this proof to represent existential quantifiers (48) and Boolean clauses (54) can be combined to give a third direct NP-hardness proof for the anaphora problem, where each pronoun is no more than four-ways ambiguous and no elliptical contexts are used. Such a proof requires significantly fewer agreement features than used in the proof of lemma 4.1.1.

The epilogue to this proof is a demonstration of how the preceding reduction might concretely represent the QBF formula $\forall x \exists y[(\overline{y} \vee \overline{x} \vee y), (x \vee \overline{x} \vee y)]$ in an actual English sentence.

There are two minor difficulties, that are entirely coincidental to the English language: the English plural pronoun *they* is unspecified for gender; there are no entirely neutral arguments in English, that can be the antecedent of any pronoun. Rather than construct our example in a different language, say Italian, let us make the following allowances. To overcome the first difficulty, let *they*$_0$ be the masculine plural pronoun, and *they*$_1$ the feminine plural pronoun. To overcome the second difficulty, we observe that a plural pronoun can always have a split antecedent, as in example (55), and that the condition of local obviation holds between *they* and *him*. That is, *they* and *him* cannot share an antecedent when they are locally obviative.

(55) John$_1$ suggested to Tom$_2$ that they$_{\{1,2\}}$ tell him$_{*1/*2}$ to leave.

We will use split antecedents below.

The given formula has two variables, x and y, which we will identify via the plural/singular number distinction: plural pronouns represent literals of x, while singular pronouns represent literals of y. Negation will be represented via the the masculine/feminine gender distinction: masculine pronouns for negative literals, feminine pronouns for positive literals. These correspondences are summarized in the table:

$$\begin{array}{r|l} \overline{x} \mapsto \mathit{they}_0 & \overline{y} \mapsto \mathit{he}_0 \\ x \mapsto \mathit{they}_1 & y \mapsto \mathit{she}_1 \end{array}$$

The constructed sentence consists of four components:

- The VP-ellipsis construction (46) to represent $\forall x$:

  (56) [[$_{\mathrm{NP}_0}$ some stewards] [$_{\mathrm{VP}}$ say [$_{\mathrm{S}}$ ...]]]
  and so do [[$_{\mathrm{NP}_1}$ some stewardesses] [$_{\mathrm{VP}}$ e]]

- The extraposed strong crossover configuration (53) to represent $\exists y$:

  (57) [$_{\mathrm{S}}$ at the airport [$_0$ a KGB man] met [$_1$ Jane], [$_{\mathrm{S}}$ who$_k$ [... t_k] and [... t_k]]

- The pigeonhole construction (54) to represent $(\overline{y} \vee \overline{x} \vee y)$ using split antecedents.

  (58) [$_{\mathrm{S}}$ the officer, the agent, and the mechanic suspected
  [he$_0$ expected them$_0$ to talk to her$_1$]]

  There are three locally available antecedents, all singular and unspecified for gender. The three pronouns in the embedded clause are obviative, and require at least four singular antecedents. Therefore, at least one of the pronouns must be linked to an argument outside the construction (58).

- A second pigeonhole construction to represent $(x \vee \overline{x} \vee y)$, again using split antecedents.

  (59) $[_S$ the crew, the pilot, and the co-pilot knew [they$_1$ traded them$_0$ to her$_1$]]

  There are three locally available antecedents: one is plural neuter (*the crew*), and the remaining two are singular neuter. The three pronouns in the embedded clause are obviative, and require at least one plural antecedent and three singular antecedents. Therefore, at least one of the pronouns must be linked to an argument outside the construction (59).

The resulting sentence, in all its glory, is:

(60) $[[_{NP_0}$ some stewards] $[_{VP}$ say
$[_S$ at the airport $[_0$ a KGB man] met $[_1$ Jane], $[_{S'}$ who$_k$
$[_S$ the officer, the agent, and the mechanic suspected
[he$_0$ expected them$_0$ to talk to her$_1$ about t_k]]
and
$[_S$ the crew, the pilot, and the co-pilot knew
[they$_1$ traded them$_0$ to her$_1$ for t_k]]]]]]
and so do
$[[_{NP_1}$ some stewardesses] $[_{VP}$ e]]

This concludes the presentation of the lemma 4.2.1.

4.2.3 Ellipsis reconsidered

In the previous section, we proved that the anaphora problem is PSPACE-hard. The thesis we are defending states that language comprehension is NP-complete. Therefore, the thesis predicts that there is a defect in the linguistic analysis that led to the PSPACE-hardness result. The thesis also tells us exactly where to look for the defect: we must reexamine that part of the analysis that allowed us to simulate a computation outside of $\mathcal{NP}$. In the case of a reduction from QBF, the defect must be in that part of the analysis used to simulate the unnaturally powerful universal quantifier. Therefore, let us reexamine the copy theory 4.1 of ellipsis.

A copy operation naturally makes two predictions; neither holds.

The first prediction is that the original (overt) structure and its copy will obey the same post-copying linguistic constraints, including agreement and the linking conditions. (If agreement and the linking conditions did not apply after copying, then it would always be possible to vacuously satisfy those constraints, simply by postponing all linking until after copying had applied. Therefore, agreement and the linking conditions must apply both before and after copying.) This expected post-copying equivalence is violated. Although overt pronouns must agree with their antecedent on gender and number (61a), copied pronouns can disagree with their antecedents, as in (61b):[6]

(61)
a. Tom$_1$ read his$_{1/*2}$ book and Barbara$_2$ read his$_{1/*2}$ book (too).
b. Tom$_1$ [read his$_1$ book] and Barbara$_2$ did [e] too.
([*read his*$_{1/2}$ *book*])

Moreover, although overt anaphors must have local antecedents in (62a), copied anaphors need not, as shown in (62b):

(62)
The prisoner$_1$ shot himself$_1$ before [the executioner$_2$ could shoot himself$_{*1/2}$].
a. The prisoner$_1$ [shot himself$_1$] before the executioner$_2$ could [e].
([*shoot himself*$_{1/2}$])

The second prediction is that processes that apply after copying, such as linking, will apply independently in both the original (overt) structure and its copy. This expected post-copying independence is also violated. In particular, linking is not independent in both the original structure and its copy, as shown by example (63), which is incorrectly predicted to have five read-

[6] The difficulty of obtaining the covariant interpretation for *Barbara*$_1$ *read her*$_1$ *book and Eric did too*, or for *We*$_1$ *ate our*$_1$ *vegetables and so did Bob*, does not weaken my criticism of the copy theory 4.1. My criticism is based on the necessity of discriminating (61a) and (61b), which the copy theory is unable to do. In order to account for the possible invariant$\sim$covariant contrast between masculine and feminine pronouns, we suggest in appendix B that the thematic-position assigned to some anaphoric elements α will inherit the agreement features of α, and in these cases α must agree with both of its antecedents. An alternate approach, to say that *he* is the "default bound variable," would incorrectly suggest that the covariant interpretation of *she* is never available.

ings (two when linking precedes copying, four when linking follows copying, and one overlap).

(63) Bob [introduced Felix to his neighbors] and so did Max [e].

In particular, there should be a reading where the overt *his* refers to *Felix* and the null/copied *his* refers to *Max*. However, this reading is not available. In fact, only three interpretations of (63) are attested (two invariant, one covariant), as shown in (64):

(64)
a. Bob_1 [introduced Felix_2 to his_2 neighbors] and so did Max_3 [e].
([*introduced* $\textit{Felix}_2$ *to* $\textit{his}_{*1/2/*3}$ *neighbors*])
b. Bob_1 [introduced Felix_2 to his_1 neighbors] and so did Max_3 [e].
([*introduced* $\textit{Felix}_2$ *to* $\textit{his}_{1/*2/3}$ *neighbors*])

In other words, a pronoun must link to the same position in both the visible verb phrase and its understood copy. This is not real copying, but a kind of logical predicate-sharing that can always be represented without copying.

Let us therefore propose the following predicate-sharing theory 4.2 of ellipsis:

Theory 4.2 *The logical form of ellipsis is constructed by sharing the same thematic predicate between the overt and ellipsed structures; obviation is a relation between argument positions in a thematic predicate; an anaphoric element may link to an argument or to an argument position.*

Evidence. First, verbs are thematic functions from their direct and indirect objects to a verb phrase; a verb phrase is function from the inflection and the subject to a logical proposition, that is realized syntactically as a clause. For example, the expression *Felix hates vegetables* would be assigned then logical form (65).

(65) $(\lambda x.[x \text{ hates vegetables}])\ (\lambda P.[(\text{Felix } P)])$

Second, VP-ellipsis is the sharing of one VP predicate between two clauses. One way to represent the logical form of an elliptical structure to lambda-abstract the VP predicate. For example, the surface form (66a) would be assigned the logical form representation (66b):

(66) a. [Felix [ate dinner]] and so did [Tom [e]]

b. (λx.[x ate dinner]) (λP.[(P Felix) and so did (P Tom)])

Third, obviation is a relation between the argument positions in the VP predicate, as illustrated in (67b) for the surface form (67a).

(67) a. Romeo$_1$ wants Rosaline to [love him$_{*1}$] before wanting himself$_1$ to [e].

b. (λx_i.[x_i to love him$_{*i}$])
(λP.[Romeo$_j$ wants [(Rosaline P)] before wanting [(himself$_j$ P)]]) $\Rightarrow$ [$i \neq j$]

This logical form representation accounts for all the cases of invisible obviation, without an unnaturally powerful copying operation.

Fourth, an anaphoric element may link to an argument directly (68b), resulting in the invariant interpretation, or indirectly, to an argument position in the VP predicate (68c), resulting in the covariant interpretation.

(68) a. Felix$_1$ [hates his$_1$ neighbors] and so does Max [e]

b. (λx.[x hates his$_i$ neighbors]) (λP.[(Felix$_i$ P) and (Max P)])

c. (λx_i.[x_i hates his$_i$ dinner]) (λP.[(Felix P) and (Max P)])

□

This predicate-sharing theory 4.2 correctly predicts that the example (63) has exactly three interpretations, one for each of the three possible verbal predicates shown in (69).

(69) a. (λx.[x introduced Felix to Felix's neighbors])
(λP.[(P Bob) and (P Max)])

b. (λx.[x introduced Felix to x's neighbors])
(λP.[(P Bob) and (P Max)])

c. (λx.[x introduced Felix to Bob's neighbors])
(λP.[(P Bob) and (P Max)])

While predicate-sharing is conceptually simple, an extensive investigation is needed to confirm such a linguistic theory. This is the task of appendix B.

The predicate-sharing theory 4.2 gives us the upper bound predicted by the complexity thesis:

Theorem 10 *The anaphora problem is in $\mathcal{NP}$ for nonelliptical structures, and for elliptical structures with predicate-sharing.*

Proof. Covert arguments in a structure are either coreferential with an overt argument in the structure (for example, control PRO or wh-trace), in which case the they may be coalesced with their overt antecedent, or they are assigned an arbitrary interpretation (PRO_{arb}), in which case they do not participate in the graph of referential dependencies and may be ignored entirely. Therefore, the number of obviation relations is at most quadratic in the number of overt arguments, an upper bound that is obtained in the case of a complete obviation graph. The logical forms licensed by the predicate-sharing theory 4.2 are nearly the same size as their corresponding surface forms, because we can always lambda-abstract the shared predicate, if the structure is elliptical. Otherwise, the structure is nonelliptical and logical and surface forms are the same size, because operators that map surface forms to logical forms, such as quantifier scope assignment, do not increase the number of arguments and therefore cannot increase the size of the graph of referential dependencies beyond quadratic. Next, each anaphoric element is nondeterministically linked either to an argument in the set A of available antecedents, or to an open thematic position. Clearly this may be done in nondeterministic polynomial time. Finally, we check that the linking conditions are satisfied, including invisible obviation, in deterministic time proportional to the number of links, verifying the semantics of obviation by propagated "referential value" markers along the links, checking for cyclic dependencies, and so forth. $\square$

Theorem 11 *The anaphora problem is NP-complete.*

Proof. By theorems 10 and 9. $\square$

As we saw above in section 4.2.3, and again in greater detail in appendix B, the predicate-sharing theory is strictly superior to the copy-and-link theory. That is, the predicate-sharing theory assigns better structural descriptions to the class of elliptical utterances than the copy-and-link theory does, and no utterances are assigned better structural descriptions by the copy-and-link theory. However, the significance of the predicate-sharing theory goes beyond merely the number of linguistic examples correctly reclassified.

Recall that our central scientific goal is to understand the comprehension, production, and acquisition of human language; generative theory is interesting only in so far as it advances this goal. The solution to a PSPACE-hard problem may be exponentially large in the size of the problem statement. (Unlike problem in $\mathcal{NP}$, PSPACE-hard problems do not have efficient witnesses. An efficient witness is a short correctness proof for a solution. In the case of the anaphora problem, a permissible graph of referential dependencies serves as the correctness proof.) If anaphora comprehension were PSPACE-hard, as it is according to the copy-and-link theory of ellipsis, then the mental representations required to produce and comprehend elliptical anaphora would be infeasibly large. Language users could not even comprehend the utterances that they themselves produced. And the generative theory of anaphora would not yield a plausible account of language comprehension and production.

But by reducing the complexity of anaphora from PSPACE to $\mathcal{NP}$, we prove that the anaphora problem has efficient witnesses, and in turn show that generative theory is the basis of a plausible account of anaphora comprehension and production.

4.3 Analysis of linguistic theories

It is informative to contrast the approach of this report, the direct analysis of human language, with a related approach, the analysis of linguistic theories. In the latter approach, we study the theory-internal computational problems posed by the theory, such as "compute this predicate defined by the theory," or "ensure that this constraint of the theory is satisfied." Chapter 2 has examined the computational structure of generative phonological theories in some detail. This approach is also exemplified by Giorgi, Pianesi, and Satta (1989) in their complexity investigation of the binding theory of Chomsky (1986).

The central danger of such an approach is the risk of irrelevance, which increases whenever we lose sight of the computational problems of language comprehension, production, and acquisition. Different theories talk about vastly different things, and hence it is impossible to achieve any kind of invariance with respect to either phenomena or theory. Moreover, the computational properties of even the perfect linguistic theory have at best an indirect connection to the computational properties of human language.

The central computational problem posed by all generative linguistic theories, of which all theory-internal problems are subproblems, is to enumerate all and only the possible structural descriptions (that is, possible linguistic representations). That is, linguistic theory itself poses a computational problem, the problem of constructing the ith representation in the enumeration, given the index of enumeration i. (Equivalently, we may think of i as the encoding of a possible linguistic representation, that must be verified by the linguistic theory as being permissible.) As elaborated in chapter 2, the computational problems of enumerating or verifying representations have at best an indirect connection to human language, which is the process of constructing structural descriptions of evidence.

Even worse, complexity analyses of linguistic theories are likely to be irrelevant. For example, consider the many different theories of referential dependencies. They are stated in vastly different terms: as constraints on syntactic or discourse-level representations, in terms of the goals and intentions of speakers and hearers, or even in terms of the objective "meaning" of utterances in relation to the external world. Let us examine three very similar theories of anaphora, that nonetheless have widely divergent complexities.

One approach requires all referentially-dependent elements, including pronouns, to have linguistic antecedents if at all possible (Higginbotham 1983, or at least its spirit). The theory-internal computational problem posed by such a theory is to link every pronoun to an argument, subject to conditions of obviation and acyclicity. As proven in this chapter, the decision problem posed by this approach is NP-complete.

A second approach postulates a condition of obviation combined with free indexing of all arguments (Chomsky 1986). The corresponding theory-internal problems posed are (i) to decide if a given indexing is permissible (verification) and (ii) to ensure that pronouns may indexed without violating the obviation condition (satisfaction). The verification problem is clearly easy, requiring us to simply compute the obviation relations from the phrase structure and check that each pronoun is assigned a different index than any of the arguments that it obviates. Because obviation may always be satisfied by trivially assigning a different index to every pronoun, the decision problem for satisfaction requires constant time, that is, always answer YES.

A third approach only handles cases of so-called bound anaphora, where a pronoun is interpreted as a variable bound by a quantifier, as in [*every man*] *ate his dinner* (Reinhart 1983). The theory-internal verification problem posed is to ensure that every pronoun interpreted as a bound variable is c-commanded by a natural quantifier. The problem of checking an existing structure is efficient, requiring time proportional to the size of the structure. However, even when every pronoun is required to have a linguistic antecedent, no pronoun need ever be interpreted as a bound variable, and hence the corresponding decision problem for satisfying Reinhart's theory only requires constant time.

But the theory-internal problems corresponding to the second two approaches are of no independent interest, being entirely irrelevant to human language. In studying the computational structure of human language, the only relevant problems are language comprehension, production, and acquisition. The computational problem posed by pronouns in the act of comprehension is to compute their antecedents using no new information. If language user fails to do this, then he has failed to comprehend either the pronouns or the utterance that contains them. The fact that a language user can fail to comprehend an utterance in constant time by assigning it an inadequate or incomplete representation is of no interest.

Even if our sole interest is in the computational structure of linguistic theories, then we should still study the complexity of LC problems. Studying LC problems allows us to more easily compare linguistic theories, and to study the complex interactions among the different parts of a linguistic theory. Either a particular LC problem is posed by a particular linguistic theory, or it is not. If it is not posed by the theory, then that theory is empirically inadequate, and we understand exactly why and how the theory is inadequate. Otherwise the LC problem is posed by the theory, and no matter how it is posed by the particular theory—no matter how it is disguised or carved up into different parts of the theory, whether in phonology, syntax, discourse, pragmatics, semantics, or what have you—then that linguistic theory inherits the computational complexity and structure of the LC problem. This is because complexity theory classifies *problems*, not algorithms or particular ways of solving those problems. As long as a linguistic theory poses an LC problem, the problem of assigning representations to utterances according to that theory is at least as complex as the LC problem is.

This is well-illustrated by the anaphora problem studied in this chapter. As long as a particular theory of language has an empirically adequate description of the observed facts of obviation and antecedence, then the comprehension problem for that linguistic theory inherits the structure of the anaphora problem examined here. This is true no matter how this description is couched, whether in terms of constraints on a syntactic relation of coindexing or linking, in terms of syntax or discourse, in terms of speaker-hearer intentions or other pragmatic considerations, or even in terms of a Montague-like compositional theory of semantic types. If the theory provides an empirically adequate description of the language user's knowledge of utterances, then it will inherit the inalienable computational structure of that knowledge.

Chapter 5

References

Baker, Mark. 1985. *Incorporation: a theory of grammatical function changing*. Ph.D dissertation, MIT Department of Linguistics and Philosophy. (Published by University of Chicago Press, 1987.)

Barton, Ed, Robert Berwick, and Eric Ristad. 1987. *Computational complexity and natural language*. Cambridge: MIT Press.

Beattie, James. 1788. *The Theory of Language*. Menston, England: The Scolar Press Limited. (English Linguistics 1500-1800 Collection, No.88.)

Berwick, Robert, and Amy Weinberg. 1984. *The grammatical basis of linguistic performance: language use and acquisition*. Cambridge: MIT Press.

Browne, Thomas Gunter. 1795. *Hermes Unmasked*. Menston, England: The Scolar Press Limited. (English Linguistics 1500-1800 Collection, No.201.)

Burzio, Luigi. 1988. "On the non-existence of disjoint reference principles." Colloquim, LSA Meeting, December 1988.

Burzio, Luigi. 1989. "The role of the antecedent in anaphoric relations." Second Princeton Workshop on Comparative Grammar, Princeton University, April 27–29.

Chandra, A., D. Kozen, and L. Stockmeyer. 1981. "Alternation." *J. ACM* 28(1):114–133.

Canny, John F. 1988. *The complexity of robot motion planning.* Cambridge: MIT Press.

Chiat, Shulamuth. 1986. "Children's pronouns." in U. Wiesemann (ed.), 1986, pp.381–404.

Chomsky, Noam. 1951. *Morphophonemics of modern Hebrew.* M.A. dissertation, University of Pennsylvania. (Published with revisions, Garland Publishing, New York, 1979.)

Chomsky, Noam. 1955. *The logical structure of linguistic theory.* mimeographed, Harvard. (Published in part by Plenum Press, 1975).

Chomsky, Noam. 1956. "Three models for the description of language." *I.R.E. Transations on Information Theory*, Vol. IT-2, pp. 113–124. Reprinted, with corrections, in R.D. Luce, R. Bush, and E. Galanter (eds.), *Readings in Mathematical Psychology*, Vol.II. New York: Wiley, 1965.

Chomsky, Noam. 1965. *Aspects of the theory of syntax.* Cambridge: MIT Press.

Chomsky, Noam. 1968. *Language and mind.* New York: Harcourt Brace Jovanovich.

Chomsky, Noam. 1980. "On binding." *Linguistic Inquiry* 11(1):1–46.

Chomsky, Noam. 1980. *Rules and representations.* New York: Columbia University Press.

Chomsky, Noam. 1981. *Lectures on government and binding.* Dordrecht: Foris Publications.

Chomsky, Noam. 1982. *Some concepts and consequences of the theory of government and binding.* Cambridge, MA: MIT Press.

Chomsky, Noam. 1986. *Barriers.* Cambridge: MIT Press.

Chomsky, Noam. 1986. *Knowledge of language: its origins, nature, and use.* New York: Praeger Publishers.

Chomsky, Noam. 1989. "Some notes on economy of derivation and representation." MIT Working Papers in Linguistics, vol.10, pp. 43–74.

Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English.* New York: Harper & Row.

Clements, George and S. Jay Keyser. 1983. *CV Phonology: a generative theory of the syllable.* Cambridge: MIT Press.

Cordemoy, Géraud de. 1667. *A Philosphical Discourse Concerning Speech, Conformable to the Cartesian Principles.*

Curry, Haskell B. 1961. "Some logical aspects of grammatical structure." *Proceedings of Symposia in Applied Mathematics*, volume XII, pp. 56–68.

Dahl, Osten. 1972. "On so-called 'sloppy identity'." *Gothenburg Papers in Theoretical Linguistics* 11, University of Göteberg, Sweden.

Dahl, Osten. 1974. "How to open a sentence: abstraction in natural language." *Logical Grammar Reports* 12, University of Göteberg, Sweden.

Finer, Daniel. 1984. *The formal grammar of switch-reference.* Unpublished Ph.D dissertation, University of Massachusetts at Amherst.

Fodor, Jerry. 1983. *The modularity of mind.* Cambridge: MIT Press.

Fukui, Naoki. 1986. *A theory of category projection and its applications.* Unpublished Ph.D dissertation, MIT Department of Linguistics and Philosophy.

Gazdar, Gerald, Ewan Klein, Geoffry Pullum, and Ivan Sag. 1985. *Generalized phrase structure grammar.* Oxford, England: Basil Blackwell.

Giorgi, Alessandra, Fabio Pianesi, and Giorgio Satta. 1989. "The computational complexity of binding theory's satisfiability and verification problems." Trento, Italy: unpublished manuscript, IRST.

Haik, Isabelle. 1987. "Bound VPs that need to be." *Linguistics and Philosophy* 10:503–530.

Hale, Kenneth. 1982. "The logic of Damin kinship terminology." in Oceania Linguistic Monograph no.24. *The Languages of Kinship in Aboriginal Australia*, ed. by Jeffrey Heath, Francesca Merlan, and Alan Rumsey.

Halle, Morris. 1961. "On the role of simplicity in linguistic descriptions." *Proceedings of Symposia in Applied Mathematics*, volume XII, Structure of Language and its Mathematical Aspects, pp. 89–94.

Halle, Morris. 1962. "Phonology in generative grammar." *Word* 18(1–2):54–72.

Halle, Morris. 1985. "Speculations about the representation of words in memory." In *Phonetic Linguistics, Essays in Honor of Peter Ladefoged*, V. Fromkin, ed. New York: Academic Press.

Halle, Morris and G.N. Clements. 1983. *Problem Book in Phonology.* Cambridge: MIT Press.

Herder, Johann Gottfried von. 1827. *Treatise upon the origin of language*. London: Camberwell Press.

Higginbotham, James. 1983. "Logical form, binding, and nominals." *Linguistic Inquiry* 14:395–419.

Higginbotham, James. 1985. "On semantics." *Linguistic Inquiry* 16:547–593.

Higginbotham, James. 1989. "Eludications of meaning." *Linguistics and Philosophy* 12(3):465–517.

Higginbotham, James. 1990. "Reference and Control." ms., Department of Linguistics, MIT.

Hopcroft, John and Jeffrey Ullman. 1979. *Introduction to Automata Theory, Languages, and Computation*. Reading, MA: Addison-Wesley.

Jakobson, Roman and Morris Halle. 1956. *Fundamentals of Language*. The Hague: Mouton.

Johnson, C. Douglas. 1972. *Formal aspects of phonological description*. The Hague: Mouton.

Kayne, Richard. 1981. "Unambiguous paths." In *Levels of syntactic representation*, R. May and J. Koster, eds. Dordrecht: Foris Publications, pp. 143–183.

Kayne, Richard. 1984. *Connectedness and binary branching*. Dordrecht: Foris Publications.

Keenan, Edward. 1971. "Names, quantifiers, and sloppy identity problem." *Papers in Linguistics* 4(2):211–232.

Kenstowicz, Michael and Charles Kisseberth. 1979. *Generative Phonology*. New York: Academic Press.

Kitagawa, Yoshi. 1986. *Subjects in Japanese and English*. Unpublished Ph.D dissertation, Department of Linguistics, University of Massachusetts at Amherst.

Kitagawa, Yoshi. 1989. "Copying identity." ms., Department of Linguistics, University of Massachusetts at Amherst.

Koopman, H. and D. Sportiche. 1985. "Theta theory and extraction." in *GLOW newsletter* 14:57–58.

Koopman, H. and D. Sportiche. 1986. "A note on long extraction in Vata and the ECP." *Natural Language and Linguistic Theory*, 4(3):357–374.

Koster, Jan. 1987. *Domains and dynasties: the radical autonomy of syntax.* Dordrecht: Foris Publications.

Kuroda, S.-Y. 1988. "Whether we agree or not: a comparative syntax of English and Japanese." *Linguisticae Investigationes*, 12:1–47.

Lambek, Joachim. 1961. "On the calculus of syntactic types." *Proceedings of Symposia in Applied Mathematics*, volume XII, pp. 166–178.

Larson, Richard. 1988. "On the double object construction." *Linguistic Inquiry* 19(3):335–391.

Lasnik, Howard. 1976. "Remarks on coreference." *Linguistic Analysis* 2(1):1–22.

Lasnik, Howard. 1981. "Treatment of disjoint reference." *Journal of Linguistic Research* 1(4):48–58.

Lasnik, Howard. 1989. "On the necessity of binding conditions." in H. Lasnik, *Essays on Anaphora*, Dordrecht: Kluwer Academic.

Lasnik, Howard, and Mamoro Saito. 1984. "On the nature of proper government." *Linguistic Inquiry* 15:235–289.

Lawler, E.L. 1976. "A Note on the complexity of the chromatic number problem." *Information Processing Letters* 5(3), 66–67.

Levin, Beth and Malka Rappaport. 1989. "-er nominals: implications for the theory of argument-structure." in E. Wehrli and T. Stowell, eds., *Syntax and the lexicon*, New York: Academic Press.

Levin, Leonid. 1973. "Universal sequential search problems." *Problems in Information Transmission* 9:265–266.

Luria, A. R. 1968. *The Mind of a Mnemonist: a little book about a vast memory.* New York: Basic Books.

Mahajan, Anoop Kumar. 1989. "Agreement and agreement phrases." MIT Working Papers in Linguistics, vol.10, pp. 217–252.

Marantz, Alec. 1984. *On the nature of grammatical relations.* Cambridge: MIT Press.

Marantz, Alec. 1989. "Projection vs. percolation in the syntax of synthetic compounds." ms., Department of Linguistics, UNC Chapel Hill.

Marantz, Alec. 1990. "Asymmetries in double object constructions." talk at MIT.

Marr, David. 1980. *Vision.* San Francisco: W.H. Freeman.

May, Robert. 1985 *Logical Form: its structure and derivation.* Cambridge: MIT Press.

McCarthy, John. 1981. "A prosodic theory of nonconcatenative morphology." *Linguistic Inquiry* 12:373–418.

McCawley, James D. 1967. "Meaning and the description of languages." appears in *Grammar and Meaning*, Tokyo: Taishukan Publishing Co., 1973.

Minsky, Marvin. 1969. *Computation: finite and infinite machines.* Englewood Cliffs: Prentice Hall.

Partee, Barbara H. 1975. "Montague grammar and transformational grammar." *Linguistic Inquiry* 6(2):203–300.

Pesetsky, David M. 1982. *Paths and categories.* Unpublished Ph.D dissertation, MIT Department of Linguistics and Philosophy.

Peters, Stanley, and R.W. Ritchie. 1973. "On the generative power of transformational grammars." *Information Science* 6:49–83.

Piera, Carlos. 1985. "Gaps in gaps in GPSG." *Linguistic Inquiry* 16:681–683.

Plátek, Martin and Petr Sgall. 1978. "A scale of context sensitive languages: applications to natural language." *Information and Control* 38(1):1–20.

Pollock, Jean-Yves. 1989. "Verb movement, universal grammar, and the structure of IP." *Linguistic Inquiry* 20(3):365–424.

Pritchett, Bradley L. 1988. "Garden path phenomena and the grammatical basis of language processing." *Language* 64(3):539–576.

Reif, John H. 1979. "Complexity of the mover's problem and generalizations." *Proceedings of the 20th Annual Symposium on Foundations of Computer Science*, 421–427. New York: IEEE Computer Society.

Reif, John H. and Micha Sharir. 1985. "Motion planning in the presence of moving obstacles." *Proceedings of the 26th Annual Symposium on Foundations of Computer Science,* 144–154. New York: IEEE Computer Society.

Reinhart, Tanya. 1983. *Anaphora and semantic interpretation.* Chicago: The University of Chicago Press.

Rissanen, Jorma. 1978. "Modeling by shortest data description." *Automatica*, 14:465–471.

Ristad, Eric. 1986. *Complexity of Linguistic Models: a computational analysis and reconstruction of generalized phrase structure grammar.* S.M. dissertation, MIT Artificial Intelligence Lab. (Portions published in Barton, Berwick, and Ristad 1987.)

Ristad, Eric. 1986. "Computational complexity of current GPSG theory." *Proceedings of the 24th Annual Meeting of the ACL*, New York, NY, pp.30–39, June 10–13, 1986.

Ristad, Eric. 1988. "Complexity of Human Language Comprehension." MIT Artificial Intelligence Lab, memo 964.

Ristad, Eric and Robert Berwick. 1989. "Computational consequences of agreement and ambiguity in natural languages." *Journal of Mathematical Psychology* 33(4):379–396.

Ross, John Robert. 1967. *Constraints on variables in syntax.* Ph.D dissertation, MIT Department of Linguistics and Philosophy. (Published in 1986 as *Infinite Syntax!*, Norwood NJ: Ablex.)

Rounds, William. 1975. "A grammatical characterization of the exponential time languages." *Proceedings of the 16th Annual Symposium on Switching Theory and Automata.* New York: IEEE Computer Society, pp. 135–143

Sag, Ivan Andrew. 1976. *Deletion and Logical Form.* Unpublished Ph.D dissertation, MIT Department of Linguistics and Philosophy.

Sagey, Elizabeth. 1986. *The Representation of Features and Relations in Non-linear Phonology.* Unpublished Ph.D dissertation, MIT Department of Linguistics and Philosophy.

Sells, Peter. 1987. "Aspects of logophoricity." *Linguistic Inquiry* 18(3):445–479.

Sportiche, Dominique. 1986. "A theory of floating quantifiers and consequences." Proceedings of NELS 17.

Sportiche, Dominique. 1988. "A theory of floating quantifiers and its corollaries for constituent structure." *Linguistic Inquiry* 19(3):425–449.

Stabler, Edward P., Jr. 1983. "How are grammars represented?" *The Behavioral and Brain Sciences* 6:391–421.

Stowell, Timothy. 1981. *Origins of phrase structure.* Unpublished Ph.D dissertation, MIT Department of Linguistics and Philosophy.

Torrego, Esther. 1984. "On inversion in Spanish and some of its effects." *Linguistic Inquiry* 15:103–129.

Turing, Alan. 1936. "On computable numbers with an application of the Entscheidungs-problem" *Proc. London Math. Soc.* 42(2):230–265.

von Humboldt, Wilhelm. 1836. *Linguistic Variability and Intellectual Development.* Translated by George C. Buck and Frithjof A. Raven, 1971, Miami Linguistics Series No.9, Coral Gables, Florida: University of Miami Press.

Wasow, Thomas. 1972. "Anaphoric relations in English." Ph.D dissertation, MIT Department of Linguistics and Philosophy. (Published in 1979 as *Anaphora in Generative Grammar*, Ghent: E. Story–Scientia.)

Wax, Mati and Kailath, Thomas. 1985. "Decentralized processing in sensor arrays." *IEEE Trans. ASSP* 33(4):1123–1129.

Wiesemann, Ursula, ed. 1986. *Pronominal systems.* Tübingen: Gunter Narr Verlag.

Williams, Edwin. 1977. "Discourse and logical form." *Linguistic Inquiry* 8:101–139.

Williams, Edwin. 1978. "Across-the-board rule application." *Linguistic Inquiry* 9(1):31–43.

Williams, Edwin. 1980. "Predication." *Linguistic Inquiry* 11(1):203–238.

Williams, Edwin. 1989. "The anaphoric nature of θ-roles." *Linguistic Inquiry* 20(3):425–456.

Yao, Andrew. 1988. "Computational information theory." in *Complexity in Information Theory*, Yaser Abu-Mostafa, ed., New York: Springer-Verlag, pp. 1–15.

Appendix A

Philosophical Issues

In this chapter, we examine two philosophical issues arising from the research described in the body of the report. First, we consider the implications of an NP-completeness thesis for human knowledge. Next, we discuss the idealizations to unbounded inputs and unbounded distinctions, which have played an important role in the complexity analyses.

A.1 Implications of a complexity classification

What are the implications of placing human language in the abstract hierarchy of computational complexity? And what is the exact significance of proving $\mathcal{NP}$ lower bounds on language comprehension/production? The consequences of our thesis, that language is NP-complete, are both practical and theoretical.

The central theoretical consequence of the NP-completeness thesis is to offer a broad new (and very different) perspective on human language, where things previously obscured now become clear. By placing language in the much-studied complexity hierarchy, we better understand it's overall computational structure, by analogy to the other equivalent combinatorial problems in its complexity class. We see that language computations are not like two-person adversary games (PSPACE), nor are they like pointer-following (LSPACE) or directed search in a feasible space of possibilities ($\mathcal{P}$). Rather, human language is like blind search in an exponentially large space ($\mathcal{NP}$), to find efficient witnesses.

An second (indirect) theoretical consequence of the thesis has been to refine the linguistic theory. In order to carry out a direct complexity analysis, we must present strong empirical arguments about the facts of linguistic knowledge. Each direct analysis in this report has improved on current understanding. In chapter 2, repeated complexity analyses led us to reorganize the overgeneral architecture of segmental phonology, expose (for the first time) the unnatural rule interactions allowed in segmental model, reveal the importance that the methodological directive "omit directly-predictable information" plays in phonological processes, and increase our understanding of the SPE evaluation metric. In chapter 4, we falsified the standard nondistinctness theory of anaphoric agreement and elucidated the grammar of referential dependencies. Next we discovered the phenomena of invisible obviation, demonstrated the necessity of revising the binding conditions accordingly, falsified the widely-accepted copy-and-link theory of syntactic ellipsis, and proposed an alternative predicate-sharing theory that is strictly empirically superior (as shown in appendix B). Each discovery arose naturally out of the complexity investigation.

We may also ask, what are the real-world implications of the $\mathcal{NP}$ lower bounds for natural language parsers and for theories of language processing?

A parser is an algorithm that assigns structural descriptions to linguistic expressions, according to a linguistic theory that is typically represented as a grammar. An expression is a string of abstract symbols, typically words and occasionally morphemes. A parser is *correct* if it assigns to every input expression exactly the structural descriptions that the generative linguistic theory does. Given our current understanding of nondeterminism, the NP-hardness of language comprehension means that correct natural language parsers require an exponentially-increasing amount of time to parse expressions of linearly-increasing length. In short, correct parsers must be intractable. (This empirical consequence is nothing new; the intractability of existing parsers is well-documented.)

A theory of language processing is an explicit computational model of the language user, that attempts to explain (or at least describe) the comprehension, production, and acquisition of languages. Sometimes such a theory is called a performance theory, or a theory of sentence processing. *The complexity analyses in this report demonstrate conclusively that the relation between competence and performance, between a generative theory and a theory of processing, is not one of limited ability.* This fact is contrary to the frequently expressed and widely held beliefs of linguists, psycholinguists, and computational linguists. If linguistic performance was the limited ability to use linguistic competence, then two consequences would accrue. The first is that language users would have difficulty processing those utterances that are truly computationally difficult. The second is that language users should not have difficulty processing computationally trivial utterances.

There are infinite classes of sentences that may be easily parsed (that is, assigned correct structural descriptions by a simple and very fast algorithm), yet these sentences are extremely difficult for language users to process. One such class is those sentences with trivial obviation graphs, such as complete or edge-free graphs. Computing the referential dependencies for such utterances is trivial, yet language users cannot do it. (In fact, they appear to have difficulty processing utterances with multiple antecedents, regardless of the obviation relations involved.) A second instance is garden path sentences, such as *the horse raced past the barn fell*, which are quickly parsed by simple algorithms, yet seem extremely difficult for language users to process.

There are also infinite classes of complex sentences that cannot be efficiently parsed by any known algorithm, yet these sentences are processed effortlessly by human language users. One such class is sentences containing many local ambiguities, such as lexically ambiguous words. Most sentences are in this class, and no language user has any difficulty processing them. However, such sentences bring current parsers to their knees, because it is not known how to correctly resolve lexical ambiguities locally, without building a complete structural description and therefore being forced to examine an exponential number of possible parses. A second instance is utterances understood as containing phonologically covert elements (so-called empty categories, such as as traces or PRO). Detecting empty categories and computing their antecedents is extremely difficult for parsing algorithms, but effortless for humans.

It is not at all surprising that attempts to explain so-called "performance limitations" as resource-bounded competence have all failed. One fixed resource bound is never subtle enough to capture the diverse range of observed empirical facts. In order to have explanatory force, a small number of resource bounds must be postulated to explain a large number of seemingly unrelated performance limitations. (To postulate a different resource bound for every construction is merely to restate the performance facts.) To my knowledge, no one been able to explain a truly diverse range of performance facts—say from the phonology and syntax, or involving both referential dependencies and phrase structure—using one resource bound, although many have tried. Nor has anyone successfully described even a similar set of performance facts using one resource bound. This may be seen in the work of Miller and Chomsky (1963), who attempted to calculate a numerical bound on the depth of acceptable recursive phrase structure embeddings. However their work only served to demonstrate that no fixed bound could be found for the few constructions they examined, even in in the limited domain of phrase structure computations. A second example comes from the numerous failed attempts to explain garden pathing as the inability to properly resolve a local ambiguity in phrase structure attachment. The central difficulty in such an endeavor is to explain why some types of local ambiguity exhaust resources, while others don't, and why global ambiguities (which should always be more costly) do not. It is not known how to resolve such contradictions.

Even worse, an account in terms of resource limitations has never been plausibly motivated, that is, shown remotely relevant to human language processing. A theory of resource utilization makes exactly one fundamental prediction: that the resource-consuming process must for some input at some critical point exhaust the available resources, at which point the process will crash. Those inputs that exceed the critical point will be rejected, even though they are very similar to other inputs that do not exceed the critical point. In order to demonstrate the plausibility of an explanation in terms of resource limits, someone must exhibit examples on both sides of such a critical point. This has yet to be done.

Nor can performance limitations be explained as errors in competence. The language device cannot be said to make systematic or pervasive errors, because such errors can exist only with respect to a designer's intentions or goals, and the language device was not designed. In short, systematic "errors" cannot be errors in performance, only empirical inadequacies of a particular competence theory. A real performance error, then, must be intermittent and unexpected. And if such errors are not to be accounted for by the competence theory, as is widely-assumed, then they cannot logically constitute evidence for or against the competence theory. This is exactly the *a priori* segregation of evidence into relevant/irrelevant that Chomsky (1986) has so powerfully argued against. Empirical evidence for or against a scientific theory might in principle be found anywhere. Linguistics is no different; the one scientific theory of human language must explain performance errors, because such errors are relevant evidence for the theory. Language errors cannot have their own scientific theory. To see this, consider Becker (1979), who shows how the independent tiers of the autosegmental model can explain facts about speech errors such as that "when vowels or syllables or parts of syllables or whole words are substituted or transposed, there is no change in the stress contour of the sentences." (Franklin, 1971:42)

What is the relation between a generative theory of linguistic knowledge and a constructive theory of language, that explains comprehension, production and acquisition? It seems to me that a constructive theory will result from the generative theory under the information-theoretic interpretation (outlined in chapter 1), refined by increasingly subtle principles and limited by the current state of acquisition. Empirical facts thought of today as performance limitations will be explained tomorrow either as interactions among the refined linguistic principles, or as the incomplete acquisition of linguistic knowledge.

Some so-called performance limitations will be understood as the interaction of increasingly subtle linguistic principles. One such an account is due to Pritchett (1988), who explained garden path constructions in terms of invariants in the computation of thematic structure. A second instance is due to Idsardi (1989), who accounted for a range of classical performance limitations (PP attachment ambiguities, garden paths, and multiple center embeddings) in terms of a linguistic constraint on the mapping between a syntactic relation (government) and a phonological structure (the intonation phrase).

A second (more powerful) class of explanations may be obtained by thinking of the generative theory as the theory of acquirable structural descriptions, and performance limitations as temporary, accidental limits in the current state of acquisition. On this view, language users learn how to pair utterances with their permissible structural descriptions only after repeated exposure to the relevant evidence. Let us consider some examples. The naive language user does not easily recognize the ambiguities inherent in many utterances, such as ambiguities in lexical choice or quantifier scope, but once these ambiguities are pointed out and successfully acquired, they are effortlessly detected and produced in novel utterances. Other examples come from constructions on the frontiers of linguistics research, such as parasitic gaps, strong crossover, and ellipsis. Language users have great difficulty comprehending these constructions on their initial exposure. After repeated exposure, however, these constructions are easily comprehended. We would say that knowledge of the binding conditions is innate, but that the language user must acquire anaphoric morphemes and learn how to compute antecedence and obviation in particular structural configurations, such as strong crossover or ellipsis. A third class of examples, such as center embedding and garden paths, comes from psycholinguistics. Although it is seldom discussed, the most striking fact about these constructions is that after sufficient practice, the language user no longer has difficulty processing them.

A.2 Unbounded idealizations

A central assumption in this work has been the idealization to an unbounded number of input instances, computational resources, and linguistic distinctions. These 'unbounded idealizations,' from a finite set of finite objects to an infinite set of finite objects, are as central to linguistics as they are to computer science. Generative linguistic theory and theoretical computer science make the same idealizations to unbounded inputs and computational resources because they result in the best explanations and empirical predictions.

The first idealization, from a necessarily finite set of inputs to an abstract infinity of inputs, results in better linguistic theories. Finite sets may be characterized by simply listing their elements or by bounding a finite characterization of some infinite superset. The latter idealization to an infinite set gives us a simpler, more predictive, and more interesting characterization than any simple listing could. The idealization to unbounded inputs gives us potent insights because it necessitates a finite characterization of an infinite set, which is only possible if we have discovered significant structure in that set.

The second idealization, from a class of computations that each uses a finite amount of time and space to infinite computational resources, is central to computer science: "To properly capture the notion of a computation we need a potentially infinite memory, even though each computer installation is finite." (Hopcroft and Ullman, 1979:14) In linguistics, Chomsky (1956) and others have convincingly argued that human language is not a finite state system, despite the empirical fact that language users have only finite capabilities. Although every linguistic computation only uses a finite amount of resources, viewing human language as a finite state system—as a computation whose available resources are bounded by a fixed constant—does not give us the most explanatory linguistic theory.

In general, we make idealizations to simplify the impossibly complex real world, and therefore idealizations are never subject to direct empirical confirmation. An idealization is justified only if it results in the best scientific theory, with the best explanations and predictions, not if it is true or not.

Consider the Newtonian idealization to point masses. Clearly, it is empirically false: there has never been a point mass, nor will there ever be. However this point-mass idealization is useful because it simplifies the computation of interactions among massy objects without distorting the outcome of that computation. However, when two objects are very close, or when the computations are very sensitive, then the point-mass idealization breaks down, and must therefore be abandoned. The sole justification of an idealization is its utility: arguments about the *a priori* plausibility of an idealization, although perhaps persuasive, are not ultimately relevant.

Unbounded idealizations are no different. In this finite world, there will never be an infinity of anything. However, the idealization to an infinite set of finite objects (an *unbounded idealization*, hereafter) can be an extremely useful simplification of a finite set of finite objects whose size can vary. An unbounded idealization is especially useful when the infinite set is bounded by an order-of-growth function in some natural parameter. For example, in order to restrict the amount of resources used by any given computation while preserving the idealization to infinite resources, we can bound resources by a function $f(n)$ in the input length n. Thus, although a $f(n)$-resource bounded computation in principle has access to an infinite amount of computational resources, it may use no more than $f(n)$ units of a given resource on any actual length-n input. Crucially, in an unbounded idealization, although objects can be arbitrarily large, each object is finite.

The idealization to an unbounded number of linguistic features is no different from any other unbounded idealization. Features are a method of representing significant distinctions, where each feature represents an independent dimension wherein elements can differ. (The relevant parameter is the number of significant distinctions, not the number of features.) The unbounded-feature idealization does not claim that a language user is capable of making an infinite number of linguistically-significant distinctions.

Rather, it claims that language users are best seen as being capable of making any finite number of distinctions because the number of empirically-observed distinctions is quite large and varies from language to language, and even from language user to language user. In fact, linguistic features are intuitively equivalent to computational space, and therefore the unbounded feature idealization is properly included in linguistic theory's uncontroversial idealization to infinite computational resources.

The goal of a complexity analysis is to characterize the amount of time and space needed to solve a given problem in terms of all computationally relevant inputs. Therefore, the unbounded-feature idealization is justified on complexity-theoretic grounds if the number of linguistic features affects the complexity of a linguistic processes such as language comprehension. The proofs in this report conclusively establish that the number of significant distinctions is a significant parameter of the complexity of a linguistic process, and therefore the idealization is justified in the framework of complexity theory.

A central goal of linguistics is to characterize the productive portions of our linguistic abilities. Therefore, the unbounded-feature idealization is justified on linguistic grounds if the number of linguistically-relevant distinctions is productive. A set of linguistic objects, such as the set of lexical entries, is productive if the set is uniform, variable, and large. By *uniform*, I mean that linguistic process are not sensitive to the exact size of the set, nor is each member of the set associated with its own idiosyncratic processes—rather, linguistic process apply uniformly to a significant subset of the set. By *variable*, I mean that the number and type of linguistic objects varies from theory to theory, language to language, and even speaker to speaker. By *large*, I mean that the set of linguistic objects is not restricted to a handful of such objects. If the set of linguistically-relevant distinctions is uniform, variable, and large, then it is linguistically productive. This work makes unbounded-distinction idealizations for two different class of features: syntactic features and phonological features. Let us consider each in turn.

A.2.1 Unbounded agreement features

The set of syntactic distinctions is **uniform**—that is, syntactic features are not associated with their own peculiar idiosyncratic agreement process, fundamentally different from all other agreement processes. In the linguistic theory of *generalized phrase structure grammar*, there are only three (overlapping) classes of agreement features (HEAD, FOOT, and CONTROL), to which agreement processes apply. In *lexical functional grammar* and related unification grammars, the sole agreement process (unification) applies uniformly to any subset of features, and most commonly applies to all features together ($\downarrow=\uparrow$). In the *government-binding theories* of Chomsky (1981; 1982; 1986), agreement processes apply uniformly to the unbounded vector of "φ-features."

The set of syntactic distinctions is also **variable**—different languages employ different distinctions, and different theories often postulate wildly different features. I am trying to get results that are invariant across a wide range of linguistic theories. The significance, then, of the fact that the set of syntactic distinctions varies from theory to theory is that this set will most likely be explicitly variable in the 'true' linguistic theory. As mentioned in chapter 4, pronouns in different languages are marked for a wide range of distinctions, and these vary considerably from language to language. In so-called nonconfigurational languages such as Latin, nouns express many more overt case distinctions than in configurational languages such as English. The number of agreement features specified on reflexives varies from language to language: the Russian object reflexive *sebja* is featureless, whereas Modern English reflexives are fully specified for the person, gender, and number of their antecedent (Burzio 1988).

Finally, in syntactic theories that concern themselves with agreement processes the number of distinctions induced by agreement features is certainly **large**. For example, Finnish is known to have sixteen distinct cases, while the Guinness *Book of World Records* states that Tabassaran has 35 different cases, all subject to agreement constraints. In the *Aspects* transformational grammar model, syntactic classes are characterized by at least ten binary features (nominal, verbal, manner, definite, aux, tense, aspect, predicate, adjective, predicate-nominal); prepositional phrases are characterized along an unbounded number of syntactically-relevant dimensions ("Direction, Duration, Place, Frequency, etc" p.107); nouns are distinguished by at least ten syntactically-active features (common, abstract, animate, human, count, det, gender, number, case, declension-class); and verbs are distinguished not only by such features as object-deletion, transitive, and/or progressive, but by their ability to distinguish all other syntactic categories in their selectional restrictions. *Government-binding theories* of syntax are similarly capable of enforcing agreement with respect to a large number of distinctions: for example, selectional agreement occurs along such dimensions as theme (agent, patient, goal, proposition, etc.) and case (nominative, accusative, objective, oblique, genetive, ergative, etc.), in addition to all the distinctions of the Aspects model. In *generalized phrase structure grammar* some agreement features, such as `PFORM` and `SUBCAT`, are capable of making an unbounded number of distinctions—and even if all GPSG features were restricted to two values, GPSG agreement processes would still be sensitive to the more than 10^{775} distinctions made by GPSG's complex feature system (Ristad 1986). *Lexical-functional grammar* has agreement processes sensitive to the literally infinite number of distinctions that LFG's feature system is capable of making (because syntactic categories in LFG may themselves contain an arbitrary number of syntactic categories).

In short, linguistic support for the idealization to an unbounded number of syntactic agreement features is quite significant. Now let us consider whether the same is true for phonological features.

A.2.2 Unbounded phonological features

The set of phonological distinctions is **uniform** with respect to agreement (and other phonological processes) because phonological agreement processes such as assimilation and harmony apply to natural classes of phonological features. That is, no feature has its own idiosyncratic phonological agreement process: rather, one or two phonological agreement processes apply to all natural classes of features, as determined by a language-universal feature geometry (Sagey 1986).

The set of phonological distinctions is **variable** because the set of phonetic segments (and articulatory features) varies from language to language, as do all abstract phonological distinctions such as degrees of tone, vowel height, sonority, and stress. The domain of assimilation processes also varies from theory to theory, language to language, and even from speaker to speaker, as do morpheme classes.

Finally, the number of phonologically significant distinctions is **large**. For one, the human articulatory apparatus can produce a large number of articulatorily–distinct and acoustically–distinct segments. The transcription key in Halle and Clements (1983), lists 52 consonants marked with 12 additional distinctions and 21 vowels marked with 7 additional distinctions, for a total of 771 (= 624+147) purely phonetic distinctions. Their system employs 21 distinctive features, Chomsky and Halle (1968) use 28 distinctive features, and the articulatory tree geometry of Sagey (1986) employs 21 nodes. Phonological processes are additionally sensitive to the distinctions created by the sonority hierarchy; syllable structure (onset, nucleus, coda, appendix, branching/nonbranching structure, number of feet, syllable weight, etc.); tone (a range of discrete steps from highest to lowest in addition to rising and falling tones, and tonal downsteps); stress (degree and type of foot); and so forth. Morphological processes are sensitive to all those phonological distinctions, plus a set of morpheme class distinctions that is itself uniform, variable and large, and hence best seen as unbounded. For example, there are upwards of twenty noun classes in the Bantu languages, and no reason to believe 'noun class 1' in a Bantu language is in any sense the same as 'noun class 1' in a Romance language.

The number of articulatory (phonetic) distinctions would seem to be bounded by human physiology. But there is significant evidence that the bound is not a constant, even with a fixed set of primary articulators. Many features such as vowel height, tone, and sonority may be best seen as the arbitrary discretization of an infinite continuum, a kind of scale. Some languages have six degrees of vowel height, while others have only three; and certainly every language can have its own sonority hierarchy and tonal inventory. Moreover, there is no reason to believe that that the language faculty is incapable of using additional articulators, were they made available. For example, speaking often is accompanied by the meaningful use of hand gestures and facial expressions and the movement of secondary articulators such as the jaw. Thus, although the number of muscles in our bodies is a (large) constant, the language faculty does not appear to be tied to a fixed set of muscles (witness sign languages) or muscular movements, and therefore the idealization to an unbounded number of articulatory features may be empirically correct, in addition to being theoretically justified on grounds of productivity (being uniform, variable, and large). In fact, the language faculty may be maximizing the number of perceptually observable distinctions in the domain of a given sensory-motor system (Stevens and Keyser 1989). Therefore, if the human motor system were capable of producing additional perceptually-distinct segments, the language faculty might employ them.

In conclusion, there is significant support for the idealization to an unbounded number of linguistic distinctions in both phonology and syntax. To assume otherwise is to vastly complicate linguistic theory. To argue otherwise is to present a finite language-universal list of all possible linguistically significant features, a project which has yet to begin and unlikely to finish in this century.

A.2.3 Limiting unbounded idealizations

It does, however, seem reasonable to limit the number of distinctions by some sharp order-of-growth in a natural parameter. The natural parameter for language learner might be the amount of time spent acquiring the language; in the case of a computational complexity analysis, the natural parameter is the size of the input to the reduction. The polynomial time bound on reductions limits us to specifying a feature system with no more than a polynomial $f(n)$ number of symbols, which can make at most $k^{\frac{f(n)}{k+1}}$ distinctions for k-ary features, which is maximal when the features are binary. Some stricter limits include no more than an exponential number of distinctions $2^{k \cdot n}$ (linear number $k \cdot n$ of binary features) or a polynomial number of distinctions n^k (logarithmic number $k \cdot \log n$ of binary features). It is desirable to limit the number of distinctions available to a reduction because this forces us to use other unbounded linguistically-significant distinctions, based on other linguistic structures, in order to simulate increasingly complex computations. In each proof, I explicitly state the number of distinctions required for that proof to succeed.

Appendix B

Structure of Elliptical Dependencies

The goal of this appendix is to provide an analysis of referential dependencies in elliptical contexts, such as VP-ellipsis (70a) and CP-ellipsis (70b), that does not make use of a copy operation.

(70)
a. Felix [hates his neighbors] and so does Max [*e*].
b. Felix told Kyle [that he hates his neighbors] and Max told Lester [*e*].

I argue that the facts of elliptical dependencies can be accounted for by two representational innovations. First, ellipsis is analyzed as identically-composed thematic-structure shared between the overt structure and the corresponding null structure. Second, the two relations of referential dependency, link and obviate, are generalized to hold between positions in the thematic-structure as well as positions in the phrase-structure.

Before proceeding, let us establish some terminology. We say two elements β and β' in different structures *correspond* when they are in equivalent positions and receive equivalent interpretations, assuming an appropriate notion of structural equivalence. In example (70), *Felix* and *Max* correspond, as do *Kyle* and *Lester*.

As we saw in chapter 4, the central theoretical problem posed by ellipsis is that the invisible structure must be an independent copy of the overt structure; yet at the same time, it cannot be.[1]

At beginning of section 4.2, I presented evidence that the overt structure must be copied to the position of null structure in the syntax, that copying is a recursive process, and that anaphoric elements may be linked to their antecedents either before or after the copying (the copy-and-link theory 4.1). Next, I falsified this theory by showing that the null structure is not an independent copy of the overt structure, because (i) the original and its copy do not obey the same post-copying linguistic constraints, and (ii) processes that apply after copying, such as linking, do not apply independently in both the original and its copy. To resolve this apparent paradox, that copying is both necessary and impossible, I sketched an empirically superior predicate-sharing theory 4.2 at the end of section 4.2.

In this appendix, I fill in the details of such theory, and defend it. Briefly, I propose that the invisible VP shares the thematic-structure of the overt VP, but not its phrase-structure or phonology. I also generalize the two relations of referential dependency, link and obviate, to hold between positions in the thematic-structure or phrase-structure. Let α be an anaphoric element in the overt VP, β an argument of the head of that VP, and β' the corresponding argument of the invisible VP. The invariant interpretation of α arises when α links to the phrase-structure position of β. The covariant interpretation arises when α links to the thematic-position assigned to β, because then α also links to the thematic-position assigned to β'.

Now let α locally obviate the thematic-position assigned to β, according to binding condition B. Then α also locally obviates the corresponding argument β' because the same thematic-position is assigned to both β and β'. The cases of invisible obviation arise when α is linked to some antecedent γ and β' is coreferential with γ. Then α is both obviative and coreferential with β', which is a contradiction. The details of this analysis may be found below in section B.4.3.

It is of course possible to develop an alternate analysis, that does not refer to a level of thematic-structure, and does not generalize linking and obviation to thematic-positions. Such an analysis is sketched in section B.3; it is

[1] For historical reasons, ellipsis phenomenon has been called "deletion." Essentially, one of two underlyingly "nondistinct" substructures in a structural description could be deleted in the D-structure to S-structure derivation (Ross 1967; Keenan 1971), in the logical form to S-structure derivation (McCawley 1967), or in the S-structure to PF derivation (Sag 1976).

considerably less elegant. My central motivation in this research, however, is to accumulate evidence for the constructive complexity thesis for human language. In the introduction to the report, I argue that human language has the structure of an NP-complete problem. That is, the process of constructing linguistic representations is bounded above by $\mathcal{NP}$ and below by NP-hardness. As proved in chapter 4, the copy-and-link analysis of ellipsis leads to a complexity outside of $\mathcal{NP}$ (in fact, to PSPACE-hardness). By eliminating the recursive copy operation from linguistic theory, we provably reduce the complexity of representing ellipsis from PSPACE-hardness to inside $\mathcal{NP}$. The fact that such a reduction in complexity is possible constitutes empirical evidence for the $\mathcal{NP}$ upper-bound.

This remainder of appendix is organized as follows. In the next section, section B.1, previous work is reviewed in an attempt to illuminate the inherent structure of an adequate account of elliptical dependencies. We begin with the earliest theories of VPE, confront these theories with empirical difficulties, and in this manner move to successively more sophisticated theories. Section B.2 presents the phenomenon of invisible obviation in detail. Section B.3 discusses the necessary structure of an adequate theory of ellipsis. Section B.4 proposes an explicit system of representation as it applies to VPE, and section B.5 illustrates it for the tricky cases of invisible obviation, invisible crossover, recursive ellipsis, and nonsubject covariance.

B.1 Previous work reconsidered

It is the central insight of early work on VP-ellipsis that both overt and null VPs correspond to identical underlying predicates: either the pronoun *his* inside the identical predicates refers to a constant (either *Bob* or *Felix* in example (63)), resulting in the invariant interpretation, or it refers to an argument of the predicate (in this case, the external argument), resulting in the covariant interpretation. In this chapter, we accumulate evidence for a refinement of this view.[2] If this is so, then the central research questions are how to represent the predicates, and what constitutes identity of

[2]The "identity of predicates" observation has been made in some form or other, apparently independently, by a wide range of authors including McCawley (1967), Keenan (1971), Lasnik (1976), Sag (1976), Williams (1977), and Reinhart (1983). Keenan's work is particularly valuable for the simplicity of its presentation, and Sag's for the breadth of its empirical investigation.

predication?

B.1.1 Covariance reduced to predication of subjects

The guiding idea of both Sag (1976) and Williams (1977) is to reduce anaphoric covariance to the predication of subjects. This idea may be described informally as follows. At some level of representation after S-structure—either at LF or in the discourse grammar—a VP is represented as a one-place predicate, where the external argument of the VP is bound by a λ-variable inside the VP, as in ($\lambda x.[x$ *eat his*$_i$ *dinner*]). Exercising some amount of charitable reconstruction, we may say that pronouns are assigned free "referential" indices at D-structure. The grammar contains an Pronoun Rule that optionally replaces an anaphoric element coindexed with the subject with a variable bound by the λ-operator, as in ($\lambda x.[x$ *eat* x *dinner*]). At some subsequent level of representation, the λ-expression corresponding to the overt VP is copied to the position of the null VP. The covariant interpretation is obtained when the Pronoun Rule applies; the invariant interpretation when it does not.

Although the particular mechanism of λ-abstraction is not a natural component of the current principle-and-parameter framework, this idea may be easily implemented in a number of other ways using mechanisms that have been independently motivated. To illustrate the central issues, we consider two mechanisms: predicate structure and VP-internal subject. In either case, the LF representation of VP-ellipsis is interpreted (after LF) as if the VP predicate appeared in the position of both overt and null VPs.

First, we may appeal to a suitably modified version of Williams' (1980) *predicate structure*, where the subject-predicate relation is represented by coindexing each predicate and its subject. The covariant interpretation, where the anaphoric element α refers to the argument of the predicate, is represented by assigning the same variable index to the predicate and its embedded referentially-dependent element α, as in *John*$_1$ *[ate his*$_i$ *dinner]*$_i$. The invariant interpretation, where α refers to the matrix subject, is represented by assigning the same constant index to α and the matrix subject, as in *John*$_1$ *[ate his*$_1$ *dinner]*$_i$. (Here we temporarily depart from our convention of using subscripts to represent speaker-hearer judgements.)

Second, highly articulated phrase structure can give rise to a linking ambiguity. For example, we might postulate a VP-internal subject position, following Fukui (1986), Kitagawa (1986), Koopman and Sportiche (1985;1986), and Kuroda (1985). Now every embedded pronoun coreferential with the clausal subject may be ambiguously linked directly to the subject, or to the empty position in the VP that is itself linked to the subject. In the latter case, we may obtain the covariant interpretation; in the former case, we obtain the invariant interpretation when the subject is a logical constant, such as a proper noun or definite NP. A typical surface form is:

(71) $[_S[_{NP}\text{Felix}]_1\ [_{VP}[_{NP}e]_1\ [_{V_1}\text{hates}\ [_{NP}\text{his neighbors}]]]_2]$
and $[\text{so does}\ [\text{Max}\ [_{VP}e]_2]]$

When the subject of first conjunct is a logical constant such as the proper noun *Felix*, then the embedded pronoun *his* can be linked either directly to that logical constant to obtain the invariant reading, or to the VP-internal specifier position $[_{NP}e]_1$ (which is itself a logical variable linked to the subject) to obtain the covariant reading.

B.1.2 The problem of nonsubject covariance

The central prediction of any such theory is that the covariant interpretation is only available for anaphoric elements coreferential with the subject of a predicate that contains them. This prediction appears to be false. As Reinhart (1983:152) observes, covariant readings are available when the relevant ellipsis is not VP and when the antecedent not a subject:

(72) a. We paid [the professor]$_1$ his$_1$ expenses, but not [the student]$_2$.
([*we didn't pay [the student]$_2$ his$_{?1/2}$ expenses*])
b. The nurse referred Siegfried$_1$ to his$_1$ doctor, and Felix$_2$ too.
([*the nurse referred Felix$_2$ to his$_{1/2}$ doctor*])
c. You can keep Rosa$_1$ in her$_1$ room for the whole afternoon, and Zelda$_2$ too.
([*you can keep Zelda$_2$ in her$_{1/2}$ room ...*])

The simplest solution to this difficulty is to assign a new phrase structure to these constructions, where what were objects become subjects. Then the Sag–Williams analysis, which reduces covariance to the predication of subjects, would still apply, *mutatis mutandis*. One such approach, due to Kayne (1981;1984), analyzes the double objects of the verbs in (72) as small clauses, as in (73).

(73) We believe [[$_{SC}$ John V [a genius]] and [$_{SC}$ Bill [e]]] too.
([*a genius*])

Kayne suggests that small clauses of the form [$_{SC}$ NP NP] contain an embedded abstract verb-like element V that expresses the thematic relationship between the two objects. (In the case of (73a), Kayne would postulate an underlying abstract 'be' element, ie., [believe [NP [is NP]]]; in (74a), Kayne would postulate an abstract 'have' element, ie., [pay [NP [has NP]]].) Then Reinhart's examples (72) would be assigned the phrase structures shown in (74):

(74) a. We paid [[$_{SC}$ the professor$_i$ [V his$_i$ expenses]$_1$] but not [$_{SC}$ the student [e]$_1$]].
b. The nurse referred [[$_{SC}$ Siegfried$_i$ [V to his$_i$ doctor]$_1$] and [$_{SC}$ Felix [e]$_1$] too].
c. You can keep [[$_{SC}$ Rosa$_i$ [V in her$_i$ room for the whole afternoon]$_1$] and [$_{SC}$ Zelda [e]$_1$] too].

A second approach, due to Larson (1988), would assign the highly articulated phrase structures in (75) to Reinhart's examples from (72):

(75) a. We paid$_i$ [[$_{VP}$ [the professor]$_j$ [$_{V_1}$ [t_i t_j] [his expenses]]] but not [$_{VP}$ [the student]$_k$ [e]]].
b. The nurse referred$_i$ [[$_{VP}$ Siegfried [t_i [to his doctor]]] and [$_{VP}$ Felix [e]] too].
c. You can keep$_i$ [[$_{VP}$ Rosa [t_i [$_{VP}$ [in her room] [t_i [for the whole afternoon]]]]] and [$_{VP}$ Zelda [e]] too]].

These novel phrase structures also hold the promise of assisting our analysis of covariant interpretations in CP-ellipsis, as in (76). (Keenan (1971) analyzes these constructions, which he calls S'-deletion, and demonstrates that the covariant coreference relations in the elliptical clause may be arbitrarily complex.)

(76) John$_1$ told Betty$_2$ [that he$_1$ thought she$_2$ was drunk] and Orville$_3$ told Naomi$_4$ [e] (too)
([*that he$_3$ thought she$_{2/4}$ was drunk*])

As before, both the agent (subject) and benefactor (object) of *tell* would be underlyingly subjects of predicates, and hence both would be available as the antecedent of a covariant interpretation.

Observe that CP-ellipsis (77a) and PP-ellipsis (77b) must be distinguished from the corresponding NP-ellipsis, as in (78).

(77) a. Sally told John [$_{CP}$ that cookies had been baked] and Deb told Andy [e] too.
b. Sally told John [$_{PP}$ about the fresh-baked cookies] and Deb told Andy [e] too.

(78) a. *? Sally told John [$_{NP}$ an interesting story] and Deb told Andy [e] too.
b. * Sally gave John [$_{NP}$ fresh-baked cookies] and Deb gave Andy [e] too.

This distinction appears to be related to underlying differences in thematic-structure. In particular, the *tell* of (77) is like *inform* in that it permits an optional theme argument, as in *Sally informed John*, whereas the *give* and *tell* of (78) are like *relate* in that both require an obligatory theme argument, as in * *Sally related/gave John*. Thus, it seems that ellipsed arguments do not satisfy obligatory selectional constraints.

These facts present a serious difficulty for an approach that attempts to reduce either nonsubject covariance or CP-ellipsis to V1-ellipsis under the Kayne/Larson analysis of double object constructions. In such an approach, the benefactive antecedent β is in the specifier-of-VP position, and the ellipsed structure is a V1. So a rule of V1-ellipsis is needed. However, the unacceptable example (80a) would then be assigned the permissible structure in (80b or c), showing that not all V1 constituents may be the target of ellipsis.[3]

[3] A further technical difficulty, particular to Larson's analysis of double object constructions, occurs in cases of Heavy NP Shift, which Larson analyzes as $V1 \rightarrow V0$ reanalysis followed by Light Predicate Raising. In these cases, the combined analysis incorrectly predicts that the covariant interpretation is not available when the indirect object is the antecedent (79a), simply because the indirect object has been incorporated into the V0 complex and is no longer a subject, as shown in (79b):

(80) a. * Sally gave John fresh-baked cookies and Deb gave Andy [e]

b. [Sally gave [John [$_{V1}$ V fresh-baked cookies]]] and [Deb gave [Andy [$_{V1}$ e]]] too.

c. [Sally gave$_i$ [[$_{VP}$ John$_j$ [$_{V1}$ [t_i t_j] [fresh-baked cookies]]]]] and [Deb gave$_k$ [[$_{VP}$ Andy$_l$ [$_{V1}$ e]]]] too.

For this reason, nonsubject covariance and CP-ellipsis remain open problems in this approach.

B.1.3 Covariance reduced to bound anaphora

Reinhart's solution to the difficulties posed by nonsubject covariance and CP-ellipsis is to reduce the covariant interpretation of anaphora to the bound-variable interpretation of anaphora. This solution, following an earlier suggestion due to Lasnik (1976:20), is based on the observation that the covariant interpretation of an anaphoric element α coreferential with an argument β is available if and only if α can be interpreted as a bound variable in the scope of a QNP in the position of β:

(81) a. We paid [every man]$_i$ his$_i$ expenses.

b. The nurse referred [every victim]$_i$ to his$_i$ doctor.

c. You can keep [some woman]$_i$ in her$_i$ room for the whole afternoon.

Reinhart crucially distinguishes bound anaphora from pragmatic or accidental coreference. Accidental coreference is an extra-syntactic relation between two NPs, either of which may or may not be referentially dependent. Bound anaphora is a syntactic relation between an NP β and an anaphoric element α that is understood as a variable bound by β. It is represented by coindexing α and β, subject to the following conditions: (i) β c-commands α; and

---

(79) a. We paid to [the professor]$_1$ the most outlandish expenses that he$_1$ had ever incurred, and to [his$_1$ student]$_2$ too.
([*we paid to [his$_1$ student]$_2$ the most outlandish expenses that he$_2$ had ever incurred*])

b. We [$_{V1-V0}$ paid to the professor]$_i$ [$_{VP}$ [the most outlandish expenses that he had ever incurred] t_i]

Moreover, the pronoun is no longer c-commanded by its antecedent, which will also incorrectly block the desired covariant interpretation for these structures.

(ii) if α is a pronoun, then β cannot be dominated by the minimal governing category mgc(α) dominating α; (iii) otherwise α is a reciprocal or reflexive, and then β must be dominated by mgc(α). The semantic interpretation of a CP containing β is given by the rule (82), which λ-abstracts β from CP and replaces all α coindexed with (and hence c-commanded by) β by the λ-variable that replaced β.

(82) $[_{\mathrm{CP}}\ \Phi] \Rightarrow [_{\mathrm{CP}}\ \beta\ (\lambda x.[\Phi^{\beta}/x])]$

Example (81a) is assigned the surface structure (83a) by the coindexing rule and the semantic interpretation (83b) by the rule (82):

(83) a. We $[_{\mathrm{VP}}$ paid [every man]$_i$ [his$_i$ expenses]].
b. ([Every man] (λx.[we paid x [x's expenses]])).

It seems that this system is meant to apply to VPE as follows. The CP that contains the overt VP is interpreted by λ-abstracting some of its arguments to form a λ-expression E; the CP that contains the null VP is interpreted by λ-abstracting its overt arguments, and then applying them to E. The covariant interpretation would arise when a λ-abstracted argument β has been coindexed with anaphoric elements in the syntax, as in (84a); the invariant interpretation when β is "accidentally" coreferential with an anaphoric element, as in (84b):

(84) a. ([The professor] (λx.[we paid x [x's expenses]]))
but not ([The student] (λx.[we paid x [x's expenses]]))
b. ([The professor]$_1$ (λx.[we paid x [his$_1$ expenses]]))
but not ([The student] (λx.[we paid x [his$_1$ expenses]]))

This proposal is missing many crucial details; Reinhart does not supply them. Perhaps they can be supplied. There is also the question of the adequacy of the proposed theory of anaphora.[4] Nonetheless, let us assume that it is correct in order to evaluate this approach to covariance.[5]

[4] For one, I am not convinced that the claimed disjoint-reference consequences follow from the proposed pragmatic Gricean theory based on speaker/hearer intentions. Lasnik (1989) discusses these issues and other empirical failings—do not overlook his fn.1.

[5] Even assuming that the proposed technical system (bound anaphora only if c-command holds) covers the central cases, Reinhart (chapter 8) and others have observed

It is not clear what conditions permit an argument in the elliptical structure to be applied to the λ-abstracted predicate constructed from the overt clause. For example, Reinhart's system fails to explain the existence of an independent tense marker in the elliptical clause. That is, why should the subject of the elliptical VP require an inflected agentive *do*, and why can this inflection differ from the corresponding inflection in the overt VP?

(87) a. Felix hated his neighbors, but Max (still) does.
b. Felix hates his neighbors, and Max did too.

Example (88) demonstrates that this is not a question of the case filter applied to the subject of the elliptical clause:

(88) Felix$_1$ may hate his neighbors, but not Max$_2$.
(ie., *Max$_2$ doesn't hate his$_{1/2}$ neighbors*)

Recall also the cases of nonsubject covariance in (72), where the overt antecedent of the covariant pronoun is entirely by itself in the elliptical clause.

---

that bound anaphora are available even when c-command does not obtain. In addition, Reinhart's correlation (covariant interpretation if and only if bound anaphora interpretation) has exceptions. For example, Reinhart predicts that (85b) has a covariant interpretation precisely because (85a) has a bound variable interpretation:

(85) a. Zelda thought about [every man]$_i$ on his$_i$ wedding day.
b. Zelda thought about [every man]$_i$ on his$_i$ wedding day and about Felix$_2$ too.
([*thought about Felix$_2$ on his$_2$ wedding day*])

However when a proper noun replaces the QNP antecedent in this example, the covariant interpretation is crucially *not* available for some speakers:

(86) * Zelda thought about Siegfried$_1$ on his$_1$ wedding day and about Felix$_2$ too.
([*thought about Felix$_2$ on his$_{1/2}$ wedding day*])

(In fact, my informants claim that (86) can only mean that "Zelda thought about Felix too.") This contrast presents difficulties for Reinhart's correlation, as well as for other theories. The natural solution is to appeal the distinction between quantifiers and proper nouns, ie., the former are assigned scope while the latter are not. There are a number of ways to implement this proposal—for example, perhaps anaphoric dependencies (theta-links, in this case) must be established under c-command between the scopal marker of the quantifier and the anaphor *qua* variable. I return to these difficulties below in section B.4.3.

A central property of Reinhart's system, and of the other systems we have considered, is a fundamental asymmetry between overt and null structures. Relations of anaphoric antecedence are established the overt structure, subject to syntactic constraints obtaining in that structure, and then applied to the argument(s) of the null structure. So, if an anaphoric element α links to an argument β in the overt structure, then the copy of α in the null structure will also be allowed to link to the argument corresponding to β.

B.2 Invisible obviation

Now consider the discourse (89), and its variant (37), repeated here as (90).

(89) Ann: Romeo$_1$ wants Rosaline$_2$ to [love him$_1$].
Ben: Not any more—now Romeo$_1$ wants Juliet$_3$ to [e].
([*love him*$_1$])

(90) Ann: Romeo$_1$ wants Rosaline$_2$ to [love him$_i$]. ($i = 1$)
Ben: Not any more—now Rosaline$_2$ wants Romeo$_1$ to [e].
([*love him*$_i$], $i \neq 1$)

In both examples, Ann's use of the pronoun *him* is most naturally understood as referring to *Romeo*. Yet when Ben replies in example (90), the coreferential interpretation ($i = 1$) is no longer possible in Ann's statement. This "invisible" relation of local obviation can also be created entirely within a sentence, with the pronoun understood as first including but later obviating the argument *Romeo*:

(91) Romeo$_1$ wanted Rosaline$_2$ to [love him$_i$] before wanting himself$_1$ to [e].

Similarly, an R-expression is "invisibly obviative" from its local c-commanders, as in (92), where pragmatic considerations strongly favor a coreferential interpretation that can only be excluded by syntactic principles.

(92) Sue likes Narcissus$_1$ and he$_{*1}$ does [e] too.

There are a number of subtleties, however, the most interesting of which is that invisible obviation is entirely a local phenomena, as illustrated in (93).

(93) a. He_{*1} knew Juliet loved Romeo_1.
b. The nurse [knew Juliet loved Romeo_1] before he_1 did [e].
([*know Juliet loved* $\textit{Romeo}_1$])

Although the pronoun *he* must be obviative from the R-expression *Romeo* that it overtly c-commands in (93a), it need not be obviative from the R-expression that it invisibly c-commands in (93b).

In fact, the domain of invisible obviation is exactly the local domain of binding condition B. Let position i c-command position j in a phrase marker. Then invisible obviation holds between positions i and j if a pronoun in position j would be obviative from an argument in position i (unless, of course, there is an anaphor is position j).

This Invisible Obviation Condition (IOC), a descriptive generalization that follows from deeper principles discussed below, is illustrated by the following examples for both pronouns (96b,97b) and R-expressions (96c,97c) c-commanded in position j:[6,7]

[6]It appears that both overt and invisible condition C effects between two R-expressions can be overcome with heavy phonological stress, as in (94).

(94) a. BILL_1 wanted BILL_1 to kiss Mary.
b. Sue [wanted BILL_1 to kiss Mary] and BILL_1 did [e] too.

whereas invisible condition C effects between an R-expression and a c-commanding pronoun are inviolable, regardless of the amount of stress (95).

(95) a. * He/HE_1 wanted Bill/BILL_1 to kiss Mary.
b. * Sue [wanted Bill/BILL_1 to kiss Mary] and he/HE_1 did [e] too.

[7]The examples in (96b) and (97b) are constructed using a unique antecedent *Bill* to more clearly reveal the invisible obviation configuration. However, the IOC appears to overlap in these examples with an independent (not understood) constraint that excludes some cross-conjunct antecedences, as in *Bill wanted* $\textit{him}_{*1}$ *to win and* $\textit{Tom}_1$ *wanted himself to win (too)*.

(96) a. $Bill_1$ wanted him_{*1} to kiss Mary.
b. Sue [wanted him_{*1} to kiss Mary] and $Bill_1$ did [e] too.
c. Sue [wanted $Bill_1$ to kiss Mary] and he_{*1} did [e] too.

(97) a. $Bill_1$ wants PRO_1 to love him_{*1}.
b. Sue wants Mary to [love him_{*1}] and $Bill_1$ wants PRO_1 to [e] (too).
c. Sue wants him_1 to [love Mary] and $Bill_1$ wants PRO_1 to [e] (too).

Examples (98) demonstrate that the nonlocal obviation defined by binding condition C is not relevant to the IOC.

(98) a. $Bill_1$ wanted Mary to kiss him_1.
b. He_{*1} wanted Mary to kiss $Bill_1$.
c. Sue [wanted Mary to kiss him_1] and $Bill_1$ did [e] too.
d. Sue [wanted Mary to kiss $Bill_1$] and he_1 did [e] too.

The fact that the IOC should be defined relative to condition B and not in terms of (the negation of) condition A is illustrated with a prepositional adjunct in (99), and with a possessive NP in (100).

(99) a. $Bill_1$ saw a snake near him_1/$himself_1$.
b. He_{*1} saw a snake near $Bill_1$.
c. Tom [saw a snake near $Bill_1$] before he_1 did [e].

(100) a. $Bill_1$ knew that pictures of him_1/$himself_1$ would be on sale.
b. He_{*1} knew that pictures of $Bill_1$ would be on sale.
c. Sue [knew that pictures of $Bill_1$ would be on sale] before he_1 did [e].

As noted in the introduction, the invisible structure is not an invisible pronoun, simply because there is no invisible obviation when an overt pronoun is used (39b,102b) instead of ellipsis (39a,102a):[8]

[8]The effects of invisible obviation are most pronounced when the invisible pronoun obviates an anaphoric element that must have a local antecedent, such as an anaphor.

(102)
a. Juliet$_1$ thought that the Friar$_2$ [poisoned her$_1$] without realizing she$_{*1}$ did [e].
b. Juliet$_1$ thought the Friar$_2$ [poisoned her$_1$]$_3$ without realizing that she$_1$ did it$_3$.

If the null structure were simply an empty pronoun at LF, then there would be no way to explain the lack of invisible obviation in (39b, 102b).

The following examples are particularly interesting because they demonstrate that local obviation in the overt structure is preserved in the null structure, even when it is embedded one level, as in (103), or more than one level, as in (104).

(103)
a. Bill$_i$ [wants PRO$_i$ to love him$_{*i}$].
b. Sue$_i$ [wants PRO$_i$ to love Bill$_1$] and he$_{*1}$ does [e] too.
c. Sue$_i$ [wants PRO$_i$ to love him$_{*1}$] and Bill$_1$ does [e] too.

(104)
a. Bill$_i$ [expects PRO$_i$ to want PRO$_i$ to love him$_{*i}$].
b. Sue$_i$ [expects PRO$_i$ to want PRO$_i$ to love Bill$_1$] and he$_{*1}$ does [e] too.
c. Sue$_i$ [expects PRO$_i$ to want PRO$_i$ to love him$_{*1}$] and Bill$_1$ does [e] too.

Contrast these examples to the examples (105), which show that the nonlocal obviation of condition C is not similarly preserved under embedding.

The only configurations with this property require an elliptical infinitival VP, where the subject of the null infinitival VP is an anaphor, as in (101a). In such a configuration, however, it is not possible to directly pronominalize the overt VP (101b), perhaps for reasons having to do with the case filter. Instead, we must introduce the agentive *do*, as in (101c).

(101)
a. Romeo$_1$ asked the apothecary$_2$ to [kill him$_{*1/*2}$] before telling himself$_1$ to [e].
b. * Romeo asked the apothecary to [kill him] before telling himself to it.
c. Romeo$_1$ asked the apothecary$_2$ to [kill him$_{1/2}$]$_3$ before telling himself$_1$ to do it$_3$.

These examples raise doubts as to whether we can consistently view both *do* and *to* as realizations of I0.

(105) a. $Bill_i$ [wants PRO_i to know if Mary loves him_i].

b. He_i [wants PRO_i to know if Mary loves $Bill_1$]. $(i \neq 1)$

c. Sue_i [wants PRO_i to know if Mary loves $Bill_1$] and he_1 does [e] too.

d. Sue_i [wants PRO_i to know if Mary loves him_1] and $Bill_1$ does [e] too.

These facts are exactly in accordance with the IOC.[9]

These apparently novel examples provide powerful evidence for the structure of elliptical dependencies. For one, they argue against any non-syntactic account: neither the discourse grammar of Williams (1977) nor the semantic interpretation of Reinhart (1983) are able to maintain the purely syntactic (ie., sentence-level) distinction between local and nonlocal obviation necessary for the IOC. Second, they argue against the standard asymmetric account, where the coreference relations in the overt structure are simply imposed on the null structure in a manner that satisfies "identity of predication." Third, they also constitute new empirical evidence for the existence of an explicit *relation of obviation* that, like relation(s) of antecedence, is computed in the syntax at S-structure and subject to semantic interpretation (Lasnik 1976;1981; Chomsky 1980; Finer 1984; Higginbotham 1985).[10]

---

[9]Chomsky (1981) and other authors have accounted for strong crossover phenomenon as a condition C effect. The fact that condition C effects do not arise in elliptical structures gives us a direct empirical test for this hypothesis. The facts of invisible strong crossover in (106) mean that strong crossover cannot be due to condition C.

(106) a. The man who_i $Mary_1$ said that he_{*i} likes t_i.

b. The man who_i he_{*i} said that $Mary_1$ likes t_i.

c. The man who_i $Mary_1$ [said that she_1 likes t_i] and who_i he_{*i} did [e] too.

[10]Finer (1984) exhibits a class of human languages with a "switch-reference" system, where the relations of coreference and obviation between subjects are overtly expressed in the phonetic form of an utterance, by distinct morphemes.

We discuss the details of our representation below.[11,12]

B.3 Necessary structure of an explicit theory

In order to represent the elliptical structures in the T-model with the standard binding theory and using coindexing, we would need to postulate an S-structure to LF mapping with following six properties.

1. The mapping must include a copy operation capable of copying the entire overt structure to the position of the null structure, even when the null structure is in a different sentence in the discourse, a not-insignificant revision of the T-model, which is (was?) a theory of sentence grammar.

2. This copy operation must be able to replace an anaphoric element α with a variable that sometimes inherits the agreement features of α, as evinced by examples (61) above.

3. In order to account for the invariant interpretation of an anaphoric element α, the copy operation must be able to sometimes copy the referential index of α.

4. In order to account for the covariant interpretation of α without overgenerating as in (63, 64), the copy operation must be able to assign the copied α the referential index of the argument that corresponds to β when it does not copy the referential index of β.

5. In order to account for the lack of invisible condition A or nonlocal condition C effects, we must confine binding conditions A and C to

[11] The representation for anaphora we propose is similar to that of Chomsky (1980), with the crucial difference that our obviation is a relation between positions: phrase-structure obviation may be nonlocal and is always overt, while thematic-obviation is local and may be invisible. Chomsky's "anaphoric indices" are relations among arguments and hence would not work for the preceding examples of invisible obviation.

[12] An explicit relation of obviation is independently motivated on conceptual grounds. Representations have a degree of permanence beyond conditions on those representations; conditions should not apply beyond the "creation" of those representations. Obviation between two positions is a condition that must be satisfied in the semantic interpretation of a linguistic representation in the context of the discourse, and therefore is a relation of its own, not merely a precondition on the construction a syntactic relation of coreference or antecedence.

S-structure.

6. In order to account for the IOC, condition B must be enforced at LF, and R-expressions must be replaced with pronouns at S-structure, but only when the LF copy of an R-expression is c-commanded by a coreferential argument in its minimal governing category.

Let us therefore consider an alternate approach.

We have seen that the overt and null structures are symmetric with respect to certain relations of referential dependency (antecedence, local obviation) while being asymmetric with respect to conditions on those relations (agreement, binding condition A). That is, the conditions are strictly enforced in the overt structure, but blithely ignored in the null structure. This strongly suggests that there is really only one underlying representation, that the overt and null structures correspond to the same underlying thematic function.

A more elegant representation for ellipsis, then, is to segregate phrase structure from thematic structure and posit a relation of local obviation that holds between thematic-positions. Then we would simply say that the overt and null structures share the same thematic structure and hence they share the same relations of linking and local obviation. Invariant and covariant interpretations are accounted for by linking at the levels of phrase- and thematic-structure, respectively. The IOC is accounted for by defining condition B in terms of nonanaphors, as the obviation of thematic-positions.

Let us now make our representations explicit.

B.4 The proposed system of representation

B.4.1 Segregate phrase and thematic structure

We propose to segregate phrase structure from thematic structure as follows. Phrase structure consists of a set of labeled points (syntactic constituents) and the familiar relations of immediate domination, precedence, and so forth.

Thematic structure consists of arguments, functions from thematically-typed argument positions to functions and to arguments, and the relations defined on those objects, including relations of thematic discharge between argument positions and arguments (theta-marking), as well as relations entirely between argument positions (theta-binding and theta-identification).[13] Arguments are identified by integral indices; functions are identified by integral subscripts on the generalized function symbol f. For example, "$f_{10}(1:\texttt{role1}, 2:\texttt{role2})$" identifies the particular function $f_{10}(\cdot)$, a function of two thematically-typed arguments. For clarity, only the theta-position that is currently being discharged is depicted, as in "$f_{10}(2:\texttt{role2})$" when the theta-position 2 of theta-type role2 is being discharged. By convention, the (curried) arguments of a given function are assigned successive indices starting with 1.

Motivated by Marantz (1990), I propose thematic-functions of order 3. That is, a V0 function $f_i^2(\cdot)$ in combination with its most affected internal arguments (such as inalienable possessor, theme/patient, instrument, or affected object locative) returns a new verb-stem thematic function $f_i^1(\cdot)$ (Marantz's "event 1"). The verb-stem function $f_i^1(\cdot)$ combines with the next most affected internal arguments (such as benefactive, directional locative, or alienable possessor) to return a new VP thematic function $f_i^0(1:\texttt{tense}, 2)$, which is a function from an I0 tense and an external argument (typically an actor) to an entirely saturated function (an argument of thematic-type 'proposition').[14] This will play a role in my account of why a verb plus the benefactor is never the target of ellipsis, see section B.5.1.

The relation between phrase and thematic structures is represented as the pairwise connection between elements in the structures. For example, (107) is assigned the structure described in (108-110), where we have suppressed

[13] As mentioned in chapter 4, a thematic position must inherit the agreement features of the argument that saturates it, for some arguments. The set of pronouns to which this process applies varies from speaker to speaker.

[14] The notation used for thematic structure (ie., functions of various orders) does not matter formally because all choices are formally equivalent. Anything done by one can be done by the other with ambiguity (overlays). Marantz (1989) distinguishes two representations of thematic functions. In one, the verb is a 0-order function of all arguments contained in the proposition (internal arguments, event, and external argument). This approach makes use of "thematic-grids" (cf., Stowell 1981; Levin and Rappaport 1988; Higginbotham 1985;1989). Marantz contrasts this with the use of higher order functions, which have the advantage of naturally representing the fact that less affected arguments are assigned a compositional theta-role, resulting from a verb in combination with its more affected internal arguments (cf., Chomsky 1981; Marantz 1984). To more naturally capture compositional thematic-role assignment, I use higher-order functions. Another reason for using higher-order thematic-functions is to define a c-command relation on thematic-positions, which will greatly simplify the statement of the binding theory.

many important details not relevant in this context.

(107) Felix hates his neighbors.

The possessive morpheme ['s] is a function of two arguments, the possessor and the possession (ie., the possessed thing). First the pronoun *he* is theta-marked with the theta-position $f_{20}(1)$. Then $f_{20}(2)$ theta-binds $f_{30}(1)$, resulting in argument 4 (the closed NP *his neighbors*).

(108) $[_{\mathrm{NP}}[_{\mathrm{NP}}[_{\mathrm{N0}}\mathrm{he}, 3]\ [_{\mathrm{N0}}\text{'s}], f_{20}(1:\mathtt{owner})], f_{20}(2:\mathtt{possession})]$
$[_{\mathrm{N0}}\mathrm{neighbors}, f_{30}(1:\mathtt{object})], 4]$

Next, the V0 function $f^2_{10}(1:\mathtt{object})$ of the verb *hate* theta-marks argument 4, and returns the VP function $f^0_{10}(\cdot)$:

(109) $[_{\mathrm{VP}}[_{\mathrm{V0}}\mathrm{hate}, f^2_{10}(1:\mathtt{object})]\ [\textit{his neighbors}, 4],$
$f^0_{10}(1:\mathtt{event}, 2:\mathtt{actor})]$

The phrase-thematic structure detailed in (108) is italicized and summarized in (109). Finally, $f^0_{10}(1)$ theta-marks argument 2 (past tense I0), and $f^0_{10}(2)$ theta-marks argument 1 (the subject *Felix*), resulting in argument 5 (a proposition):

(110) $[_{\mathrm{IP}}[_{\mathrm{NP}}\mathrm{Felix}, 1]$
$[_{\mathrm{I1}}[_{\mathrm{I0}}[\text{-past}], 2]\ [_{\mathrm{VP}}\textit{hate his neighbors}, f^0_{10}(1:\mathtt{event})],$
$f^0_{10}(2:\mathtt{actor})], 5]$

B.4.2 Two relations of referential dependency

We further propose two relations of referential dependency, link and obviate, defined on constituents (phrase structure points) and on theta-positions (the argument positions of thematic functions). The linking of theta-positions is favored over the coindexing of theta-positions by the same reasons that favor linking over coindexing in the phrase structure.[15] It is needed to

[15]There are arguments both ways; whatever arguments can be made for linking, can also be made for theta-linking. Lasnik (1989) argues for a relation of coindexing on the basis of condition C effects. Binding condition C is somewhat mysterious in a linking theory, because R-expressions need never have antecedents, yet must have links to induce condition C effects.

reveal more information about split antecedents and the interpretation of direct/indirect antecedence. For expository convenience, we write "α theta-links to β" when we mean that "the theta-position assigned to α links to the theta-position assigned to β."

We significantly simplify the binding theory by representing all arguments of the verb in one order-3 thematic-function. This reformulation of the binding theory does not solve many well-known problems for the standard formulation, such as the permissibility of *John$_i$ has himself$_i$ for him$_i$ to blame t_i*.

Let the argument α be assigned the theta-position $f_i(j)$ and be governed by the c-commanding function $f_k(\cdot)$. Then condition A requires a link between $f_i(j)$ and some $f_k(m)$ for [+anaphor] α. Condition B states that $f_i(j)$ obviates all $f_k(m)$ for [-anaphor] α. (Recall that condition B must be stated in terms of [-anaphor], rather than the widely assumed [+pronominal], in order to obtain the IOC effects for pronouns as well as R-expressions.) When α is controlled PRO, then it is obligatorily theta-linked to its controller, always resulting in the covariant interpretation. All such referentially-dependent α must be linked or theta-linked to some β at S-structure, subject to these binding conditions. Binding condition C requires that an R-expression obviate all c-commanding phrase structure positions (in the domain of the head of its chain, if strong crossover is reduced to condition C effects for variables).

For example, $\mathsf{obviate}(f^2_{10}(1),f^0_{10}(2))$ holds in example (108-110), and when the antecedent of *his* is *Felix*, then either $\mathsf{link}(3,1)$ or $\mathsf{link}(f_{20}(1),f^0_{10}(2))$. We leave unanswered the question of whether $\mathsf{obviate}(f_{20}(1),f_{20}(2))$ should be included in this list of referential dependencies, as it seems it should be.

Following Higginbotham (1985:573-5), $\mathsf{link}(\alpha,\beta)$ is interpreted to mean 'include some values of β in the values of α,' while $\mathsf{obviate}(\alpha,\beta)$ is interpreted to mean 'α and β cannot share any values in the structure in which they occur.'[16]

The necessity of explicitly representing the invariant interpretation using one (shared) structure prevents us from making our coreference relation entirely between thematic positions. Alternately, the necessity of explic-

[16] This does not quite work for cases of split antecedents, where the verb seems to have a significant effect on whether pronoun obviation is enforced. For example, (111a) is perfect, (111b) only slightly degraded, and (111c) entirely unacceptable; yet all are excluded by condition B.

itly representing all coreference relations at S-structure, including covariant interpretations, prevents us from making antecedence a relation entirely between phrase structure positions. Therefore, both instances of the link relation are needed in the theory in order to explicitly represent a perceivable distinction between invariant~covariant interpretations. The use of one linking relation defined on two different types of points should not raise objections on the grounds of parsimony when both are necessary, as well as independently motivated. The linking of phrase-structure positions is motivated by Higginbotham (1983) for anaphoric antecedence, while the linking of thematic-positions is motivated by Higginbotham (1985;1989) as a primitive semantic operation. The mathematician should not find the the generalization of linking to include thematic positions objectionable, because it has little effect on the computational and generative power of the theory. (In fact, it is a straightforward proof that representing ellipsis is in the complexity class $\mathcal{NP}$, a theorem that is not so obvious when a copying operation is used instead.)

B.4.3 Ellipsis as a shared thematic-function

An elliptical structure, then, is a proposition P' that contains a thematic function $f_i(\cdot)$ that is not phonologically realized. Rather, it is borrowed from a proposition P that appears earlier in the discourse, where $f_i(\cdot)$ is the result of combining an overt verb with some of its arguments in P.

The borrowed thematic function must be composed in the same way, as shown in (112) for a raising verb, and in (113) for passive.

(112) a. A man [arrived with his mother] and a woman did [e] too.
b. * There [arrived a man with his mother] and a woman did [e] too.

---

(111) a. John$_1$ suggested to Bill$_2$ that he$_2$ shoot them$_{\{1,2\}}$.
b. ? Navarre$_1$ suggested to Benedict$_2$ that he$_2$ persuade them$_{\{1,2\}}$ to abjure sensual pleasures.
c. * John$_1$ suggested to Bill$_2$ that he$_2$ tell them$_{\{1,2\}}$ to leave.

(113)
a. A state college [granted Charlie a degree] and a private college did [e] too
b. * Charlie [was granted a degree] and a private college did [e] too

According to the principle of full interpretation, a logical constant is licensed only if it saturates a thematic-position, whereas a logical operator is licensed only if it binds a logical variable that saturates a thematic-position. Consequently, a logical variable is licensed only if it both saturates a thematic-position and is bound by an operator. An elliptical structure is licensed only if all overt elements that it contains are licensed. Therefore, the elements in an elliptical structure are subject to two constraints: (i) each logical argument (constant or variable) must be assigned a thematic-position by the shared thematic function; (ii) each logical operator with scope over the shared thematic function $f_i(\cdot)$ must bind a thematic-position that is free in $f_i(\cdot)$. The denotational semantics of an elliptical structure is given by substitution.

The principle of full interpretation, then, establishes a element-wise bijection between some elements in P and some in P'. An element $\alpha \in P$ and an element $\alpha' \in P'$ are said to *correspond* iff (i) both are logical arguments that saturate the same thematic-position $f_i(j)$, for some j; or (ii) both are logical operators that bind the same thematic-position $f_j(k)$, for some $f_j(k)$ free in $f_i(\cdot)$.

Argument correspondence

Let us now examine, in detail, our representations for the central case of VP-ellipsis. VP-ellipsis results in correspondences between two I0 arguments, and between two external arguments. The example (114a) of VP-ellipsis is assigned the partial representation (114b-d). The shared VP thematic function $f_{10}^{0}(\cdot)$ appears in (114c).[17]

[17] For clarity, I have not represented the thematic structure of the coordinating conjunction. However, it seems clear that coordinators are higher-order functions, from a sequence of thematic functions $f_i(\cdot)$, ..., $f_j(\cdot)$ of identical structure to a new function $f_{j+1}(\cdot)$, also with the same structure. This, in any event, is the intuition underlying "across-the-board" constraints.

(114)
a. Felix hates his neighbors and Max does too.
b. $[_{NP}[_{NP}[_{N0}$he, 3] $[_{N0}$'s], $f_{20}(1 : \texttt{owner})$], $f_{20}(2 : \texttt{possession})$]
$[_{N0}$neighbors, $f_{30}(1 : \texttt{object})$], 4]
c. $[_{VP}[_{V0}$hate, $f^2_{10}(1 : \texttt{object})$] [*his neighbors*, 4],
$f^0_{10}(1 : \texttt{event}, 2 : \texttt{actor})$]
d. $[[_{IP}[_{NP}$Felix, 1]
$[_{I1}[_{I0}$[-past], 2] $[_{VP}$*hate his neighbors*, $f^0_{10}(1 : \texttt{event})$],
$f^0_{10}(2 : \texttt{actor})$], 5]
[and $[_{IP}[_{NP}$Max, 6]
$[_{I1}[_{I0}$[-past], 7] [$f^0_{10}(1 : \texttt{event})$],
$f^0_{10}(2 : \texttt{actor})$], 8], 9]

A perceivable ambiguity arises when β is the external argument of a VP in a VPE structure. In this case, a link gives rise to the invariant interpretation, while a theta-link yields the covariant interpretation. (In other cases, link and theta-link can both result in an invariant interpretation.) Thus, the partial representation in (114) may be completed in one of two ways:

(115)
a. obviate($f^2_{10}(1)$,$f^0_{10}(2)$), link(3,1)
b. obviate($f^2_{10}(1)$,$f^0_{10}(2)$), link($f_{20}(1)$,$f^0_{10}(2)$)

The link in (115a) gives the invariant interpretation, while the theta-link in (115b) gives the covariant interpretation. (Recall that $f_{20}(1)$ is assigned to argument 3, the pronoun *he*.)

Operator correspondence

We just saw how the outer arguments of an elliptical structure can share the thematic function created in a distinct proposition by a verb in combination with its inner arguments. The proposed system allows another possibility, where an operator with scope over a thematic function shares that thematic function with a corresponding operator at LF. That is, if α is an operator that binds a variable e in an overt proposition P, and α' is the corresponding operator with scope over an ellipsed proposition P', then α' will be interpreted as if it also binds the variable e in P'. Recall that logical constants may become operators as the result of a focusing process. Focus may be reliably correlated with stress, at least in English.

This solves the difficulty discussed above in the context of the Sag–Williams proposal, which is the problematic cases of nonsubject covariance and CP-ellipsis. Observe that when the corresponding arguments receive parallel phonological focus, then both may be antecedents of the covariant pronoun in (116a). However, when only one argument is focused, as in (116b), or when the focus is applied elsewhere (116c), then the covariant interpretation is no longer available. (In fact, (116c) can only mean that "the bursar paid the student," never "the bursar paid the student expenses.")

(116)
a. The bursar paid the PROFESSOR$_1$ [his$_1$ expenses] and the STUDENT$_2$ too.
b. ? The bursar paid the PROFESSOR$_1$ [his$_1$ expenses] and the student$_2$ too.
c. * The BURSAR paid the professor$_1$ [his$_1$ expenses] and the student$_2$ too.

Given these facts, the explanation of nonsubject covariance would seem to lie not in reducing nonsubjects to subjects, but in reducing corresponding antecedents to corresponding logical operators. That is, focused elements are assigned scope at LF, and when the corresponding thematic arguments of a coordinate structure are focused, then the covariant interpretation becomes possible. The LF representation assigned to the covariant interpretation of (116a) would involve proposition-ellipsis and look something like (117):

(117) [[$_{[+\text{FOCUS}]}$ the professor]$_i$ [the bursar paid e_i his$_i$ expenses]$_1$] and [[$_{[+\text{FOCUS}]}$ the student]$_i$ [e]$_1$]

The examples of CP-ellipsis would be assigned a similar structure, where the focused actor and focused benefactor become logical operators with scope over the shared proposition *that he thought she was drunk*, perhaps as in (118) for the example (76).

(118) [John$_i$ [Betty$_j$ [e_i told e_j [that e_i thought e_j was drunk]$_3$]]] and [Orville$_i$ [Naomi$_j$ [e_i told e_j [e]$_3$]]] too

The covariant interpretation of the anaphoric elements in the shared proposition is enhanced by the "parallelism cue" *too*, as well as by the equivalence between matrix verbs.

The possibility of operator correspondence resolves an open problem, namely the fact that an anaphoric element may receive a covariant interpretation even when its antecedent is not an argument of the shared thematic function, provided that the corresponding antecedents are focused, as in (119), or they are inherently logical operators, as in (120).

(119) a. TOM_1 [said that Sue [kissed him_1] before BILL_2 asked Mary to [*e*]].
([*kiss* $him_{1/2}$])

b. TOM_1 [wanted Sue to [kiss him_1] before BILL_2 asked Mary to [*e*]].
([*kiss* $him_{1/2}$])

(120) a. Which man_1 [said that Sue [kissed him_1] before which boy_2 asked Mary to [*e*]].
([*kiss* $him_{1/2}$])

b. Every man_1 [wanted Sue to [kiss him_1] before some boy_2 asked Mary to [*e*]].
([*kiss* $him_{1/2}$])

(The relevant interpretation is the one where the adjuncts are understood as being associated with the higher verb, ie., *say/want before*.)

An interesting property of the example (121) is that a verb with its benefactive argument, ie., *inform Mary*, is in some sense equivalent to a verb without any of its internal arguments, ie., *say*.

(121) TOM_1 said [that Sue kissed him_1] and BILL_2 informed Harry_3 [*e*]
([*that Sue kissed* $him_{1/2/*3}$])

We know on independent grounds that the benefactor does not saturate the first argument position of the verb; rather it is an argument of the verb plus its theme/patient (Marantz, 1990). That is, the complex predicate analysis of double object constructions states that *inform*, *inform S*, and *inform NP S* correspond to possible thematic functions, whereas *inform NP* cannot. The covariant interpretation of example (121) cannot be accounted for as theta-linking, given the complex predicate analysis suggested in section B.1.2, but may be accounted for straightforwardly as operator correspondence in the analysis proposed in this section.

The central conceptual problem with the proposed system is one of parsimony. To put things in the worst possible light, linking can result in an invariant interpretation when the antecedent is a logical constant, or in a covariant interpretation when the antecedent is logical operator. Likewise, theta-linking can result in a covariant interpretation, or an invariant interpretation when the antecedent is a logical constant outside the domain of the shared thematic-function. Thus, it would seem that theta-linking is entirely unnecessary, that we can always account for the covariant interpretation as linking to a logical operator.

However, this is not the case. What needs to be explained is the complex interaction among (i) the phonology (stress/unstressed antecedents), (ii) the logical type of the antecedent (operator/argument), (iii) the domain of the shared thematic-function (includes/excludes antecedent), and (iv) the invariant and covariant interpretations. Linking, theta-linking, and argument and operator correspondence are all needed in order to account for the complex array of facts we have seen so far. The next section presents additional evidence.

B.5 The space of elliptical structures

In this section, we exercise the proposed system. First, we enumerate the empirical consequences of our decision to use thematic functions of order-3. Next, we show how the proposed system accounts for the tricky cases of invisible crossover, invisible obviation, and recursive ellipsis.

B.5.1 Possible domains of ellipsis

Given our decision to use thematic functions of order-3, we predict that the V0 function, the verb-stem function, the VP function, and the saturated proposition function are the only possible targets for ellipsis. Let us briefly consider each in turn.

True V0-ellipsis is only possible for verbs of only one argument, the external argument, as in (122).

(122) a. Bill [left] and John did [e] too
b. Bill wanted to [leave] and John wanted to [e] too

(It is of course difficult to distinguish V0-ellipsis from verb stem- or VP-ellipsis in these cases.) Other examples of V0 ellipsis, such as (123), are excluded by independent principles of the grammar, perhaps the case filter applied to internal arguments.

(123) a. * John [saw] Mary and Bill did [e] Kate (too).
b. * John [gave] Mary books and Bill did [e] Kate records.

Gapping structures, as in (124a), cannot involve true V0 ellipsis, because both the verb and its I0 tense argument are gapped. This means we cannot employ a theta-linking analysis. Rather, we must follow Pesetsky (1982) in analyzing gapping as LF argument raising, perhaps by the mechanism of focus, combined with proposition-ellipsis, as sketched in (124b). That is, gapping is analyzed as the correspondence of logical operators.

(124) a. John saw Mary and Bill, Kate.
b. [John$_1$ [Mary$_2$ [t_1 saw t_2]$_3$]] and [Bill$_1$ [Kate$_2$ [e]$_3$]]

Verb-stem ellipsis is exemplified in (125) for the verb plus theme/patient.

(125) John [donated money] to the Red Cross and Bill did [e] to the Boy Scouts.

Again, the verb plus directional locative cannot be ellipsed without violating the case filter. However LF argument raising may be combined with proposition ellipsis, as shown in (126) to create the appearance of verb-stem ellipsis.

(126) a. John took a bus to New York and Tom, a plane.
b. [John$_1$ [a bus$_2$ [t_1 took t_2 to New York]]] and [Tom$_1$ [a plane$_2$ [e]]]

The verb plus benefactive cannot be ellipsed unless the theme/patient is also ellipsed (127).

(127)
a. * John [donated] money [to the Red Cross] and Bill did [e] time.
b. * John [gave the Red Cross] money and Bill did [e] time.

Nor can the verb plus directional locative be ellipsed unless the instrument is also ellipsed (128).

(128) * John [took a plane] to New York and Tom did [e] to Los Angeles.

It is possible, however, to ellipse the verb plus benefactive when the verb is passivized, raising the theme (129a). When the I0 morpheme is missing, as in (129b), the utterance must be analyzed as operator correspondence combined with proposition ellipsis.

(129)
a. Money was [given (to) the Red Cross] and time was [e] too.
b. Money [was given the Red Cross] and time [e] too.

Again, the gapping structures in (130) must be distinguished from true ellipsis of the verb plus benefactive or verb plus directional locative, which are both impossible.

(130)
a. John donated money to the Red Cross and Bill, time.
b. John took a plane to New York and Tom, to Los Angeles.

Proposition ellipsis is exemplified by gapping structures, and by the examples of nonsubject covariance, as discussed above in section B.4.3.

B.5.2 Invisible crossover

Recall that Higginbotham's (1985) theta-binding relation of thematic-discharge holds between a determiner and the open theta-position associated with nominals. Chomsky (1982) suggests that the impossibility of iterating determiners may be related to a prohibition against vacuous quantification. This motivates Higginbotham to equate theta-binding to the quantification of theta-positions, in order to block iterated determiners. The theta-link relation of referential dependency proposed here and Higginbotham's theta-binding relation are sufficiently similar to suggest that theta-binding reduces to theta-linking, perhaps under the stricter locality constraints applying generally to all relations of thematic-discharge. If this is so, as certainly seems plausible, then our system equates the covariant interpretation of an anaphoric element to the quantification, and "invisible crossover" effects should accrue in VPE. One such example—first noted by Dahl (1972;1974) and first explained by Sag (1976:137) as crossover violation—is (131a), whose null structure has the three interpretations shown in (131b-c).

(131)
a. Bill_1 [believed that he_1 loved his_1 wife] and Harry_2 did [*e*] too.
b. ([*believed that* he_2 *loved* $his_{1/2}$ *wife*])
c. ([*believed that* he_1 *loved* $his_{1/*2}$ *wife*])

In order to obtain interpretation (131b), *he* must be theta-linked to *Bill*, while *his* may be either linked or theta-linked to *Bill* in the overt structure, resulting in an ambiguity in the null structure. However, in order to obtain interpretation (131c), *he* must be linked to *Bill* in the overt structure. Now in order to obtain the covariant interpretation of *his* in the null structure, *his* must be theta-linked to the external argument of the matrix VP (*Bill*) in the overt structure. But this is excluded as a crossover configuration in the first conjunct of (131a): the theta-position assigned to $Bill_1$ "quantifies" the theta-position assigned to his_1, crossing over a pronoun (he_1) that his linked to the argument *Bill*.

B.5.3 Invisible obviation

Invisible obviation structures are illustrated in the representation (134) of our previous example (91). The verb *want* corresponds to a function of one internal argument, $f_i^?(1:\texttt{proposition})$.[18] The shared thematic function $f_{20}^0(\cdot)$ appears in (134b).

[18] The bare VP examples in (132) suggest that the PP headed by *before* should be viewed as an adjunct to the VP, rather than as an argument of the I0 event or as the innermost argument of the V0 function, as some have suggested (cf. Larson 1988, fn.11).

(132)
a. [[Run six miles]$_1$ before lunch]$_2$ is what Randy did.
b. [Do $\text{that}_{1/*2}$ after eating breakfast] is what Rod did.
c. [Do $\text{that}_{1/*2}$] is what Rod did after eating breakfast.

This is because the antecedent of the *that* pronoun, whatever it is, is not understood as including the PP. Hence the thematic function corresponding to *before* should be

(134)
a. Romeo$_1$ wanted Rosaline to [love him] before wanting himself$_1$ to [e]

b. [$_{\mathrm{VP}}$[$_{\mathrm{V0}}$love, $f^2_{20}(1:\mathtt{object})$] [$_{\mathrm{NP}}$him, 5],
$f^0_{20}(1:\mathtt{event}, 2:\mathtt{actor})$]

c. [$_{\mathrm{VP}}$[$_{\mathrm{V0}}$want, $f^2_{10}(1:\mathtt{proposition})$]
[$_{\mathrm{IP}}$[$_{\mathrm{NP}}$Rosaline, 3]
[$_{\mathrm{I1}}$[$_{\mathrm{I0}}$[-tns,+to], 4] [$_{\mathrm{VP}}$*love him*, $f^0_{20}(1:\mathtt{event})$],
$f^0_{20}(2:\mathtt{actor})$], 6],
$f^0_{10}(1:\mathtt{event}, 2:\mathtt{actor})$]

d. [$_{\mathrm{IP}}$[$_{\mathrm{NP}}$PRO, 7]
[$_{\mathrm{I1}}$[$_{\mathrm{I0}}$[-tns,+ing], 8]
[$_{\mathrm{VP}}$[$_{\mathrm{V0}}$want, $f^2_{30}(1:\mathtt{proposition})$]
[$_{\mathrm{IP}}$[$_{\mathrm{NP}}$himself, 9] [$_{\mathrm{I1}}$[$_{\mathrm{I0}}$[-tns,+to], 10] [$f^0_{20}(1:\mathtt{event})$],
$f^0_{20}(2:\mathtt{actor})$], 11],
$f^0_{30}(1:\mathtt{event})$],
$f^0_{30}(2:\mathtt{actor})$], 12]

e. [$_{\mathrm{IP}}$[$_{\mathrm{NP}}$Romeo, 1]
[$_{\mathrm{I1}}$[$_{\mathrm{I0}}$[+past], 2]
[$_{\mathrm{VP}}$[$_{\mathrm{VP}}$*want Rosaline to love him*, $f^0_{10}(1:\mathtt{event})$]
[$_{\mathrm{PP}}$before [$_{\mathrm{IP}}$*PRO want+ing himself to*, 12]]
$f^0_{10}(1:\mathtt{event})$],
$f^0_{10}(2:\mathtt{actor})$], 13]

The associated referential dependencies are:

(135)
a. obviate($f^2_{20}(1)$,$f^0_{20}(2)$), link$_1$($f^0_{20}(2)$,$f^0_{30}(2)$), link$_2$($f^0_{30}(2)$,$f^0_{10}(2)$)

b. obviate($f^2_{20}(1)$,$f^0_{20}(2)$), link$_1$(9,7), link$_2$($f^0_{40}(2)$,$f^0_{10}(2)$)

$f_{40}(1:\mathtt{event}, 2:\mathtt{event})$, and in the structure (134), $f_{40}(1)$ either theta-marks or theta-identifies $f^2_{30}(1)$, and $f_{40}(2)$ is theta-identified with $f^2_{10}(1)$. The same argument may be applied to manner adverbials, which raises difficulties for any theory that analyzes them as the innermost argument of the verbal function:

(133)
a. [[Run six miles]$_1$ quickly]$_2$ is what Randy did.
b. [Do that$_{1/*2}$ slowly] is what Rod did.
c. [Do that$_{1/*2}$] is what Rod did slowly.

The relation of obviation is between *him* and *Rosaline* (the other obviation relation, obviate($f^0_{20}(2)$,$f^0_{10}(2)$), between *Rosaline* and *Romeo*, has been suppressed for clarity). link_1 is from the anaphor *himself* to controlled *PRO*; link_2 is from controlled *PRO* to *Romeo*. By the semantics of the link relation, both arguments 9 and 7 must include the value of argument 1 in their values. By the semantics of the obviate relation, none of arguments assigned $f^2_{20}(1)$ (that is, argument 5) may share a value with any of the arguments assigned $f^0_{20}(2)$ (that is, arguments 3 and 9).

If we enforced coreference between *him* and *Romeo*, we would have to add a linking relation to either (135a) or (135b), either link(5,1) or link($f^2_{20}(1)$,$f^0_{10}(2)$). The effect of adding either link is to include the values of argument 1 in the values of argument 5; but then the arguments 9 and 5 share the values of argument 1, which is expressly forbid by the obviate($f^2_{20}(1)$,$f^0_{20}(2)$) relation.

B.5.4 Recursive ellipsis

The last of the tricky cases are the examples of recursive ellipsis. The example (35) from chapter 4, reproduced here as (136), has the three possible interpretations paraphrased in (137).

(136)
a. Jack_1 [[corrected his_1 spelling mistakes]$_i$ before the teacher_2 did $[e]_i]_j$ and Ted_3 did $[e]_j$ too.

(137)
a. Jack_1 corrected his_1 spelling mistakes before the teacher_2 corrected his_1 spelling mistakes and Ted_3 corrected his_1 spelling mistakes before the teacher_2 corrected his_1 spelling mistakes.
b. Jack_1 corrected his_1 spelling mistakes before the teacher_2 corrected his_2 spelling mistakes and Ted_3 corrected his_3 spelling mistakes before the teacher_2 corrected his_2 spelling mistakes.
c. Jack_1 corrected his_1 spelling mistakes before the teacher_2 corrected his_1 spelling mistakes and Ted_3 corrected his_3 spelling mistakes before the teacher_2 corrected his_3 spelling mistakes.

The purely invariant and covariant interpretations (137a,b respectively) present no problems for the standard identity-of-predication analysis. They are analyzed in the current proposal as linking and theta-linking. The mixed interpretation (137c), which has been considered indomitable evidence against identity-of-predication and for a copying operation, is also straightforwardly analyzed in the current proposal as linking with operator correspondence between *Jack* and *Ted*, after argument raising. This analysis is confirmed by the fact that both *Jack* and *Ted* must be heavily stressed before the interpretation (137c) becomes available.

B.6 Conclusion

The central technical goal of this chapter has been to contribute to the development of a linguistic theory that is explicit, appropriate, and maximally-constrained.

In order to be appropriate and maximally-constrained, the theory should not use a copy operation. One natural consequence of copying is that the original and its copy are *independent* with respect to subsequent processes. That is, once the original is copied, a subsequent processes will apply to both the original and its copy independently. But as we saw, the overt and null VPs are not truly independent in a VPE structure, and therefore the copy operation is inappropriate and should be allowed into a restrictive theory only as a last resort.

In order to be explicit, a linguistic theory must represent all perceivable linguistic distinctions. As Chomsky (1965:4–5) observes, "a fully adequate grammar must assign to each of an infinite range of sentences a structural description indicating how this sentence is understood by the ideal speaker-hearer. This is the traditional problem of descriptive linguistics. . . ." In general, distinct interpretations must correspond to distinct linguistic representations. In particular, all referential dependencies—such as perceived coreference between an anaphor or pronoun and its linguistic antecedent—must be represented by an explicit syntactic relation of coreference, even if the binding conditions are stated in terms of obviation. To do otherwise will result in a less than adequate theory.