The Philosophy of Generative Linguistics

Peter Ludlow

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First published 2011

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
MPG Books Group, Bodmin and King's Lynn

ISBN 978–0–19–925853–6

10 9 8 7 6 5 4 3 2 1

Contents

Preface ix
Acknowledgments xi
List of Abbreviations xiii
Introduction xv

1. Linguistic Preliminaries 1
   1.1 Transformational Grammar from ST to EST 1
   1.2 Government and Binding Theory 26
   1.3 The Principles and Parameters Framework 31
   1.4 The Minimalist Program 36
2. The Ontology of Generative Linguistics 44
   2.1 E-Language, I-Language, Ψ-Language 44
   2.2 Having Linguistic Rules and Knowing Linguistic Facts 48
   2.3 Levels of Explanation in the Theory of Grammar 55
   2.4 Abstracta and Non-isomorphic Representation 55
   2.5 Types and Tokens 58
   2.6 Derivation vs. Representation 60
3. Data, Intuitions, Judgments 64
   3.1 Linguistic Phenomena, Linguistic Data, Linguistic Theory 64
   3.2 Linguistic Intuitions are Linguistic Judgments 69
   3.3 Linguistic Judgments are Reliable (enough) 71
   3.4 Linguistic Judgments as Scientific Experiments 77
   3.5 On the Alleged Priority of the Data 82
4. A Role for Normative Rule Governance? 90
5. Worries about Rules and Representations 101
   5.1 Quinean Indeterminacy Arguments 102
   5.2 Kripke/Wittgenstein Concerns about Rules 105
   5.3 Externalism about Syntax? 117
6. Referential Semantics for Narrow Ψ-Languages 129
   6.1 The Compatibility of Referential Semantics and Narrow Ψ-Languages 129
   6.2 Chomsky's Incompatibilist Arguments 131
   6.3 The "Bite the Bullet" Strategy and Chomsky's Response 134

6.4 The Compatibilist Bites Back 137
6.5 The Prospects for a Non-referential Semantics 142

7. Best Theory Criteria and Methodological Minimalism 152
7.1 Simplicity Criteria 152
7.2 Formal Rigor 162
7.3 Minimal Effort and Optimal Switching Points 170

Appendix: Interview with Noam Chomsky 174
Bibliography 192
Index of Names 205
Index of Terms 208

Preface

This is not a book about everything that calls itself linguistics, or even everything that calls itself *generative* linguistics. There are many enterprises that go by the name 'linguistics' in the academy today, and no doubt philosophical issues arise for all of those enterprises. There are likewise many interesting aspects of generative linguistics that I can only touch on in passing here—generative phonology being a notable case in point. My goal in this book is to explore the philosophical questions that arise in the conduct of one particular branch of inquiry, beginning with the Standard Theory of Chomsky (1965) and extending to the *Principles and Parameters* (P&P) framework of current generative linguistics, understood broadly to include both *Government-Binding Theory* as initially articulated in Chomsky (1981) and the *Minimalist Program* as initially outlined in Chomsky (1995a).

I should note that my aim in this book is not in *arguing for* the generative framework so much as explaining its motivation, describing its basic mechanisms, and then addressing some of the many interesting philosophical questions and puzzles that arise once we adopt the general theoretical approach. Of course my interest in generative linguistics is driven by my belief that it has been and continues to be successful, but very little that I say here hangs on whether the program is ultimately true. Even the best theories are overthrown or modified beyond recognition, but that doesn't mean we shouldn't be interested in understanding and interpreting those theories and making sense of the practices being engaged in by those working within the theory, nor that we shouldn't be interested in understanding when those practices were successful and when they failed.

I won't address all of the interesting foundational questions that arise—for example whether the language faculty evolved slowly due to selectional pressures or whether it is a "spandrel" in the sense of Gould and Lewontin (1979), or whether linguistic structures might serve as the language of thought, or even questions about the proper place to draw boundaries between syntax, semantics, and pragmatics. Such questions are interesting to me, and there were even chapters devoted to these questions in earlier versions of this book, but in the end I decided that other questions came first—questions about the ontology of linguistics, about the nature of data, about language/world relations, and questions about best theory criteria. The issues addressed here are thus, in my opinion, more basic.

Of course many of the philosophical concerns that arise within generative linguistics arise in other areas of linguistics and indeed arise even in other areas of philosophical investigation, ranging from epistemology (for example the role of intuitions/judgments in theorizing) to ethics (the plausibility of being normatively guided by rules that

we don't have clear access to). Accordingly, those working in these areas are encouraged to join in the discussion. It is my hope that, among other things, this work will bring the philosophy of linguistics to a wider philosophical audience and show that we have many shared philosophical questions. Similarly, it is also my hope that the philosophical issues addressed here are laid out clearly enough so that linguists will feel comfortable engaging these issues as well. I appreciate the difficulties in speaking to both linguistic and philosophical audiences at the same time (or at least I do now). If I have had some success in this effort, this book will contribute to more frequent and more productive exchanges between linguists and philosophers on foundational issues.

1

Linguistic Preliminaries

1.1 Transformational Grammar from ST to EST

To a first approximation, linguistics is concerned with how humans pair sounds (or other perceivable forms) with meanings.[1] The theory of grammar is the theory of the mechanisms by which this is accomplished. We can think of there being three components to the grammar—a phonological component, a syntactic component, and a semantic component. The phonological component has to do with the representations that are relevant to the production and perception of speech (e.g. the combination of features by which vowels are formed, the metrical (rhythmic) patterns of the language, etc.). The syntax has to do with the form or structure of sentences (for example, that a sentence might consist of a subject and predicate or a noun phrase and a verb phrase). The semantic component has to do with the assignment of meanings to linguistic forms. For the most part in this book I will be concerned with the enterprise of generative *syntax* (although the relation to semantics will be explored in Chapter 5). Much could be written about the philosophy of generative phonology, but that will not be covered here. Furthermore, discussion of generative syntax that follows is not intended to be complete, nor even balanced to highlight the most important empirical results and discoveries. My mission in this chapter is simply to introduce those aspects of empirical research that raise the most interesting philosophical questions and those aspects that can inform our discussion in subsequent chapters.

Generative linguistics can best be appreciated if we examine it against the backdrop of American structuralism, in particular as articulated by Bloomfield (1933, 1939). American structuralism adopted a number of key assumptions from logical positivism, including the following.

(i) All useful generalizations are inductive generalizations.
(ii) Meanings are to be eschewed because they are occult entities—that is, because they are not directly empirically observable.

[1] Special thanks are due to Rob Stainton for comments on this chapter.

(iii) Discovery procedures like those advocated in logical positivism should be developed for the proper conduct of linguistic inquiry.
(iv) There should be no unobserved processes.

All of these assumptions were rejected at the inception of generative linguistics (see Chomsky 1975a: introduction, for a detailed discussion). As regards discovery procedures, for example, Chomsky rejected them while still a matriculating graduate student, then holding a position in the Harvard Society of Fellows:

> By 1953, I came to the same conclusion [as Morris Halle]: if the discovery procedures did not work, it was not because I had failed to formulate them correctly, but because the entire approach was wrong . . . [S]everal years of intense effort devoted to improving discovery procedures had come to naught, while work I had been doing during the same period on generative grammars and explanatory theory, in almost complete isolation, seemed to be consistently yielding interesting results. (1975a, 131)

More generally, Chomsky rejected methodological first principles, arguing that one ought to adopt whatever resources are at hand that work, expressing this idea as follows in the 1958 Texas Conference on Problems of Linguistic Analysis in English, in an exchange with James Sledd of UC Berkeley.

> CHOMSKY: The process that I use for investigating language is the one that I was taught. It is described in Harris' *Methods*. I use it for want of a better, though I know that it does not give me all the results that I want. When I can go beyond the method, fine, I go beyond it. Suppose I suddenly get a bright idea that such and such a reformulation will work better, I try it. I don't sit down and see how I could have gotten to the same result by defining the methods in a more and more rigorous way. If you don't have any bright ideas, then you simply experiment with the data, using the methods you have been taught. . . .
>
> . . .
>
> SLEDD: You are doing a kind of restatement linguistics, and so feel that you can make use of whatever others of us do, whenever it helps you to more knowledge of the language. You take what is useful, and put it into your machine. Isn't that so?
>
> CHOMSKY: Exactly. And I don't care how I get from data that someone has presented to me, to a useful formulation. Questions of that sort don't interest me, though they might be important for some other field, say the psychology of invention.

Generative grammarians also rejected the assumption that all processes should be "observable" processes—early theories of transformational grammar offering key examples of unobservable processes. For example, in the "Standard Theory" of generative grammar as developed in Chomsky (1965), the grammar is divided into two different "levels of representation"—initially called *Deep Structure* (hereafter DS) and *Surface Structure* (hereafter SS). Transformational rules map from DS representations onto SS representations, so that the picture of the grammar is something like the following.

$$\text{DS} \xrightarrow{\text{transformations}} \text{SS}$$

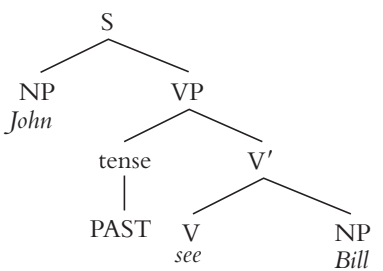
The names 'Deep Structure' and 'Surface Structure' were probably unfortunate in that they led to the imputation of properties to these levels of representation that were not intended.

To put it simply, DS representations are merely representations that are generated by what is known as a *Context-Free Phrase Structure Grammar*. Here is an example of a toy (emphasis on *toy*) context-free phrase structure grammar:

(i) S → NP VP
(ii) VP → tense V′
(iii) V′ → V (NP)
(iv) NP → *John*
(v) NP → *Bill*
(vi) NP → det N′
(vii) N′ → (adj) N
(viii) V → *see*
(ix) V → *like*
(x) tense → *PAST, PRES, FUT*

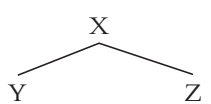
These phrase structure rules are in effect instructions for building (generating) tree representations for linguistic structures. For example, the following tree is derived from successive applications of rules (i–v) above.

(1)



From time to time we will find it useful to "linearize" trees like this for expository purposes, but the resulting structural descriptions will encode the same information. We thus take structures like (a) and (b) to be notational variants of each other.

(a)



(b) [x Y Z]

The linearized version of (1) above would thus be (1′).

(1′) [$_{S}$ [$_{NP}$ John][$_{VP}$ [$_{tense}$ PAST] [$_{V'}$ [$_{V}$ see][$_{NP}$ Bill]]]]

As a notational shortcut, I will sometimes omit structure within brackets. So, in cases where fine structure is unimportant, I might represent (1′) as in (1″), or something with

an intermediate amount of structure. The general principle is to show as much structure as necessary to make the point; omitting structure in a representation like (1″) is not to say the structure isn't there in the linguistic object being represented.

(1″) [S John saw Bill]

Crucially, for generative linguists the objects of analysis in linguistic theory were not the so-called *terminal strings* of words, but rather *phrase markers*—structured objects like (1) or (1′). As we will see, structures like (1) carry significantly more information than an unstructured string of words. For example, the adverb 'reluctantly' can be inserted between 'John' and 'saw', but not between 'saw' and 'Bill' (unless comma intonation is added suggesting that 'reluctantly' is a parenthetical aside). One possible explanation is that the VP forms a *constituent*, which cannot be split by the adverb.

Grammars that generate the same set of unstructured strings of words are said to have the same *weak generative capacity*. If two grammars generate the same structures (including tree structure) we say they have the same *strong generative capacity*. Generative linguistics has, since the beginning, been concerned with *strong* generative capacity.

As I said above, the simple toy grammar above is a context-free phrase structure grammar. This is a kind of grammar that is familiar to computer scientists, and the properties of such formal languages are fairly well understood. From the beginning, Chomsky has argued that such grammars are *not* sufficient to describe the structure of natural languages.[2] We could also add that such grammars do not provide us the resources to describe important relations holding between linguistic forms in natural language. For example, we might think that there is a kind of natural relationship between sentences like 'John saw Bill' and 'Bill was seen by John' and that we would like to capture this relationship. The nature of such relations can be formalized using *transformations*.

In the Standard Theory transformational rules operated on these DS representations (in effect, on the products of context-free phrase structure rules) to yield SS representations. Here are some examples of transformational rules.

Passivization: NP1 V NP2 → NP2 be-en V by NP1

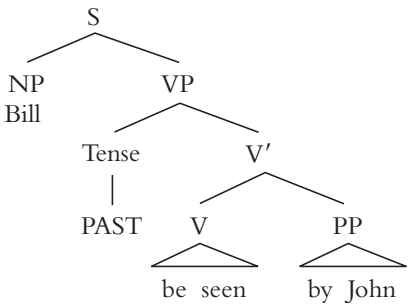
Affix Hopping: Aff V → V-Aff

Passivization converts a DS in active form to an SS representation that we would recognize as being in passive form. Affixes are elements like tense and aspectual markers, which (on Chomsky's early theory) are generated in a preverbal position at DS. Affix hopping moves them into a post-verbal position as suffixes. In this case, affix

[2] As with most claims in linguistic theory, this one has been challenged. Context-free phrase structure grammars can be generalized to account for phenomena like wh-movement and displaced constituents (see Gazdar et al. 1985), although there are arguably limits as to whether such grammars can account for all natural language constructions (see for example a worry about Swiss German raised in Shieber (1985)). I set this question aside here, as the debate does not impinge on the kinds of issues I'm addressing in this book.

hopping will move the affix '-en' from a preverbal position to one in which it is attached to the verb 'see', yielding 'seen'. So, for example, the successive operations of Passivization and Affix Hopping would take a deep structure representation like (1) and (abstracting from complications with tense) yield the surface structure representation in (2).

(2)



The standard written sentence in (3) below therefore designates a complex object consisting of (at a minimum) the ordered pair of the DS and SS representations corresponding to (1) and (2).

(3) Bill was seen by John

We could think of (3) as being our written shorthand convention for representing the ordered pair consisting of the DS representation given in (1) and the SS representation given in (2).

Other examples of transformations included the following (I'm indicating the location where the noun phrase is moved from for heuristic purposes; the idea that this location should have a syntactically significant element is a later development that will be discussed shortly).

Topicalization: [. . . NP . . .] → [NP[. . . _ . . .]]

For example from the DS [John likes Bill] we get [Bill[John likes _]]

Wh-movement: [. . . WH . . .] → WH did [. . . _ . . .]

For example from the DS [John saw who] we get [Who did [John see _]]

It is natural to think that the goal here was to formalize an intuitive connection between a passive and a non-passive form of a sentence (they certainly seem to be importantly related). However, a case can be made that this was not Chomsky's principal motivation—that the motivation was actually one of theoretical simplicity. Speaking of Chomsky's earlier (1957) book *Syntactic Structures* and the introduction of transformations there, Newmeyer (1986) argues as follows:

> Chomsky's arguments for transformation rules in Syntactic Structures were all simplicity arguments . . . They all involved showing that a grammar with phrase structure rules alone required

great complexity, a complexity that could be avoided only by the position of a transformational rule. (24)

We will return to the question of simplicity in Chapter 7. For now I want to draw attention to the fact that early generative linguists were committed not only to "unobserved" processes in the guise of transformations, but also to unobserved levels of representation.

No less significant than the introduction of unseen processes and levels of representation was the nature of some of the data that generative linguistics admitted—it was not limited to corpora of utterances or written strings, but also included speakers' judgments of acceptability and meaning. Thus, example (3) is not a datum because it has been written or spoken, but rather because speakers judge that it is (would correspond to) an acceptable linguistic form. Here again, generative linguists broke with prevailing methodology in structuralist linguistics and, indeed, behaviorist psychology, by allowing individual judgments as data (more on linguistic judgments in Chapter 3).

Having broken out of the behaviorist mindset of then prevalent linguistic theorizing, generative grammar subsequently evolved in response to a number of discoveries and explanatory pressures. Some of the most remarkable advances came in the form of Ross's 1967 Ph.D. dissertation, which catalogued numerous transformations, but crucially, Ross also observed (building on a proposal in Chomsky 1962) that certain syntactic environments seemed to block the transformations.

For example, while we can extract 'the book' from 'Bill illustrated the book' to yield 'that's the book that Bill illustrated', consider what happens when we attempt a similar extraction from the following:

(4) [$_S$ Bill married the woman who illustrated the book]

The resulting (5) is clearly bad.

(5) *[$_S$ That's the book that [$_S$ Bill married the woman who illustrated __]]

Ross observed a number of these effects (which he called "islands"—as in stranded on an island) and stated conditions on movement that would explain and predict these island effects.

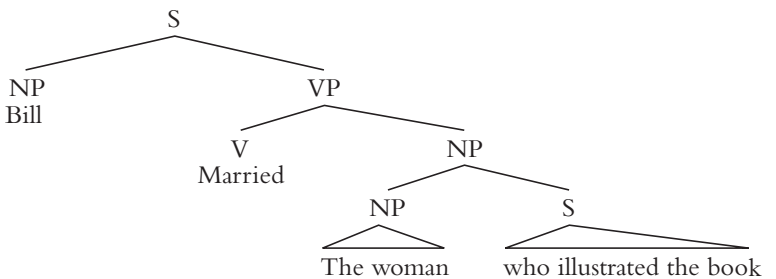
For example, he accounted for the above island (the discovery of which he attributed to Ed Klima) by positing the *Complex Noun Phrase Constraint*, which he formalized as follows:

(CNPC): No element contained in an S dominated by an NP with a lexical head noun may be moved out of that NP by a transformation. (Ross (1974, 178). Page numbers are from the excerpt of his dissertation in Harman (1974))

To explain, one node X dominates another node Y if and only if there is a direct downward path from X to Y in the tree. The CNPC says that if there is an S inside an

NP (for example a relative clause inside an NP) and the NP has a lexical head noun (in this case 'woman'), you cannot extract something from inside that S. Illustrating with (4′), we see that this is precisely what has happened. Forming a structure like (5) involves extracting 'the book' from inside an S (the relative clause 'who illustrated the book' which is contained inside an NP ('the woman who illustrated the book').

(4′)



Another such constraint was the *Coordinate Structure Constraint*:

(CSC): In a coordinate structure, no conjunct may be moved, nor may any element contained in a conjunct be moved out of that conjunct.

An example of a coordinate structure is a conjunction. To illustrate this constraint, consider the D-structure (6), and the resulting S-structure (7), where we have attempted to extract 'what sofa' via wh-movement.

(6) $[_S$ He will put the chair between $[_{NP}$ $[_{NP}$ some table] and $[_{NP}$ what sofa]]]

(7) *$[_S$ What sofa will he put the chair between $[_{NP}$ $[_{NP}$ some table] and $[_{NP}$ ___]]]

Ross took the constraints stated thus far to be universal—to hold across all human languages. But he also identified constraints for which this did not hold. One such constraint was the *Sentential Subject Constraint* (SSC):

(SSC): No element dominated by an S may be moved out of that S if that node S is dominated by an NP which itself is immediately dominated by S. (1974, 194)

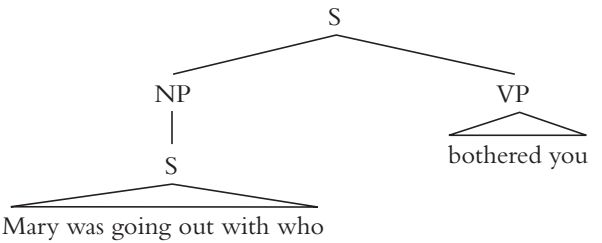
This was designed to block movement that would result in forming (9) from the DS representation (8):

(8) $[_S$ $[_{NP}$ that [S Mary was going out with who]] bothered you]

(9) *$[_S$ Who did $[_S$ $[_{NP}$ that Mary was going out with __]] bother you]]

The idea is the following: Ross is thinking of the that-clause—'that Mary was going out with who' as an NP. This NP is in turn dominated by an S, as illustrated in the following tree.

(10)



Getting the 'who' to extract requires jumping out of an S that is dominated by an NP that is in turn immediately dominated by an S. As noted above, Ross did not consider this constraint to be universal.

> That the languages whose rules I know to be subject to [SSC] far outnumber those whose rules are not so constrained suggests that a search be made for other formal properties of these latter languages which could be made use of to predict their atypical behavior with respect to this constraint. At present, however, whether or not [SSC] is operative within any particular language can only be treated as an idiosyncratic fact which must be stated in the conditions box of the language in question. (1974, 194)

Ross was noting cross-linguistic variation with respect to this constraint but he was also indicating a belief that variation across languages with respect to such properties ought to have some explanation (a theme we will return to when we discuss the Principles and Parameters framework in section 1.3).

Generative Semantics

Starting in the 1960s some generative linguists hypothesized that DS representations might be generated via *projection rules* from a level of representation which we might call Semantic Representation (or SR), that encoded meanings, or more accurately, meaning representations. According to this idea, the model of the grammar would be something like the following:

(11) SR — $_{\text{projection-rules}}$ → DS — $_{\text{transformations}}$ → SS

One central idea that crystallized with the Katz-Postal hypothesis (after Katz and Postal 1964) was the idea that the DS representations might be generated from word and sentence meanings. Thus, if two sentences had the same meaning they must have the same DS representation. If there is an ambiguity, it would correspond to there being distinct DS representations. Here is the statement from Katz and Postal (1964, 157).

> Given a sentence for which a syntactic derivation is needed; look for simple paraphrases of the sentence which are not paraphrases by virtue of synonymous expressions; on finding them, construct grammatical rules that relate the original sentence and its paraphrases in such a way that each of these sentences has the same sequence of underlying P-markers [DS representations]. Of course, having constructed such rules, it is still necessary to find *independent syntactic justification* for them. [emphasis theirs]

Probably in the background of this proposal was work on the computational theory of mind and the language of thought hypothesis which Jerry Fodor was developing at the time and which is fully articulated in Fodor (1975).

Subsequently, a number of linguists, including Ross, George Lakoff, Paul Postal, and others, developed a research program that extended this idea in fairly radical ways. The research program came to be known as *generative semantics*, and the resulting battle with Chomsky and other generative linguists has come to be known as the "Generative Semantics Wars." Although the conflicts led to some interesting reading (see Newmeyer 1986 and Harris 1993) only a couple of elements of that dispute are relevant to us.

One element concerns the range of the phenomena to be explained by the theory of grammar (we could also call this the range of the *explananda* of the theory of grammar[3]). Even if we are interested in a broad class of language related phenomena, it need not be the case that the principal object of investigation (the theory of grammar) should explain everything in this pre-theoretical domain of investigation. That is, even if we agreed that there was a broad range of language related phenomena to be explained, it does not follow that there is a single mechanism (the theory of grammar) serving as the sole *explanans* for all of them.

This point became salient in the generative semantics era because of a tendency of generative semanticists to attempt to sweep up more and more language related phenomena and take them to be targets of explanation (explananda) for the theory of grammar. Up until that point, the kinds of facts that were the targets of explanation in generative linguistics were fairly constrained, but particularly under Lakoff's influence the supposed domain of linguistics became broader and broader, and the theory of grammar was pressed into service as the potential explanans for more and more phenomena. In part, the generative semantics wars can be seen as a dispute over precisely how broad the target of theory of grammar should be.

In the end, any phenomenon remotely related to language was taken to be an explanandum of the theory. Synonymous expressions were derived from a common underlying structure (famously, 'kill' from 'cause to die'). Selectional restrictions were also argued to have syntactic explanations. Thus the oddness of 'the hot dog ate John' was not because of real world knowledge, but had an underlying syntactic/semantic explanation (McCawley 1968, Lakoff 1968, Lakoff and Ross 1976). Pronouns were derived from noun phrases by the operation of pronominalization. Soon generative semantics would be extended to account for logic, and then fuzzy logic was incorporated into the theory to account for graded judgments (Lakoff 1973). Lakoff, in Parrett (1974, 151) showed just how broadly the net was to be cast. The theory was going to accommodate

[3] My use of the term 'explananda' and 'explanans' diverges from that of, for example, Hempel, since he took the explananda and explanans to be the linguistic descriptions of the thing explained and the explanation. I'm using them to speak of the phenomena themselves.

not just syntax-semantic, phonetics-phonology, historical linguistics, anthropological linguistics, etc., which form the core of most academic programs in this country, but also the role of language in social interaction, in literature, in ritual, and in propaganda, and as well the study of the relationship between language and thought, speech production and perception, linguistic disorders, etc.

Or to put it another way, when you have a really good hammer, everything begins to look like a nail. I think it is one thing to suppose that generative linguistics intersects with these areas, but the idea that it will subsume them is an entirely different matter. In my view, matters were taken to their limit when Ross (1970) and Sadock (1969, 1970) proposed that pragmatic phenomena (and speech act phenomena) could be treated as syntactic phenomena. Thus, speech acts like assertion were analyzed as involving a syntactic operator 'I assert that'. Anyone familiar with the story of Achilles and the Tortoise will see how this strategy is bound to fail; one would need a speech act theory for the new syntactic form, which presumably would receive a syntactic account, and a regress has begun.

It is interesting to reflect on the basic assumptions about the nature of inquiry driving the generative semanticists. Of course, once one takes it to be the job of generative linguistics to subsume all of these phenomena, one quickly sees the task as hopeless. At that point, one either becomes less ambitious about the goals of generative linguistics, or settles into a kind of pessimism about generative linguistics as did Postal (1976, 203):

> This is the first in a random, possibly nonfinite series of communications designed to show beyond any doubt that there exists no linguistic theory whatever. There are apparently endless numbers of fact types not incorporable within any known or imaginable framework. In particular, what has been called the theory of transformational grammar, seems to have only the most partial relation to linguistic reality.

In the same essay Postal continues on this theme, arguing that science requires an *a priori* statement of the facts that are to be the target of the theory:

> Many people today are engaged in the attempt to construct linguistic theories. My view is that an important difficulty with all such attempts is that there is not a good a priori statement of the full range of known facts which a theory must handle. To the extent that theories are formulated in the absence of explicit awareness of this range of facts, they are dreamlike. (205)

I plan on returning to this point in my discussion of data-first approaches to linguistics in Chapter 3, but I don't mind giving away the punchline now: This attitude is deeply confused. Science begins with some pre-theoretical domain of interest, but the phenomena of interest in this domain are certainly not complete or set in stone (certainly not a priori!) and the phenomena of interest can change radically as the theory progresses. Furthermore, many phenomena which we initially take to be the target of our theory turn out not to be explananda of the theory after all. Consider, for example, the case of polio, which I will discuss in Chapter 3; many of the phenomena

that we initially called "polio" turned out to be caused by something other than the polio virus.

The generative semantics research program collapsed rapidly (it was probably out of steam by 1972), and how could it not collapse given the kind of sweeping set of phenomena that they tried to bring under the theory of grammar. It is not my intention to get into the sociology of science, but there is a very interesting question about the pros and cons of their effort. On the one hand some of the most brilliant linguists of the last 50 years sank with the program, and not all of them contributed to generative linguistics again after the demise of the program. On the other hand, they did push the envelope and while not everything they explored could be incorporated under the umbrella of generative linguistics, a surprising amount of it subsequently has been.[4] The point being that perhaps we want some members of the research community to try and expand the target of investigation, even if the effort is philosophically misguided and on balance a failure.

I don't mean to make it sound like generative semantics failed simply because its advocates bit off more than they could chew or that their theory was unconstrained or that the Katz-Postal hypothesis was anchored in some very suspect philosophy of mind (all or which are probably true). There were finely targeted empirical objections as well. For example, in 1967 Chomsky presented a talk, which was later published in 1970 as "Remarks on Nominalization," which sat like a time bomb set to go off as soon as people began to read the work and follow its line of reasoning.

We can set up the "Remarks" argument like this: It is the claim of generative semanticists that sentences having similar meanings are derived from a common structure. For example, generative semanticists took structures having nouns like 'refusal' and 'belief' to be transformationally derived from structures having the verbs 'refuse' and 'believe'. This is certainly a natural thing for a generative semanticist to think—the words have closely related meanings, and if similar meanings imply a common source (at DS or some level of semantic representation) then what could be more natural than to think than the nouns are transformationally derived from the verbs?

[4] Personally, I've been deeply influenced by their work in natural logic (see Ludlow 1995, 2002b), and have collaborated with Richard Larson and Marcel den Dikken (1997) on a proposal due to Ross and McCawley that there is an implicit clause (and verb 'have') in a sentence like 'John wants a unicorn'—that is that at some level of representation it has a structure like [$_S$ John wants [$_S$ John has a unicorn]]. The difference is that while Ross and McCawley took the superficially simpler structure to be derived from the more complex one (which they took to be its D-structure), we took it to be a positive thesis about its SS representation.

The difference between the newer proposals and the older ones tracks the difference between the generative semanticists and the so-called "interpretivists" of that period. The former took syntactic forms to be generated from, in effect, the language of thought (in the sense of Fodor 1975). The interpretivists argued for a kind of autonomy of syntax and argued that interpretation would take place on SS representations—or perhaps a level of representation that was downstream from SS. And of course, the most important element to all of this is that the mapping from DS to SS must be tightly constrained.

In "Remarks," Chomsky makes the case that some nouns come from verbs, but that not all of them do. So, Chomsky was happy to argue that gerundive constructions like 'John's believing in God' is derived, but 'John's belief in God' is not. Similarly 'John's refusing the offer' is derived, while 'John's refusal of the offer is not'.

Here is the argument.

First, gerundive nominals occur with aspectual verbs while so-called derived nominals cannot. Thus consider the contrast between (12) and (13).

(12) John's having criticized the book
John's having been refusing the offer just when Trump arrived.

(13) *John's having criticism of the book
*John's having been refusal of the offer

Second, gerundive nominals take adverbs before the gerund while so-called derived nominals take adjectives.

(14) John's sarcastically criticizing the book
John's emphatically refusing the offer

(15) John's sarcastic criticism of the book
John's emphatic refusal of the offer

Third, the gerundive nominals occur in subject raising and dative shift, but the same is not true of the so-called derived nominals.

(16) its being certain that John will win (subject raising) →
John's being certain to win

John's giving the book to Bill (dative shift) →
John's giving Bill a book

(17) the certainty that John will win ↛
*John's certainty to win

John's gift of a book to Bill ↛
*John's gift of Bill a book

If 'certainty' and 'gift' are transformationally derived, why don't they behave like the gerundives do?

On top of all that, the relation between so-called derived nominals and their corresponding verbs is highly irregular. Not every verb has a corresponding noun. Why should this be if this is supposed to be a productive process? Furthermore, when we do have verb–noun correspondences the meaning relation is often obscure (consider 'do'/'deed' and 'ignore'/'ignorance').

Other linguists weighed in with additional arguments. For example Bach (1970) took aim at the pronominalization transformation in his paper "Problominalization"

pointing out that it would lead to an infinite regress with so-called Bach-Peters sentences like (18).

(18) Every pilot who shot at it hit some MiG that chased him

It seems that the DS for this would have to be (18′).

(18′) Every pilot who shot at *some MiG that chased him* hit some MiG that chased *every pilot who shot at it*.

But these can't be the base representations, because they too have pronouns (which by hypothesis were derived) and so we are off on a regress of exponentially larger base forms.

Jerry Fodor (1970) also got into the mix with his paper arguing against "kill" being derived from "cause to die". One simple version of his argument is this: If you attach a temporal adverb to the biclausal structure you get an ambiguity, but you do not get such an ambiguity if you attach it to the single clause structure, as the following contrast shows.

(19) Uma caused Bill to die yesterday

(20) Uma killed Bill yesterday

Sentence (19) could be true even if Bill died today and not yesterday (perhaps Uma administered a slow acting poison yesterday).

Returning to "Remarks on Nominalization," probably the most significant contribution in Chomsky's paper was the introduction of so-called *X-bar theory*—a proposal that at once gave an account of the apparent similarity between verb phrases and noun phrases, offered a natural way of stating rules and generalization that were cross-categorial, and ultimately led to a way of simplifying the phrase structure component of the grammar.

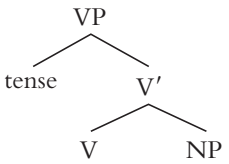
Recall that in the Standard Theory the DS representations are generated from a series of phrase structure rules like the following.

VP → tense V′
NP → det N′

The insight of X-bar theory was that you could state these kinds of relations in a more general way if you thought in terms of phrases, heads, specifiers, and complements.

For example, obviously enough, VP, NP, AP, and PP indicate *phrases* (verb phrase, noun phrase, adjectival phrase, and prepositional phrase respectively). The categories V, N, P, and A are heads. The complement of the phrase is another phrase. So, for example, consider the structure in (21).

(21)



In this structure, V is the head, NP is the complement, and tense is the specifier.

Chomsky argued that we can generalize these structures in the following way, where X is any head (V, N, A, or P):

$S \rightarrow N'' V''$
$X'' \rightarrow$ Spec-$X' X'$
$X' \rightarrow X \ldots$ ('. . .' indicating possibly another X^n phrase, for example an NP or PP)

Now here comes a very interesting insight. We can state important constraints on phrase structure and movement rules by having them apply to X′ level taxonomy—that is, we can state general constraints that only make reference to heads or specifiers or complements. For example, one generalization that we will return to when we look at the Principles and parameters framework is the idea that languages tend (with obvious exceptions like German) to be either head-initial (head to the left) or head final (head to the right).

Additionally, Chomsky suggested that the traditional categories of V, N, A, and P might be analyzed as being bundles of features—to a first approximation ±N and ±V. The resulting paradigm is the following:

	+N	−N
+V	A	V
−V	N	P[5]

The use of N and V to describe the features is probably confusing—Chomsky could just as easily have used 'foo' and 'bar' as labels for these features. Indeed, subsequent authors such as Jackendoff (1977) explored the idea that the features were ±*subject*, ±*object*, and ±*det*. The point of interest for us is that Chomsky was effectively proposing a kind of sub-atomic approach to syntactic categories (this wasn't new in generative *phonology*—phonemes had long been thought of as bundles of features). On the one hand this took the description of the property of language ever further away from the kinds of properties that were in some sense directly perceivable, but on the other hand it allowed linguists more easily to state generalizations that cut across traditional grammatical categories—some linguistic properties might apply to elements that are –N, for example (we will see an example of this in a bit).

It is interesting to reflect on the fact that Chomsky's work was moving in this "lexicalist" direction at a time when generative semanticists continued to expand the

[5] This cell is blank in Chomsky's original proposal, but subsequent work filled out the paradigm.

expressive power of their transformations (or projection rules) to incorporate larger and larger classes of data. Chomsky was clearly moving in the opposite direction; far from extending the transformational component to phenomena like synonymy and selectional restrictions, he was putting the brakes on its application even to nominals like 'refusal' and 'criticism'—nominals that were *considered* to be so obviously the result of transformational processes that it was packed into their labels: "derived nominals"!

While generative semanticists were taking the tool of transformations and extending it in less and less restricted ways to broader classes of phenomena, Chomsky was looking for probes to test the limits of that tool. "Remarks on Nominalization" was an effort to show some of those limits. As it turns out, "Remarks" was only the beginning.

The Extended Standard Theory

Ross's (1967) work on island constraints ushered in a gold rush of researchers looking to discover new transformations and island effects. But the gold rush, while productive, had a down side.

The number of transformations had begun to proliferate in a way that generative linguists found troubling. Why troubling? Early on in the development of generative grammar, Chomsky had made a distinction between the *descriptive adequacy* and the *explanatory adequacy* of an empirical linguistic theory (see Chomsky 1965, 1981, 1986). In particular, if a linguist theory is to be explanatorily adequate it must not merely describe the facts, but must do so in a way that *explains* how humans are able to learn languages; linguistics was supposed to be embeddable into cognitive science more broadly. But if this is the case then there is a concern about the unchecked proliferation of rules—such rule systems might be descriptively adequate, but they would fail to account for how we acquire a language-specific grammar (perhaps due to the burden of having to learn all those language-specific rules).

A broad range of researchers in the field thought that the Ross constraints were simply too random a collection to be plausible, and they sought ways to unify them under some simpler explanation. Not surprisingly, Chomsky was in the middle of this effort, casting it as a way of finding more general conditions on transformations (i.e. general constraints on movement) with the goal of reducing the complexity of the grammar. One ingenious hypothesis advanced in Chomsky (1973) was that we might be able to cover a large number of the facts by replacing *all* the specific transformations with a single movement rule (initially it was *move-NP* but this would soon be generalized to "move anything anywhere") and replacing all (or at least most of) the island constraints with a handful of abstract general constraints—to a first approximation, the following three constraints on movement:

Subjacency: Very roughly, an element cannot cross both an S and an NP node in a single movement. This is effectively a generalization of the Complex Noun Phrase Constraint discussed above. Consider (5′) which is (5) represented at a finer level of detail. Notice that both an NP and an S node have been jumped in the extraction from the complex noun phrase 'the woman who illustrated the book':

(5′) ⋆That's the book . . . [$_{NP}$ the woman who [$_{S}$ illustrated ___]]

Specified Subject Condition:

No rule can involve *X, Y* in the structure
$\ldots X \ldots [_{\alpha} \ldots Z \ldots \text{-}WYV \ldots]$
where *Z* is the specified subject of *WYV* in α.

This can be thought of as a generalization of the Sentential Subject Condition, which accounted for the contrast between (8) and (9), and extends also to the contrast between (22) and (23), where (23) has a specified subject and (22) only a null subject. Hence in (23) 'the candidates' cannot be linked with 'each other'.

(22) The candidates expected [PRO to defeat *each* other].

(23) ⋆The candidates expected [the soldier to defeat each other].

The Specified Subject Condition also accounts for the contrast between (24) and (25), on the hypothesis that 'John' is behaving as a kind of intervening subject in (24), but not in (25):

(24) The men saw [$_{NP}$ the pictures of each other]

(25) ⋆The men saw [$_{NP}$ John's pictures of each other]

Tensed-S Condition:

No rule can involve *X, Y* in the structure
$\ldots X \ldots [_{\alpha} \ldots Y \ldots]$
where α is a tensed sentence.

The tensed-S Condition accounted for the contrast between (26) and (27) by blocking the possibility of a binding relation between 'the candidates' and 'each other'. It does this because the lower clause "that each other would win" is tensed.

(26) The candidates expected each other to win.

(27) ⋆The candidates expected that each other would win.

This version of the theory became known as the *Extended Standard Theory* (EST). Of particular interest to the philosophy of linguistics, I think, is the fact that these constraints were becoming much more distant from the routine descriptions of language that we learn in grammar school (involving words and word order and theoretical notions like subject and predicate), so that while there were fewer constraints, it seemed ever less plausible that language learners *consciously* learned them. It became clear that the kinds of rules deployed by language users were not learned at mother's knee, but more plausibly prewired in us, or were perhaps artifacts of other prewired aspects of our cognitive architecture.

Two subsequent additions to the EST led to the criticism that, with those additions, EST was effectively recapitulating generative semantics efforts to link syntactic forms with meanings. These developments were the introduction of trace theory and the introduction and development of a level of representation called LF (to suggest logical form).

Earlier we represented the site where a moved constituent had been with an underscore "_", and as I indicated at the time this was just for heuristic purposes. A number of linguists (e.g. Postal 1970, Ross 1969, Emonds 1970) had suggested that there might be some benefit to leaving something behind—an empty node (Emonds) or a "doom marker" (Postal and Ross).

Chomsky (1973) proposed treating the extraction site as being a "trace" that would be bound either by the moved WH word or the quantified noun phrase that moved out of the position. This idea was developed in Wasow (1972) and Chomsky (1975b).

Philosophers of course immediately like this idea, since it makes SS representations look plausibly like tree-shaped versions of formulae of first order logic. In fact, Reinhart (1976) argued that one could define a notion of syntactic scope off of these tree structures using the notion of c-command as follows.

Let's say that a node α in a tree dominates a node β just in case there is a direct downward path from α to β. Then a node α in a tree *c-commands* a node β just in case the first branching node dominating α dominates β. Or to put it in simple terms, if you want to know what a node α c-commands, you go to α's mother node; everything below that mother node is c-commanded by α. The next step is to hypothesize that the c-command relation effectively plays the same role as scope in formal logic. In other words, perhaps c-command just is scope.

You can see why some linguists thought that this might be a jazzier form of generative semantics. It looks like the new trace-laden structures are being introduced in order to account for meanings. Of course, the problem with generative semantics was never that it attempted to account for meanings. From the beginning, generative linguists had been explicit about wanting to account for form-meaning pairs. This was one of the key points on which they had broken away from American structuralism after all. The problem with generative semantics was not the interest in meanings, but rather that the mechanisms were wildly unconstrained, the class of phenomena targeted by the theory was too large, and the Katz-Postal hypothesis just wasn't working.

Still, if trace was to be introduced, it would have to be empirically fruitful—that is, it would have to integrate with the explanatory structure of the syntactic theory and yield some theoretical discoveries in the bargain. As it turns out, the introduction of trace met both of the desiderata. Examples of this integration included *weak crossover* and *movement asymmetry*.

Crossover facts had been initially noticed in Postal (1971): In example (28) 'who' and 'him' can be understood as coreferential while in (29) this is not possible.

(28) Who said Mary kissed him?

(29) Who did he say Mary kissed?

Wasow's contribution was to note that these facts could be explained by the introduction of trace theory. When we look at these structures with trace introduced, we get the following distribution of facts.

(30) Who_i [e_i said Mary kissed him]

(31) *Who_i [he_i say Mary kissed e_i]

We can then hypothesize that the coreferential reading in (31) is blocked for the same reason it is blocked in a sentence like (32).

(32) *He_i said Mary kissed John_i

One early way of characterizing this is as the "leftness condition" (Chomsky 1976): A variable cannot be the antecedent of a pronoun to its left (as it would have to generate the missing binding relations in (31) and (32)).

Meanwhile, Fiengo (1977) argued that a large class of movement asymmetries could be accounted for if we thought of traces as behaving like bound anaphors. For example, contrast (33) and (34).

(33) John likes himself

(34) *Himself like John

The general account for the behavior of reflexives is that they must be c-commanded by their antecedents (within a particular local domain). Well, if traces must also be c-commanded, then any movement that involved moving an element lower in a clause would result in a structure that violates this principle, because if an operator went lower it would no longer c-command its trace.

Chomsky had gone out on a limb by proposing that all the transformations could be swept away in favor of a rule that one can move anything anywhere. Fiengo's observation helped vindicate this proposal by showing that with the introduction of trace and independently motivated principles governing reflexives, one can immediately rule out an entire class of movements—i.e., downward movement.

Quite apart from whether the introduction of trace theory was leading the field back to generative semantics, some linguists worried about the introduction of so-called "phonetically unrealized syntax". Of course, strictly speaking, *no* syntax is phonetically realized. (We don't pronounce S or NP or VP and even the phonetic realization of word and sentence boundaries is dubious—words are often resyllabified, as when Long Islanders put the g in 'Long' with 'Islanders' as though saying 'Lawn Guylanders'). But some linguists did (and still do) have issues with the idea that there could actually be inaudible syntactic objects like traces. Linguists continue to offer approaches that attempt to avoid traces. These range from Cooper (1982)—titled "Binding in

Wholewheat* Syntax (*Unenriched with Inaudibilia)" to recent work by Jacobson (1999). Now of course traces *can* be eliminated (just as Quine (1960b) showed that variables could be eliminated in first order logic). The question is whether there is good reason to eschew so-called inaudibilia in the syntax.

Concerns like this raise an issue that we will return to in Chapter 2, which is the question of whether there really is a deep and important difference between linguistic structures that have inaudibilia like trace and those that do not. As I will try to make clear in the next chapter, I think these worries about inaudibilia are largely misplaced, since they rest on a confusion about the nature of representations in cognitive science. As I will argue, the case for representations is not secured by whether they are conservatively faithful to what we imagine the articulated form of utterances must be, but rather by whether they play a helpful role in theory construction. In particle physics we don't posit quarks, spin, charm, and superstrings because they are faithful to what we see, but because they play valuable roles in our theories.

Similar considerations arise with respect to the distinction between theories that rely heavily on constraints on representations and those that rely heavily on constraints on movement. As I will argue in Chapter 2, it is an error to think that one of these approaches is conceptually more sound within cognitive science; we ought to adopt whatever approach makes our theorizing the easiest (this is the principle of methodological minimalism that I alluded to in the introduction). This becomes salient when we consider the way in which Chomsky and other generative linguists have shifted between constraints on representation and constraints on movement with relative ease.

For example, Chomsky (1976) ("Conditions on Rules of Grammar") argued that the presence of trace would allow linguists to radically rethink the constraints on movement that they had been proposing (e.g. Chomsky himself in 1973). In effect, most of the constraints on movement could be eliminated and replaced with constraints on representations. Indeed, as Chomsky saw it, the only real constraint on movement needed was subjacency. The rest of the work could be done by the specified subject condition, applying it now to traces.

This approach to constraints on representation was expanded in Chomsky's (1980b) paper "On Binding", which reformulated the constraints as follows.

The Opacity Condition (formerly the SSC): An anaphor cannot be free in the domain of the subject of NP or S′.

The Tensed-S Condition (renamed the Propositional Island Condition)

The Nominative Island Constraint: A nominative anaphor cannot be free in S′.

Here Chomsky also observed that given these constraints you could make the movement rule even more general—calling it *move-α* instead of *move-NP*. In effect, you can move anything anywhere and the constraints on representation (and the subjacency constraint on movement) would do the necessary work.

One of the interesting features of the Extended Standard Theory was the introduction of the level of representation LF (suggesting a similarity to the philosopher's notion of logical form).

DS → SS → LF

↓

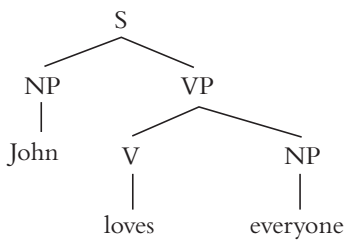
Phonetic Form

The level LF involved a rule mapping from SS to LF. Called QR, the rule simply said "adjoin quantified NP to S)." (May 1977)

For example, consider sentence (35) and its SS representation (35-SS).

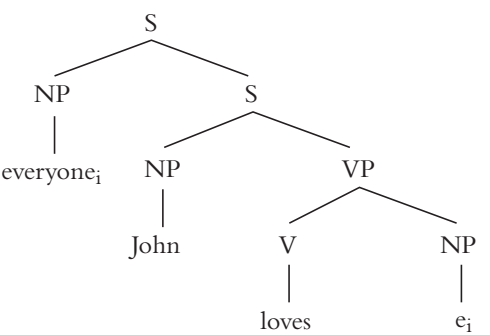
(35) John loves everyone

(35-SS)



To generate the LF representations we adjoin the NP to the topmost S node (creating a new S node) and leaving behind a co-indexed trace—in effect a bound variable.

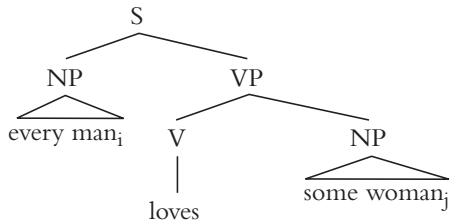
(35-LF)



Over the next decade or so a number of arguments were offered in support of QR and LF. Quite naturally, it was seen as a way of providing structural representations that could account for quantifier scope ambiguities. For example, consider sentence (36) and its SS representation.

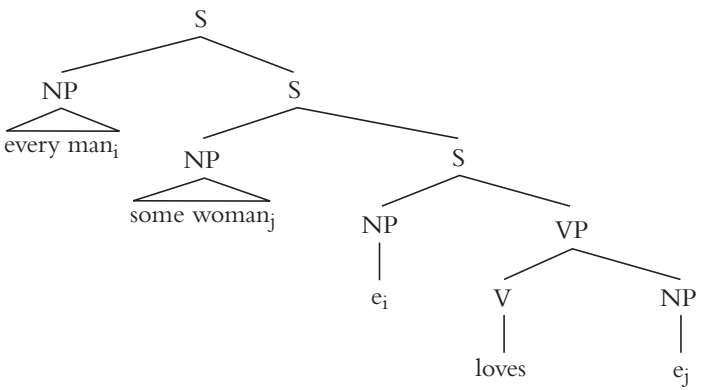
(36) Every man loves some woman

(36-SS)

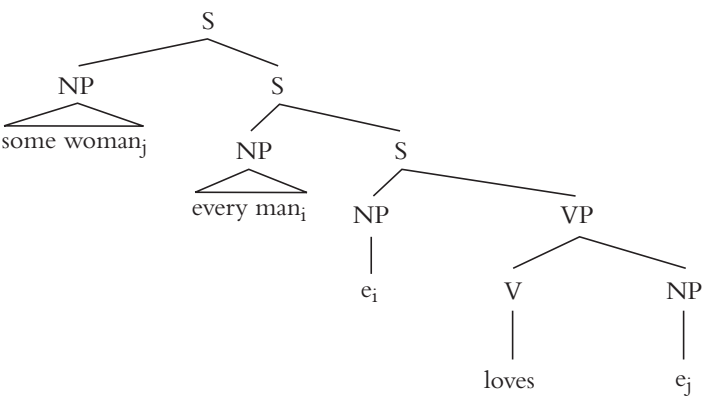


Given that either NP could raise first, this predicted two possible LF structures for the sentence, as indicated in LF1 and LF2.

(36-LF1)



(36-LF2)



Another argument for QR was that it could account for de re/de dicto ambiguity. To see this, consider (37) and the two resulting scope LF representations (37-LF1) and (37-LF2). In the case of (37-LF1) the quantifier takes scope outside of the attitude verb and generates the de re reading. In (37-LF2) it retains scope inside the attitude verb (by adjoining to the lower S) yielding the de dicto reading.

(37) John believes that a man is following him

(37-LF1) de re: $[_S$ (a man)$_i$ $[_S$ John believes that $[_S$ e_i is following him]]]

(37-LF2) de dicto: $[_S$ John believes that $[_S$ (a man)$_i$ $[_S$ e_i is following him]]]

Other arguments were more subtle. For example, consider the crossover facts adduced by Higginbotham (1980).

Higginbotham argued that there was a contrast between (38) and (39), in that (interpreting the sentence with binding relations as specified by the indices) we judge the interpretation corresponding to (38) to be awful, but the interpretation corresponding to (39) significantly better (actually, the judgments about (39) have been disputed, an issue that we will return to in Chapter 3; for now just let me say that if you are having trouble getting the reading it helps to stress 'saw').

(38) *Who$_i$ did his$_i$ sister see e_i

(39) His$_i$ sister saw John$_i$

Higginbotham suggested that the relevant generalization here was the again leftness condition of Chomsky (1976), which, recall, was that "a trace cannot be the antecedent of a pronoun to its left". Given this generalization, consider (40).

(40) His sister saw everyone

This sentence cannot mean that everyone is such that his sister saw him. The explanation is that if we propose QR, the leftness condition will automatically rule out (41).

(41) *[everyone$_i$ [his$_i$ sister saw e_i]]

Cross-linguistic evidence for LF was introduced by Huang (1982). Suppose that in Chinese, as in English, the following facts hold:

(42) 'believe' takes only [–WH] complements (that is, complement clauses without WH words)
I believe [Bill left]
*I believe [who left]
Who do you believe [e left]?
*Do you believe [who left]?

(43) 'wonder' takes only [+WH] complements
*I wonder [Bill left]
I wonder [who left]
* Who do you wonder [e left]?
Do you wonder [who left]?

(44) 'know' takes both kinds of complements
I know [Bill left]
I know [who left]
Who do you know [e left]?
Do you know [who left]?

Now consider the following facts from Chinese.

(45) Zhangsan xiang-zhidao [ta muqin kanjian shei]
Zhangsan wonder his mother see who

This must be interpreted as an indirect question as in (45′).

(45′) 'Zhangsan wondered [who his mother saw e]'

(46) Zhangsan xiangxin [ta muqin kanjian shei]
Zhangsan believe his mother see who

This must be interpreted as a direct question as in (46′).

(46′) 'Who did Zhangsan believe [his mother saw e]'

(47) Zhangsan zhidao [ta muqin kanjian shei]
Zhangsan knows his mother see who

This may be interpreted as either an indirect or a direct question, as in (47′–47″).

(47′) 'Zhangsan knows [who his mother saw e]'

(47″) Who does Zhangsan know [his mother saw e]'

Huang's insight was that these are precisely the results we would expect if we supposed that the WH elements were moving between SS and LF. The WH elements are in the lower clause at SS, but if the verb demands a –WH complement, the WH element is forced to move at LF, yielding a direct question interpretation. If the verb demands a +WH complement, the WH word must stay in place yielding an indirect question interpretation, and if it allows both complements then both readings are possible. I won't go into details here, but Huan noted other interesting effects of this invisible operator movement, including the weak crossover effects we noted above.

Perhaps the most compelling argument for LF came from the phenomenon of "antecedent contained deletion," initially noted in Ross (1969), Bouton (1970), and Sag (1976). Consider (48).

(48) John suspected everyone that Mary did

If the deleted VP is simply reconstructed, we get the following.

(49) John suspected everyone that Mary suspected everyone that Mary did

But now we have begun an infinite regress. An ellipsed VP must be inserted for 'did' again (presumably the VP 'suspected everyone that Mary suspected everyone that Mary did), leaving yet another ellipsed VP to be reconstructed.

Adapting a proposal from Sag (1976), May (1985) proposed the following. Suppose that QR takes place before the VP is reconstructed, so that the application of QR to (49) is (49′):

(49′) (everyone$_j$ that Mary did)$_i$ John suspected e$_i$

After reconstruction of the VP we now get (49″) which has just the right interpretation (everyone who is such that Mary suspected them, is such that John suspected them):

(49″) (everyone$_j$ that Mary suspected e$_j$)$_i$ John suspected e$_i$

As Larson (1987) showed, this can be extended to free relatives with missing prepositional phrases. Consider the following contrast:

(50) I'll live in whatever town you live

(51) *I'll live in that town you live

Larson's explanation was that 'whatever town you live' is a QP with a deleted PP and undergoes QR:

(50′) (whatever$_j$ town you live [PP])$_i$ I'll live in e$_i$

and after reconstruction of the PP the result was the structure in (50″):

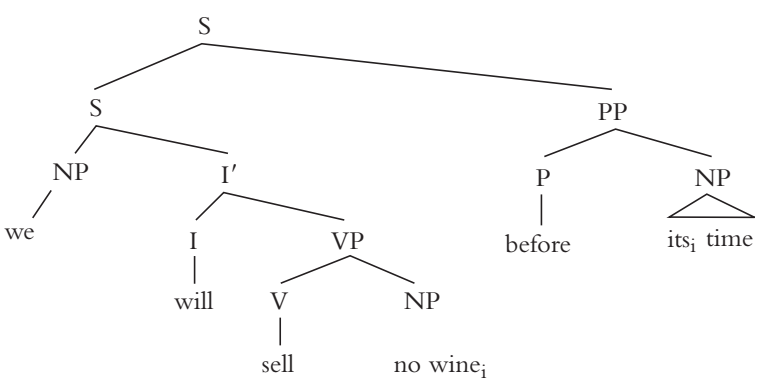
(50″) (whatever$_j$ town you live in e$_j$)$_i$ I'll live in e$_i$

Finally, there are the Orson Welles sentences, noted in Ludlow (1985), which were inspired by watching Orson Welles hawking Paul Masson wine on television.

(51) We will sell (no wine)$_i$ before its$_i$ time

Notice that the NP does not c-command the variable at S-structure

(52)



QR would account for the possibility of the very natural reading in which Orson Welles is saying no wine is such that we will sell it before it (that wine) is ready. Without QR one only predicts the reading in which one does not sell any wine before the appropriate time for some unspecified referent of 'it'.

Another interesting class of facts involves the inverse linking cases discussed in May (1977).

(53) Someone from every city despises it

May noted that this is ambiguous between a reading in which every city is such that someone from it despises it, and one in which a well travelled person despises something. Notice that binding is not possible in the latter case, presumably because the quantifier does not c-command the pronoun.

(54-a) $(\text{every city})_i[[\text{Someone from } e_i] \text{ despises } it_i]$

(54-b) $*[[\text{Someone from } (\text{every city})_i] \text{ despises } it_i]$

In Chapter 7 we will look very briefly at alternatives to QR, principally to consider the claim that if you can avoid another level of representation (LF) and rules like QR the result is a simpler or more elegant theory. As we will see, I have my doubts about whether this is in fact the case.

Semantics

Looking back to the generative semantics era it is easy to forget just how in the dark we were about semantics. While Grice's theory of pragmatics was out there, it had not been fully absorbed by most philosophers and linguists. But just as significantly, it took work such as Lewis (1972) and the work collected in Montague (1974) to show how you might construct a compositional semantic theory that began by assigning semantic values to the lexical items and then show that (directed by the syntactic tree structures) the semantics could compute the semantic values of the larger constituents from the meanings of the lexical items.

To be sure there were disagreements about the character of the semantic theory. Montague and Lewis (and Partee and her students in linguistics) had advocated a kind of model-theoretic semantics in which meanings were set-theoretic objects that helped to account for the entailment relations between sentences.

Others such as Higginbotham (1985), following a suggestion of Davidson (1967), proposed an "absolute" truth-conditional semantics that dispensed with the notion of truth in a model (and for the most part kept the resources first order and otherwise minimal). Barwise and Perry (1983) advocated a form of *situation semantics*—the semantic values would be constructed from components of the actual world.

There was also wide variation about whether more work was to be done by the semantics or the syntax. That is, some semanticists argued that one should avoid

"invisible" syntactic structure and that one could do this by introducing the appropriate model-theoretic resources into the semantics.

Despite all of this apparent variation, a widespread consensus emerged. All of these approaches are interpretivist. All of them took the semantics to be something that could "read" the meaning of the linguistic structure by computing it from the assignments of semantic values to the lexical items and computing the rest from the syntax. Thus, a clear consensus emerged around a basic thesis: that a compositional semantics assigns meanings to some level of linguistic representation (probably LF, perhaps SS). This is not to minimize the significance of the points of disagreement, and we will take up some of those in Chapter 6.

1.2 Government and Binding Theory

Work in the late 1970s was progressing in a number of disconnected albeit fruitful directions. For example, beginning with Jackendoff (1974) a great deal of work was done on X-bar theory. Philosophers and linguists like Fiengo, Higginbotham, and May worked on the development of LF and binding theory. Others continued to work on ways of accounting for the Ross Constraints. Work by Vergnaud introduced case theory into the mix.

In 1979, Chomsky gave a series of lectures in Pisa, Italy, in which he tied all of these threads together in an intricately interconnected package. The lectures were published as *Lectures on Government and Binding* in 1981 and the framework came to be known as Government-Binding Theory (or GB for short).

The theory consisted of several "modules"—called modules because their descriptions were related to narrow classes of specific linguistic phenomena. In a certain sense, however, 'module' was a funny name for them, since 'module' connotes a narrow bandwidth of communication between modules. As we will see, the interaction between modules was dense and complex, although the results were often breathtaking.

In Chomsky (1981) the components of the grammar were the following.

(i) **Bounding Theory**. This had to do with constraints on movement such as subjacency (S and NP together might be thought of as a bounding node).

(ii) **Government Theory**. This defines the relation between the head of a construction and the categories that depend on it. I will give an example in a bit, but among other things government theory interacted with trace theory in that a trace had to be "properly governed." Another example might be the notion of abstract case, which is assigned under government.

(iii) **Theta-Theory**. This dealt with the assignment and functioning of thematic roles, such as agent, patient, instrument, etc. This module interacted with government theory in that theta-assignment took place under government.

(iv) **Binding Theory**. This treated the properties that explain the relationship between grammatical elements such a pronouns, anaphors, names, and variables with their antecedents.

(v) **Case Theory**. This had to do with the assignment of abstract case—for example, nominative case might be assigned to an NP.

(vi) **Control Theory**. This determined the potential references of 'PRO'. So, for example, contrast 'John promised Bill PRO to leave' and 'John persuaded Bill PRO to leave.' In the first case 'PRO' is controlled by 'John' and in the second case it is controlled by 'Bill'.

(vii) **X-Bar Theory**. As discussed before, this was now doing the work previously done by the phrase structure component, in effect constraining the class of DS representations.

We saw that after the development of trace theory linguists searched for appropriate ways of stating structural constraints. GB provided a way of accounting for these effects, but also at the same time accounted for some puzzles in the binding theory. One interesting element of GB was the *Empty Category Principle (ECP)*, which accounted for subtle contrasts in the acceptability of a range of structures involving phonetically unrealized elements like trace and PRO. Formally, the ECP was stated as follows.

ECP: A trace must be properly governed

Proper Government: α properly governs β, if and only if α governs β, and
(i) α is N, V, or A or
(ii) α is co-indexed with β

Government was a phrase structure relation, closely related to c-command, but more local. To illustrate how the ECP worked, consider the contrast between (55) and (56).

(55) [Who_i do you think [e_i arrived]]

(56) *[Who_i do you think [that [e_i arrived]]]

In (56), the 'that' prevents the verb 'think' from properly governing the trace of movement. That is to say, the trace of movement is an empty category that in this case is not close enough to the upper V to be governed by it because of the intervening complementizer 'that'. The lower V does not c-command it (does not have scope over it) and hence also does not govern it. The wh-element 'who' is co-indexed with the trace, but is also too far away to govern it. From a semantic point of view the trace is sitting there perfectly interpretable, but a syntactic constraint has been violated. Traces (like other empty categories) for some reason like to be blanketed by nearby lexical predicates (V, N, A).

The ECP accounted for a very broad range of phenomena, including those involving LF quantifier movement. So, for example, consider the following contrast.

(57) Someone believes that everyone is a spy

(58) Someone believes everyone to be a spy

There seems to be a contrast here in that it is much easier to hear the 'everyone' as taking scope over 'someone' in (58) than it is in (57). If we think of the ECP as being an LF constraint, there is a natural explanation parallel to the explanations for the contrast between (55) and (56); in (57) the intervening 'that' prevents the trace of LF movement from being properly governed:[6]

(57′) *[everyone$_i$ [Someone believes that [e$_i$ is a spy]]]

Binding theory obviously played an important role in GB, and it interacted with other components of the grammar in very interesting ways. Chomsky formalized the binding theory according to three principles: A, B, and C.

A. A bound anaphor must be bound in its governing category
B. A pronoun must be free in its governing category
C. A lexical NP must be free

One of the interesting subsequent discoveries is that we can sharpen these principles a bit if we avail ourselves of the following taxonomy.

+anaphoric –pronominal	bound anaphors, NP-trace
–anaphoric +pronominal	pronouns
–anaphoric –pronominal	lexical NPs, logical variables
+anaphoric +pronominal	PRO

Using this taxonomy, we can revise the binding theory in terms of its basic features:

A. A [+anaphoric] NP must be bound in its governing category
B. A [+pronominal] NP must be free in its governing category
C. A [–anaphoric –pronominal] NP must be free

One of the interesting consequences of this was the question of what happens to PRO in the binding theory. Since it is both [+anaphoric] and [+pronominal] it looks like it must be both free and bound in its governing category. But notice that there is a way out. PRO *could* occur, but it would have to occur in places where it has no governing

[6] This obviously can't be the whole story, given that in 'John believes that someone is a spy' it appears this could be a de re belief about some individual. One could argue that this means that 'someone' has at least escaped the scope of 'believes' and this should leave the trace ungoverned. Alternatively one could argue that this reading is not generated scopally in this case.

category (otherwise it would violate either principle A or principle B). Are there places where PRO could have no governing category? Well yes, if it is ungoverned, and the only place it can be ungoverned is in the subject of an infinitival clause (tense would govern it, so it cannot appear in a tensed clause—hence we can't say things like [John wants PRO gone]). This became known as *the PRO theorem*, and it does a nice job of illustrating the rich deductive structure of GB. Of course given that the PRO theorem has been abandoned (Chomsky himself dropped it in Chomsky 1995a) there is an interesting question about how such a sweet result could simply disappear (and therefore whether it was a result at all). This goes to the question of explanatory loss that happens as a theory advances (sometimes called Kuhn loss). We won't address that topic head-on, but it will come up from time to time in this book.

For linguists that were used to a theory that contained a stock set of explanatory tools (for example phrase structure rules and transformations) the new framework was shocking. Even the generative semanticists, although theoretically unconstrained, had placed all the explanatory work on the transformational component (or on clearly specified projection rules). Now that component had dissolved to virtually nothing ("move anything anywhere") and the constraints on movement did not come from a single set of explicit rules, but seemed to be the result of a cacophony of various unrelated factors.

Consider, for example, the account of passive in Chomsky (1981).[7] In this new account, passive was not the result of a single transformation (or two transformations if we count affix hopping). Rather passive was the product of a number of linguistic properties interacting in subtle ways.

Take a classic example of a sentence and its passive form.

(59) Sam kissed Lilo

(60) Lilo was kissed (by Sam)

'kiss', obviously, is a transitive verb. In this way it contrasts with the following intransitive verbs (verbs that don't take subjects but no objects).

(61) *Lilo danced Sam

(62) *Lilo seemed Sam

(63) *Lilo arrived Mary

Accordingly, we can say that passivization converts a transitive verb into an intransitive verb. But it does more than that. It also converts the object of the transitive into the subject of the intransitive verb. Consider (59) and (60) again. 'Lilo' is the object in (59) but is converted into the subject in (60).

Another property of passives is that active and passive instances of the same verb differ in their form. To a first approximation, this difference is something like the

[7] My exposition of this case borrows from Roberts (1985).

addition of morpheme like '-ed' (as in 'kiss' → 'kissed'). In some cases, of course, there is a change in the stem verb ('sink' → 'sunk', 'sang' → 'sung').

As we saw earlier 'en' is the preferred affix for a broad class of cases (as in (64)), and linguists sometimes take /en/ to be the underlying form of the passive affix—it might be pronounced as '-ed' in some cases. Consider (64)

(64) Sam was given a DVD by Lilo

Example (64) shows us that ditranstive verbs—verbs like 'give' that can take an indirect object—can passivize. But now consider the following interesting cases.

(65) Sam was believed to be a musical genius

(66) Sam was considered smart

(67) Sam was spoken about

(65) and (66) tell us that so-called exceptional case-marking verbs can passivize. That is, verbs like 'believe' and 'consider' can case-mark an NP in a lower clause (as in 'I consider Lilo a singer'—tense isn't case-marking 'Lilo', the verb 'consider' is). (67) shows us that objects of PPs can passivize. We can generalize these observations as follows: the first NP case-marked by an active V cannot appear following V+en.

Pulling together all the observations made thus far, we have the following.

(i) Passives are case-intransitive
(ii) The subject of the passive is the object of the corresponding active
(iii) Passives are formed from actives by the affixation of /en/.

All of these properties can be accounted for by the interaction of case theory and theta-theory. So, for example, Chomsky (1981, 124–127) correlates the properties of passive with the following claims about case and theta-assignment in passive:

(68) NP/S is not a theta position in passives

(69) NP/VP is not assigned case in passives

DS is supposed to be a pure representation of argument structure, so the DS for our passive must be as in (70) (Lilo is the patient of kiss, and so must be governed by 'kiss' at DS).

(70) [e (was) kiss+en Lilo]

Since passive doesn't assign case inside the VP, 'Lilo' must move into subject position, otherwise the case filter will be violated.

(70′) [Lilo_i was kissed e_i]

Of course this raises the question of where the properties in (68) and (69) come from. Rouveret and Vergnaud (1980) had proposed that /en/ was of category [+V]—in effect

neutral between Verb and Adjective (recall our discussion of X-bar syntax earlier). Following standard assumptions in morphology (Lieber 1980, Williams 1981), the affix in effect becomes the head of the word and determines its categorial status. So, for example, if you affix /en/ to something that is [+V –N], the result is something that is [+V]—the categorial status of the affix trumps that of the stem.

Here we get a nice payoff from the introduction of the "subatomic" features like [+V] that Chomsky introduced in "Remarks on Nominalization". Deploying this technology a decade later, Chomsky proposed that case assigners are [–N]. Because main verbs are [+V –N] they can assign case. Passive morphology converts the verb into a non-case assigner by converting it into something that is [+V].

In effect, everything in the analysis of passive falls out from the assumption that the passive morpheme is [+V]. The account is complicated in one sense, but breathtakingly simple in another sense. You don't need a passive *transformation*, you just need the passive morpheme to have the feature [+V]. The rest is driven by independently motivated properties of the grammar—the case filter, theta-theory, standard assumptions about feature projection in morphology, proper governance, etc.

Just as the earlier generation of linguists broke from Chomsky during the "generative semantics wars," a similar break took place during the advent of GB theory. It isn't written about in quite such epic terms because the debates were relatively more civil, but they were contentious for all that. One concern expressed by a number of generative linguists was that in the move to GB too much formal rigor had been sacrificed. One can see where this impression came from. Things were very tightly defined in linguistics up until GB, but Chomsky certainly hadn't laid out a stock of rules like he had in *Aspects*.

On the other hand, just because it hadn't been done doesn't mean it *should* be done (more on this in Chapter 7). It also doesn't mean that it *couldn't* be done. Johnson (1991) and Stabler (1992) both showed that it was possible to formalize GB syntax and use the formalization to construct natural language parsers. In fact, there was nothing particularly difficult about the formalization part. We will return to the formalization question (especially with respect to GB) in section 7.2.

1.3 The Principles and Parameters Framework

As we noted earlier, from the very beginning generative linguists supposed that they were investigating the mechanisms that explained macro-level phenomena. Even through Government and Binding Theory, however, there was little insight into the character of the mechanism itself. Generative linguistics had been able to identify a compelling stock of very interesting (indeed quirky) rules, but the field had no real handle on the underlying mechanism giving rise to those rules, and hence little insight into the structure of the object of investigation itself. This changed with the emergence of the Principles and Parameters framework. Indeed, Chomsky (2000a, 8) claimed that the P&P framework "gives at least an outline of a genuine theory of language, really for

the first time." Commentators such as Smith (2000) went so far as to say that it is "the first really novel approach to language of the last two and a half thousand years."

The core idea of the P&P framework is this: Each human is born with a language acquisition device that is largely prewired, but which leaves open a handful of parameters—we can think of these parameters as toggle switches. When a child is exposed to linguistic data the parameters/switches are set, and the language the child learns is determined by the combination of parameters that have been set. Baker (2001) argues that the parametric theory provides us with a kind of periodic table that maps out the space of human languages and the relations holding between them. For example, Baker suggests that many of the apparently radical differences between English and Mohawk (an indigenous American language spoken in an area that now comprises parts of Quebec, Ontario, and New York) are determined by a single parameter setting (Baker's "polysynthetic parameter"). Baker also claims that thanks to the P&P framework the current historical position in linguistics is comparable to that of chemistry during the development of the first periodic table by Mendeleyev.

To illustrate the difference in perspective using an analogy from biology, traditional grammarians have been impressed by superficial similarities between languages, much like we might want to put whales and fish in the same taxonomy because they have the same shape and inhabit the same locations. But of course this would be an error; there are deeper differences that lead us to classify whales as mammals. On Baker's version of the principles and parameters model, the parametric settings of English and Indonesian are nearly identical, with Edo and Khmer being closer relatives to English than French is. A systematic attempt to integrate the Principles and Parameters framework into the theory of language acquisition has subsequently emerged and an industry studying the role of parameters in language acquisition has likewise emerged in psycholinguistics (see Lightfoot 1993 and Hyams 1986 for surveys).

A good example of a parameter, originally due to Rizzi (1982), is one that involves subjacency. Recall that Chomsky's initial account of subjacency effects was that an NP could not simultaneously extract from both an NP and an S node. Rizzi noticed that the variation between extraction facts in Italian and English could be accounted for if we hypothesized that the relevant bounding node in Italian was NP and so-called S′ (recall that S′ is the node that has for its daughters the complementizer 'that' and the S that it introduces for its daughters). Thus the core principle of subjacency would be stable across languages, but there would be a "subjacency parameter" as to what the bounding nodes would be in different languages.

The interesting thing about the principles and parameters model is that it seeks to find deeper level parameters that can account for a broad range of macro-level phenomena. Consider the following illustration from Baker (2001), citing early work by Richie Kayne and Luigi Rizzi.

The simplest Italian sentences look very much like the simplest English and French sentences. So, for example, we have the following paradigm.

(71) Jean arrivera (French)
Jean will-arrive

(72) Gianni verrà (Italian)
Gianni will-come

Italian falls outside of the English/French paradigm in that the subject can follow the verb. Thus we have the following contrast:

(73) Verrà Gianni
will-come Gianni

(74) *Arrivera Jean
will-arrive Jean

A second difference between English/French and Italian is that the latter is what linguists call a "pro-drop" language; in cases where we might use a pronoun to refer to someone who is already salient in the discourse, Italian allows (favors) that the pronominal element drops out altogether. Thus we have the following contrast.

(75) Verrà (Italian)

(76) *Arrivera (French)

So far the differences under discussion would be obvious to anyone what had both French 101 and Italian 101, but as Rizzi and Baker stress, the languages differ in more subtle ways too.

For example, a third difference between English/French and Italian is exemplified by the following paradigm. Consider how we might form a question from a sentence like (77)

(77) Chris will see someone in the park

In English, we replace 'someone' with the WH-word 'who' (or 'whom' if you prefer) and front it as in (78).

(78) Who will Chris see ____in the park?

French and Italian work like this in simple cases too. But now consider a more complex sentence like (79) and the French example (80).

(79) Who did you say that Chris saw ____in the park?

(80) Qui veux-tu que Marie épouse ____?
Who want-you that Marie marries ____?

In both of these examples the position being questioned is the object position—the person seen or the person married. But now notice what happens when we question the subject position.

(81) *Who did you say that ____saw Chris in the park?

(82) *Qui veux-tu que ____épouse Jean?

This is in contrast with the Italian case, where we can question the subject position with no problem.

(83) Chi credi che ____verrà?
Who you-think that ____will come?

The French and Italian versions can be repaired by avoiding the complementizers 'that' and 'que'. In English we simply drop the complementizer and in French we can convert it to 'qui' as in (81′) and (82′).

(81′) Who did you say ____saw Chris in the park?

(82′) Qui veux-tu qui ____épouse Jean?

A fourth difference between languages like French and English on the one hand and languages like Italian and Spanish on the other hand is that the former require that every tensed clause must have a subject of some kind. So, for example, in English and French we introduce a "pleonastic it" in cases like the following:

(84) It is raining (English)
Il pleut (French)

(85) Piove (Italian)
Lleuve (Spanish)

And indeed, it turns out that every language we know of goes one way or the other on this point. Some languages (like French, English, Edo) require that tensed clauses have subjects. Other languages (like Italian, Spanish, Romanian, Japanese, Navajo) don't have this requirement.

As Baker notes, the contribution of Chomsky, Kayne, and Rizzi was to show that all of these macro-level effects fall out from a single parameter—the Null Subject Parameter. The first and fourth cases we discussed fall out straightforwardly from this parameter: Because Italian and Spanish are null-subject languages they needn't bother with introducing subjects (hence 'verrà' is fine by itself) and pleonastic subjects need not be introduced in weather sentences (hence stand-alone 'piove').

What about the second paradigm? We said that English and French don't allow the subject to follow the verb, but this isn't exactly correct. The Null Subject Parameter doesn't prohibit the subject from moving to the end of the sentence; it merely requires that we introduce a replacement subject of some form, and of course we can do this in both French and English. Hence the following are fine:

(86) There appeared a boat on the horizon

(87) Il est arrivé trois hommes.
It is (has) arrived three men.

The explanation for the third paradigm is simply that if a language does not allow null subjects then any tensed clause must have a subject. Hence examples like (81) and (82) are unacceptable in English and French. Notice that this only applies to subjects. French and English don't require that sentences have objects. Hence we feel no compulsion to say 'It rained it'.

Chomsky (1981) suggested that these and two additional paradigms all fell out from a single parameter and suggested that this greatly simplified the language acquisition process for children. Simply by being exposed to a sentence like 'It is raining' or alternatively 'piove' is evidence for setting the null subject parameter. The resulting cascading effects were what Chomsky called a "parametric cluster".

As Chomsky observed, viewed from the macro level there are a number of possible variations. There is no reason why a language couldn't be pro-drop but prohibit the questioning of object positions. Strictly speaking, just based on the six facts related to this parameter, there could be $2\times2\times2\times2\times2\times2 = 64$ different Romance languages. But in reality there are only two—the French-like languages and the Italian-like languages.

Interestingly, the parametric model seems to undermine the idea that the function of the language faculty is to be a mechanism for maximizing the communicative function of language. Recall that just a minor flip of a switch (caused, for example, by language contact) can have a cascading effect with the consequence that the resulting language is very different. Baker (2003) makes the case that, far from being a high fidelity copying system, the language faculty is functionally much closer to an encrypting system. He draws on analogies to real world encryption to make his case for the Babelizing nature of the language faculty:

> Claims that a biological system has a particular function are often reinforced by comparing it to products of human engineering that have that function. When the two have detailed structural similarities, the claim that they have similar functions gains support, a classic example being the comparison between the vertebrate eye and a camera. The way human languages differ can be compared to cryptographic techniques of the 16th century. Sixteenth century cryptographers used a variety of techniques: they both replaced and rearranged symbols in systematic ways, and they performed these transformations both at the level of letters and at the level of words and phrases. This layered complexity evolved over time with the explicit purpose of defeating particular code-breaking strategies (such as frequency analysis). Natural languages also differ from one another in ways that show layered complexity, using substitutions and arrangement at multiple levels. Many of the specific tricks of the early cryptographers have striking analogies in natural language. This gives credence to the notion that natural languages have the same concealing function as man-made ciphers.

Baker's point here is that it does not appear that the function of language is to aid or enable communication. Indeed, there are other less extreme proposals for the function

of the language faculty. Chomsky has argued that the purpose—or at least the primary use—of the language faculty is for inner thought. Alternatively the selectional advantage might be something as simple as having a system that can generate certain metrical patterns—in effect for creating poetry.

An alternative, of course, is simply to say that UG is a kind of "spandrel" in the sense of Gould and Lewontin (1979)—that is not really *for* anything. It might be the by-product of the evolution of other cognitive faculties that did have some selectional value.

Discussions of this nature tend to get highly charged but this much is clear: Most of the interesting properties of the language faculty that we discussed in sections 1.1–1.3 don't appear to have anything to do with communication. The properties are weird, and quirky, and if anything they seem to inhibit communication in some respects.

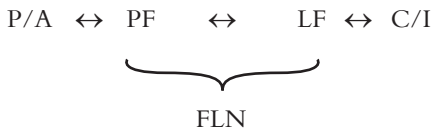
This is what generative linguists do. They study these quirky features of natural language and attempt to deduce the mechanisms by virtue of which the language faculty gives rise to these phenomena. It does not appear that fidelity of copying or aiding communication could play any sort of role in the explanation of these properties, hence syntacticians look elsewhere for the answers.

1.4 The Minimalist Program

Still working within the general principles and parameters framework, Chomsky (1995a, 2000b) articulated a research program that came to be known as *the Minimalist Program*. The headline idea behind the Minimalist Program has been the working hypothesis that the language faculty is not the product of messy evolutionary tinkering—e.g. there is no redundancy, and the only resources at work are those that are driven by "conceptual necessity." Chomsky (2001b) has been drawn to D'Arcy Thompson's (1966) proposal that the core of evolutionary theory consists of physical/mathematical/chemical principles that sharply constrain the possible range of organisms. As Chomsky (2000a, 163) puts it, "physical law provides narrow channels within which complex organisms may vary" and natural selection is only one factor that determines how creatures may vary within these constraints. Other factors (as Darwin himself noted) would include non-adaptive modifications and unselected functions that are determined from structure. In this case, the idea would be that those principles not only constrain low level biological processes (like sphere packing in cell division) but also that such factors might be involved across the board—even including the human brain and its language faculty.

Here is a way to illustrate the project. We can think of linguistics as a theory of the system that human agents have by virtue of which the appropriate representations interfacing with the human perceptual/articulatory (P/A) system are paired with the appropriate representations interfacing with the human conceptual/intentional (C/I) system. The system linking these interfaces is the *minimal* system that satisfies

"legibility" constraints or conditions imposed by both the P/A system and the C/I system. Following Hauser, Chomsky, and Fitch (2002) we can call this minimal system the *faculty of language narrowly construed* (FLN). Let's call the level of representation that interfaces with the P/A system *PF* (for phonetic form) and let's call the level of representation that interfaces with the C/I system *LF* (for logical form). We thus have the following picture:



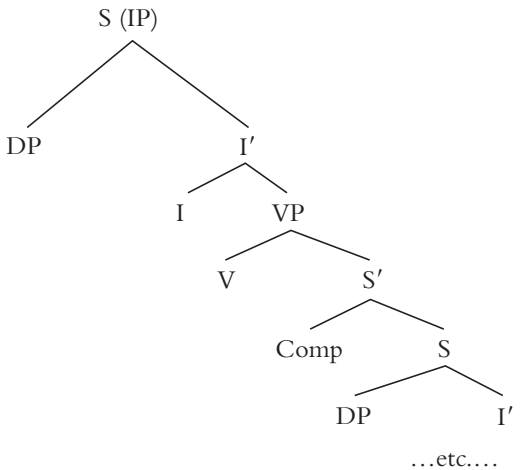
So the FLN is an optimally efficient wiring solution to the problem of linking these interfaces. One plausible story about the FLN is that it uses very simple operations like *merging* two elements (for example the words 'red' and 'ball') and moving elements as necessary (and only as necessary) to meet the interface legibility requirements. The hallmark of the resulting wiring solution is recursion, meaning that the process of merging elements in this way yields recursive structures.

What do I mean by recursion here? Consider the following toy phrase structure grammar.

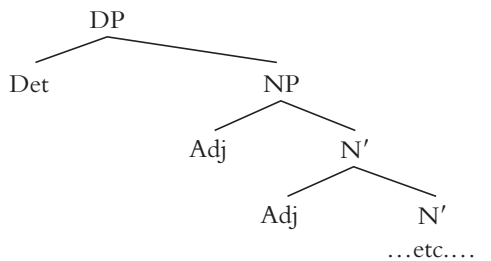
(1) XP → XP ConjP (where X is I, D, V, or P)
(2) ConjP → Conj XP (where Conj is 'and', 'or')
(3) S (IP) → DP I′
(4) I′ → I VP
(5) DP → det NP
(6) NP → N′ (PP)
(7) N′ → (adj) N′
(8) N′ → N
(9) NP → NP CP
(10) CP → wh C′
(11) C′ → C S
(12) VP → V XP

As I am using the term 'recursion', rules 1, 7, 9 would be recursive rules because the rules "call themselves", and *combinations* of rules like 3–5–9–10–11 would be recursive, because the topmost rule in the derivation is called again lower in the derivation. I am calling the output of such rules (the trees generated by such rules) recursive *structures* because the structures have "descendent" nodes that are the same type as their ancestors. For example, I take the following to be recursive structures because a category S occurs within S in the first structure and N′ occurs within N′ in the second.

(88)



(89)



As we saw in sections 1.1–1.3, linguistic structures are more complex than this and in current grammatical theory they would not be the product of rules like those I specified in the toy grammar above.

Of course, a simple merge operation and the resulting recursive structures do not, by themselves, put strong constraints on linguistic theory—nature is full of various kinds of simple processes generating recursive patterns after all (for example spiral patterns in shells and galaxies)—and some of the most interesting properties of natural language (subjacency, for example, or the basic principles of binding theory) don't seem to have anything interesting to do with recursion by itself. So the interesting properties of language appear to fall out from the fact that the system yielding these recursive structures must also meet some rather strict interface conditions. If there are no constraints imposed by the interfaces, then nothing forces the interesting properties of language into relief—we just get vanilla recursive structures, and not the quirky structures we find in natural language.

This raises interesting questions about the nature of the interface. Here, the linguistics literature is thin on detail, but I think a plausible story can be told about how the need to be legible to the semantics puts contraints on the possible kinds of recursive structures.

To be as concrete as possible in our discussion of semantics let's assume a standard approach to interpretivist natural language semantics (in the sense of section 1.2), adopting the notation and basic theoretical background articulated in Heim and Kratzer (1998).

For our purposes we assume the basic elements of a semantic theory are elements of type e and t, where we can think of elements of type e as being entities or referents and elements of type t as being truth values. Predicates of the language can be thought of as functions that map from elements of type e to elements of type t. So, for example, the predicate 'food' can be thought of as a function that maps from individual entities onto true or false. More complex elements like adjectives can be thought of as mapping from function to function etc. To a first approximation, (and borrowing some online course notes from Barbara Partee) the picture is something like this.

Syntactic category	**Semantic type (extensionalized)**	**Expressions**
ProperN	<e>	names (*John*)
S	<t>	sentences
CN(P)	<e, t>	common noun phrases (*cat*)
NP	<e>	"referential" NPs (*John, the king, he$_i$*)
	<<e, t>, t>	noun phrases as generalized quantifiers(*every man, the king, a man, John*)
	<e, t>	NPs as predicates (*a man, the king*)
ADJ(P)	<e, t>	predicative adjectives (*carnivorous, happy*)
	<<e, t>, <e, t>	adjectives as predicate modifiers (*skillful*)
REL	<e, t>	relative clauses (*who(m) Mary loves*)
VP, IV	<e, t>	verb phrases, intransitive verbs (*loves Mary, walks*)
TV	<e, <e, t>>	transitive verb (*loves*)
is	<<e, t>, <e, t>>	*is*
DET	<<e, t>, <<e, t>, t>	*a, some, the, every, no*

Semantic rules for the lexical items specify the particular functions. To illustrate, let's use the notation [[α]] to indicate the semantic value of α, and consider a couple of examples of lexical rules by which the semantics specifies the meaning of an expression.

(90) [['Schleiermacher']] = Schleiermacher

(91) [['ponders']] = $\lambda x \in D_e$. true, iff x ponders

We can read (90) as saying that the semantic value of the name 'Schleiermacher' just is Schleiermacher, and (91) as saying that the semantic value of 'ponders' is a function from any object x in the domain of things of type e (individuals) onto truth or falsehood. It maps onto truth if and only if x ponders. Of interest to us a bit later will be the rules for the connectives, which we can give as follows:

(92) [['and']] = $\lambda p \in D_t$. [$\lambda q \in D_t$. true, iff p = q = true]

(93) [['or']] = $\lambda p \in D_t$. [$\lambda q \in D_t$. true, iff p = true or q = true]

(94) [['not']] = $\lambda p \in D_t$. true, iff it is not the case that p = true

To illustrate, (92) says that the semantic value of 'and' is a function that inputs a truth value, and yields a new function that maps from truth values onto truth values. In particular, it inputs a truth value p, yielding a new function which inputs a truth value q and outputs true just in case both p and q are true.

Now, for my money, the most interesting feature of formal semantics is discovering how much one can accomplish with such limited resources. You might think that given the infinite number of syntactic constructions you might encounter there must be an infinite number of semantic rules to interpret them, or at least an awful lot of rules. But once the lexical rules are in place we get by with interpreting all those structures just using a handful of very simple semantic rules. Following Heim and Kratzer (1998), let's characterize them as follows.

Terminal Nodes (TN): If α is a terminal node, then [[α]] is specified in the lexicon.

Non-Branching Nodes (NN): If α is a non-branching node, and β is its daughter node, then [[α]] = [[β]].

Functional Application (FA): If α is a branching node, $\{\beta, \gamma\}$ is the set of α's daughters, and [[β]] is a *function* whose domain contains [[γ]], then [[α]] = [[β]] ([[γ]]).

Predicate Modification (PM): If α is a branching node, $\{\beta, \gamma\}$ is the set of α's daughters, and [[β]] and [[γ]] are both in $D_{<e, t>}$, then [[β]] = $\lambda x \in D_e$. [[β]] (x) = 1 and [[γ]] (x) = 1.

Predicate Abstraction (PA): If α is a branching node whose daughters are a relative pronoun and β, then [[α]] = $\lambda x \in D_e$. [[β]]^x

Now clearly this is just a first pass and I'm leaving out detail, but it won't matter for the discussion that follows. The point is that this semantics is rich enough to interpret all the structures that can be generated by the toy grammar I gave above. And then some. And this, to me, is impressive.

But now if we think of the semantics (or the semantics module) as being the locus of contact with the C/I interface, then these semantic rules determine the conditions for structures being legible. That is to say, if a structure is to be legible, then it must be a structure that can serve as an input for these rules. It must be visible to the rules.

For example, only binary branching and non-branching structures are visible to these rules—ternary branching structures are thus not legible. Furthermore, while these rules can "read" binary structures in which both daughters are nouns (as in 'foil ball') they cannot read structures in which there are pairs of referential expressions (expressions of type e). So these rules cannot read 'that that' or 'she she'.

But notice also that if some of these rules were missing then certain recursive structures would not be legible. For example:

If **Predicate Modification** was missing then recursive structures that involved noun–noun recursion would be out, as would structures involving NPs with recursively iterated PP modifiers, like 'dog in a city in a county in Texas . . .'

If **Predicate Abstraction** was missing, then recursive structures involving relative clauses would not be legible (at least on the Heim and Kratzer analysis of relative clause structures), since predicate abstraction is a device that takes a CP and converts it into an expression of type <e, t>. Without PA the structure corresponding to 'this is the cat that ate the rat that ate the cheese that . . .' would not be legible.

Of course, some recursive structures are made possible not by these basic semantic rules but by the lexical rules. Canonical examples of recursion involve the logical connectives.

So for example consider a lexical rule like the following, which is the 'and' used to conjoin predicates as in (sang and danced):

(95) $[[\textbf{and}]] = \lambda f \in D_{<e,t>}. [\lambda g \in D_{<e,t>}. [\lambda x \in D_{e.}$ true, iff $f(x) = g(x) = 1]]$

Without a lexical rule like this, then barring the introduction of conjunction reduction there would be no way to interpret the structures with conjoined predicates—for example a construction like 'She sings and dances and acts and . . .'.

We can put the point this way. Certain kinds of lexical rules are preconditions for a large class of recursive structures—for example, those involving logical connectives. To a first approximation, these are lexical rules that involve mapping from a function of a particular type onto a function of the same type. That is, the lexical rules must specify functions that are (i) higher order and (ii) type reflexive.

First order functions—for example those that are <e, t>—are not sufficient by themselves for recursive structures like those generated by our toy grammar to be legible. Higher order functions that are not reflexive are also not sufficient (by themselves) for recursive functions to be legible. For example determiners are higher order, but not reflexive in the sense I have indicated.

You can get recursion without type reflexivity, but only if you have a semantic derivation that forms a type reflexive theorem. So for example, sentential complements appear to be of this form: [John believes that Mary believes that . . .]. Successive application of rules can generate the type reflexive theorem. On the semantic end, what is permitting these recursive syntactic structures (like 'John believes that Mary believes that . . .') to be legible is what we could call a **derived type reflexive theorem**. In this case, what is allowing the recursive structure is that once axioms are combined in a derivation, the resulting step or theorem is type reflexive.

Summing up, we can chart out the preconditions for the legibility of recursive structures as follows.

PM: required for noun–noun compounds, and NP–PP recursion
PA: required for relative clause recursion

Type reflexive lexical rules: required for [adj [adj [. . . [N]] recursion, logical connective recursion, etc.

Derived type reflexive theorems: required for sentential complement recursion

I don't mean to suggest that this approach to the interface is canonical. Rather, I'm trying to give a plausible story about how the interface might place constraints on the grammar and give rise to some of the linguistic phenomena (e.g. the kinds of recursive structures) that we observe. Are these kinds of interface constraints plus minimal operations like merge sufficient to account for the core phenomena that we canvassed in sections 1.1–1.3? Here work is preliminary—certainly too preliminary to evaluate—but suggestive. Kayne (2002) and Zwart (2002) have offered stories about deriving basic principles of binding theory within the minimalist framework, Collins (2002) has offered an account of phrase structure that supersedes X-bar theory, Hornstein and Uriagereka (2002) have explored minimalist explanations for facts that we would ordinarily take to be LF phenomena.

Not surprisingly, a number of philosophical themes have already emerged from the Minimalist Program. One topic has to do with the interface and the question of whether a naturalistic account of semantics can be given, as seems to be required by the theory. (I take this up in Chapter 6.) Another theme—one that was also central in the days of Generative Semantics—is the suggestion that it is better to explain fewer facts if one is interested in gaining deeper explanations. That is, the price of deeper understanding may be that some things we thought were the target of linguistics turn out not to be. (I take up this topic in Chapter 3.) Yet another theme that has emerged is the idea that the Minimalist Program has the advantage of being conceptually *simpler*. (I take up this topic in Chapter 7.)

Finally, one of the themes that has emerged from the Minimalist Program, is the suggestion due to Epstein and Seely (2006), that within the research program linguistic explanation is *derivational* rather than representational. By this they mean that it is the basic operations themselves that account for linguistic phenomena, and not constraints that we might state over linguistic representations. For example, in the work on binding theory by Kayne and Zwart cited above, the idea is that binding relations are not stated as relations between objects on a phrase structure tree but rather are determined by merging operations (the antecedent and trace are related by virtue of being merged before the antecedent moves).

At least for Epstein and Seely the pendulum has swung away from Chomsky's (1976) suggestion that most of the work could be done by constraints on representations (for example on relations between traces and their antecedents). We will take up this topic in the next chapter, but I don't mind providing a spoiler here. The dispute between derivational and constraint based representational theories in linguistics tends to have the hallmarks of an earlier dispute in computer science between those who advocated declarative languages (like Prolog) and those who advocated procedural languages (like C and its dialects). In computer science the consensus emerged that the dispute came to

nothing, for the simple reason that at the lower levels the machines were doing pretty much the same thing whether the higher level language was procedural or declarative. The different kinds of programming languages provide us with different ways of reasoning about and programming the action of the machines—the best one is the one that is the easiest for us to use given the problem at hand.

I think a similar situation holds when we look at the difference between constraint based (representational) and derivational approaches to the theory of grammar. The different approaches do not reflect a deep difference in the structure of the mind/brain, but rather reflect different ways we can interpret what is happening. The choice may well depend upon the state of the theory at a particular moment (Chomsky's work, for example, seems to shift from being derivational to being constraint based and back depending upon the needs of the theory he is constructing at the moment). It is important that we not mistake aspects of the theory that are designed to help us theorize better for aspects that reflect basic properties of the mind/brain. It is especially important that we not deploy these aspects of the theory as first principles from which to theorize.

Of course, this issue is philosophically subtle, and before we can tackle it we need to sort out some questions about the ontology of linguistics. We turn to that work next.