

Order and Structure

by

Colin Phillips

B.A., Modern Languages
University of Oxford, 1990

Submitted to the Department of Linguistics and Philosophy
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
at the
Massachusetts Institute of Technology

August 1996

© 1996 Colin Phillips.
All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly
paper and electronic copies of this thesis document in whole or in part.

Signature of Author: _____
Department of Linguistics and Philosophy
August 31st 1996

Certified by: _____
Alec Marantz
Professor of Linguistics
Thesis Supervisor

Accepted by: _____
Wayne O'Neil
Professor of Linguistics
Head, Department of Linguistics and Philosophy

Introduction

Introduction

Order and Structure

by

Colin Phillips

Submitted to the Department of Linguistics and Philosophy on August 31st,
1996 in partial fulfillment of the requirements for the degree of Doctor of
Philosophy in Linguistics.

ABSTRACT

The aim of this thesis is to argue for the following two main points. First, that grammars of natural language construct sentences in a strictly left-to-right fashion, i.e. starting at the beginning of the sentence and ending at the end. Second, that there is no distinction between the grammar and the parser.

In the area of phrase structure, I show that the left-to-right derivations forced by the principle *Merge Right* can account for the apparent contradictions that different tests of constituency show, and that they also provide an explanation for *why* the different tests yield the results that they do. Phenomena discussed include coordination, movement, ellipsis, binding, right node raising and scope.

I present a preliminary account of the interface of phonology and morphology with syntax based on left-to-right derivations. I show that this approach to morphosyntax allows for a uniform account of locality in head movement and clitic placement, explains certain directional asymmetries in phonology-syntax mismatches and head movement, and allows for a tighter connection between syntactic and phonological phrases than commonly assumed.

In parsing I argue that a wide range of structural biases in ambiguity resolution can be accounted for by the single principle *Branch Right*, which favors building right-branching structures wherever possible. Evidence from novel and existing experimental work is presented which shows that Branch Right has broader empirical coverage than other proposed structural parsing principles. Moreover, Branch Right is not a parsing-specific principle: it is independently motivated as an economy principle of syntax in the chapters on syntax.

The combination of these results from syntax and parsing makes it possible to claim that the parser and the grammar are identical. The possibility that the parser and the grammar are identical or extremely similar was explored in the early 1960s, but is widely considered to have been discredited by the end of that decade. I show that arguments against this model which were once valid no longer apply given left-to-right syntax and the view of the parser proposed here.

Thesis Supervisor: Alec Marantz
Title: Professor of Linguistics

Introduction

Introduction

Acknowledgments

I left this part to the very end, thinking it might be easy, but I hadn't reckoned with the enormous number of people who deserve credit for what follows, many of whom probably don't even get a mention here.

Alec Marantz has been far more than my advisor and committee chair. It goes without saying that this thesis has greatly improved as a result of many hours of discussion with Alec and by his constant written feedback. We have argued and argued about this, and it was always valuable. But Alec has been more than my thesis advisor. We have collaborated for three years as novice neurolinguists, with Alec as our intrepid leader. Alec has been my running training partner for a number of years, and he played a large part in my 'other' big project of this year, which was to qualify for the Boston Marathon. He has always been willing to give help and advice, ranging from more regular academic matters to feeding my cat and helping me move house, with a whole lot of other things in between. He is a good friend.

I have learned a lot about doing linguistics from David Pesetsky. He seemed to know where this thesis was headed well before I did. His infectious enthusiasm made an great difference to the enjoyment I got from working on this, and he never failed to find the fatal flaws in my logic, which led to a good many matters being clarified, in my head as much as on the page. Beyond this thesis, his co-directorship (with Ken Wexler) of the joint program in linguistics and cognitive science has had an enormous impact on what I have got out of my time at MIT.

Noam Chomsky's immense influence on this work should be clear, and my discussions with him always led me to rethink whatever I had been thinking about. Perhaps this isn't so surprising, since what I'm trying to do here is something that Noam had already done long before I was born. I only wish that I had begun to meet regularly with Noam at an earlier point.

I can say for sure that I would not have written this thesis had it not been for Ted Gibson. Three years ago, before Ted came to MIT, parsing was the last thing that I was expecting to write my dissertation about. This thesis had its beginnings in a (very late) term paper I wrote for a class of Ted's, and somehow it developed from there, with Ted as my main advisor in this area for the first year. Thankfully, none of the original paper has survived to this version. Working with Ted has been extremely rewarding, and a good deal of fun, too. Particularly the material in Chapter 3 has benefitted from countless hours of discussion with Ted. Ted is the only professor with whom I have left an appointment running, so that we would not miss the final subway of the night, at 12:15am!

Beyond my committee a number of people have provided large amounts of feedback, information and other assistance with this thesis. I am amazed at how

Introduction

helpful people are around here—it is a wonderful community of linguists. Heidi Harley, Jonathan Bobaljik and Andrea Zukowski all read the entire document and provided detailed comments. Danny Fox, Uli Sauerland and Carson Schütze all read sizeable chunks and returned sizeable comments with amazing speed.

Many other people have directly contributed to the thesis, in the form of discussions, correspondence, judgements, criticisms, technical assistance, and miscellaneous wisdom: Carl Alphonse, Albert Alvarez, Kevin Broihier, Andrew Carnie, Paul Crook, Janet Fodor, Danny Fox, Bob Frank, Lyn Frazier, Paul Gorrell, Paul Hagstrom, Ken Hale, Koji Hoshi, Dick Hudson, Edith Kaan, Masa Koizumi, Howard Lasnik, Martha McGinnis, Don Mitchell, Shigeru Miyagawa, Renate Musan, Orin Percus, Norvin Richards, Uli Sauerland, Julie Sedivy, Mike Tanenhaus and Myriam Uribe-Echevarria, Ken Wexler and Kazuko Yatsushiro, Elron Yellin.

The experimental work reported in Section 2 of Chapter 3 was conducted in collaboration with Ted Gibson. Thanks to Neal Pearlmuter, Ariel Salomon, and San Tunstall for their invaluable assistance with these experiments.

Work on this thesis was supported in part by the NSF funded program ‘Language: Acquisition and Computation’ (#DIR9113607) awarded to MIT.

...and that’s just the dissertation. I also owe a great debt to the many other people who have helped me through my first six years of being a linguist and my first six years of living in the United States.

My introduction to linguistics and to the USA came when I was a visiting student at the University of Rochester in 1990-1991. I had only intended to be in this country and in linguistics for a year, but the warm welcome and the exciting intellectual environment that I found there quickly made me change my plans. Itziar Laka inspired me, because she cared about finding the right answer more than anybody I had encountered in my earlier life as a medievalist. Sandro Zucchi and Roberto Zamparelli provided many many hours of discussion and instruction. Many many other people in Rochester provided a warm welcome. The absence of any distinction between theoretical linguistics and psycholinguistics which I encountered in Rochester has had clear influences on the document that follows.

I benefitted immensely from living with Hamida Demirdache, Tom Green and Kumiko Murasugi in my first year at MIT. From remedial phonology help to forming a band, this was an extremely happy time.

I was lucky enough to spend the first 4 years of my time at MIT with a wonderful group of classmates: Masha Babyonyshev, Andrew Carnie, Heidi Harley, Shawn Kern, Masa Koizumi, Renate Musan, Orin Percus, Hubert Truckenbrodt, and adoptively Jonathan Bobaljik. Since most of them graduated and left a year ago, I have missed their company and lively discussions very much this year.

There are many many other people who have helped to make my stay at MIT incredibly rewarding. Apart from the people already mentioned and the people who I will kick myself for forgetting right after I hand this in (it is, after all *very* late as I write, or is it early?), there are Morris Halle, Tony Harris, Irene Heim, Maya Honda, Dianne Jonas, Michael Kenstowicz, LING88–95 (that’s just about everybody), Gary Marcus, all of the MITWPLers past and present, Wayne O’Neil, David Poeppel, lots of great people in the BU Applied Linguistics

Introduction

Program (you know who you are), and Linda Thomas, my companion in 'Broca's Area' for much of the writing of this thesis, Leo my furry companion at home for the rest of the writing.

Derek Gross was one of the first linguists I met in Rochester, and a friend in Boston for number of years. He showed me this year that writing a thesis is really not such a big deal. His ability to deal with cancer in a cheerful and always positive manner put everything in clear perspective for me. I wish I could thank him for this. Many people will miss him very much.

I owe an immense debt of love and thanks to Andrea Zukowski, my best friend and wife. Of all people she has had to put up with the most through the writing of this thesis. She has listened to the most unformed ideas at the most inappropriate times. She has read drafts of the thesis in great detail, raised many important questions and clarified many others. She has been a constant source of encouragement and support, and as writing took up more and more time she insisted on staying up into the small hours to keep me company. She has taught me most of what I know about bread-baking and statistics, and an enormous amount about life.

Finally, I thank my parents Kathleen and Derrick Phillips for their constant support and many years of sacrifices on my behalf. I owe them everything.

Introduction

Introduction

Table of Contents

1. Introduction.....	13
1.1 General Architecture.....	13
1.2 Outline.....	18
1.3 Some reminders.....	20
2. Constituency.....	23
2.1 Introduction.....	23
2.2 The Problem of Contradictory Constituency.....	24
2.3 Constituency in Structure Building.....	27
2.3.1 A left-to-right derivation.....	27
2.3.2 Prediction I: Uniform C-command.....	33
2.3.3 Prediction II: Left-edge constituency.....	34
2.3.4 Prediction III: Parallelism.....	34
2.4 C-Command Tests.....	35
2.4.1 Binding.....	35
2.4.2 Scope.....	36
2.5 Linear Order and Left-edge Constituency.....	39
2.5.1 VP-Fronting.....	39
2.5.2 PP-movement.....	44
2.5.3 Right Node Raising.....	48
2.5.3.1 Disguised Clausal Coordination.....	49
2.5.3.2 Right Node Non-Raising.....	51
2.5.3.3 Factorization and Ordering.....	54
2.5.4 Summary.....	58
2.6 Constituency Conflicts and Parallelism.....	59
2.6.1 An Asymmetry between VP-Fronting and VP-Ellipsis.....	60
2.6.2 Scope and Ellipsis in Japanese.....	67
2.6.3 Is Branch Right Violable?.....	69
2.6.4 Comparative Ellipsis.....	70
2.6.5 Verb-preposition units.....	75
2.7 On the Necessity of Right-Branching Structures.....	77
2.7.1 Arguments for Right-Branching VPs.....	77
2.7.2 Arguments for PP-splitting.....	78
2.8 Alternative Approaches to Contradictory Constituency.....	80

Introduction

2.8.1 Points of Agreement and Disagreement.....	80
2.8.2 Categorical Grammar.....	81
2.8.3 Pesetsky 1995: Cascade and Layered Syntax.....	85
2.9 Conclusion.....	87
Appendix 1: Constraints on Right Node Raising.....	88
A1.1 Case and Right Node Raising.....	88
A1.2 Constraints and Non-Constraints on the Coordinates.....	94
Appendix 2: Heavy NP Shift & Parasitic Gaps.....	97
3. Parsing.....	103
3.1 Issues in Parsing.....	103
3.1.1 Overview.....	103
3.1.2 Representations and transitions.....	104
3.1.2.1 Endpoints.....	105
3.1.2.2 Intermediate representations.....	105
3.1.3 Transitions: ambiguity resolution.....	106
3.1.4 Structural complexity metrics.....	108
3.1.5 Branch Right.....	111
3.2 On the Strength of the Local Attachment Preference.....	114
3.2.1 The Locality Puzzle.....	114
3.2.2 Competing Biases.....	117
3.2.3 Experimental Evidence.....	119
3.2.4 Resolving the Locality Puzzle.....	128
3.2.5 Interim Conclusions.....	131
3.3 Ambiguity resolution in English.....	131
3.3.1 Two types of structural ambiguity.....	132
3.3.2 Association ambiguities I.....	133
3.3.3 Association ambiguities II.....	137
3.3.3.1 Multiple complement verbs.....	137
3.3.3.2 PP-attachment: argument-by-category interaction.....	141
3.3.4 (In)dependence ambiguities.....	144
3.4 Some attachment preferences in head-final languages.....	147
3.4.1 Head position and PP-attachment in German.....	148
3.4.2 Japanese.....	152
3.4.2.1 How left-branching is Japanese?.....	153
3.4.2.2 How ambiguous is Japanese?.....	157
3.4.2.3 Japanese syntactic ambiguity resolution.....	160
3.4.2.4 Recovery from error.....	161
3.5 Some residual questions.....	163
3.5.1 Fragility of the predictions of Branch Right.....	163
3.5.2 One more anti-locality effect: complex NPs.....	164
3.6 Modularity.....	165
3.7 Conclusion.....	166
Appendix: Experimental Materials.....	167

Introduction

4. Morphosyntax.....	169
4.1 Introduction: Composition and Decomposition.....	169
4.2 Movement and Projection of Heads.....	172
4.2.1 The C-command Condition on Movement.....	172
4.2.2 Head Projection.....	176
4.3 (Non-)Locality in Head Movement.....	176
4.3.1 Movement and Decomposition inside VP.....	178
4.3.1.1 Two restrictions on causer arguments.....	178
4.3.1.2 Heavy PP Shift voids the restrictions.....	186
4.3.2 (Non-)locality in Participle Movement.....	188
4.3.2.1 Long Head Movement.....	188
4.3.2.2 Long-Distance Participle Fronting Crosslinguistically.....	192
4.3.2.3 Filled Comp blocks LHM in Serbo-Croatian.....	197
4.3.3 Local Head Movement.....	203
4.4 Effects of Ordering.....	206
4.4.1 Serbo-Croatian Clitics.....	207
4.4.1.1 Interfacing Phonological and Syntactic Structures.....	207
4.4.1.2 Determinants of Clitic Placement.....	214
4.4.2 A Linear Asymmetry in Clitic Placement.....	223
4.4.3 The Adjacency Penalty on Head Lowering.....	227
4.5 The Syntax of Intonation in O'odham.....	229
4.5.1 Basic Facts.....	229
4.5.2 Clitic Auxiliaries and Intonation Phrases.....	232
4.5.2.1 XP and X0 as First Position.....	233
4.5.2.2 XP-internal Auxiliaries in Wh-Questions.....	235
4.5.3 Comparison with Hale & Selkirk 1987.....	239
4.5.4 Clitic (un-)placement again.....	241
4.6 Conclusion.....	243
Appendix: More on Participle Fronting.....	243
A.1 Polish.....	243
A.2 Old Spanish.....	246
A.3 Breton.....	249
5. The Parser and the Grammar.....	253
5.1 Putting the Pieces Together.....	253
5.2 The Parser is the Grammar.....	255
5.3 Objections to the PIG Model.....	256
5.3.1 The Grammar couldn't be a Parser.....	257
5.3.1.1 Analysis-by-synthesis.....	257
5.3.1.2 Analysis-by-analysis.....	258
5.3.2 The Derivational Theory of Complexity.....	262
5.3.2.1 The Received View of the DTC.....	263
5.3.2.2 Results Supporting DTC.....	264

Introduction

5.3.2.3 Some Concerns about the DTC Experiments.....	266
5.3.2.4 Transformations that Reduce Perceptual Complexity.....	267
5.3.2.5 Other reevaluations of the DTC experiments.....	268
5.3.2.6 Conclusion.....	269
5.3.3 There are Non-Grammatical Parsing Strategies.....	270
5.4 More Objections to Parser–Grammar Identity.....	272
5.6 On the Competence–Performance Distinction.....	275
5.7 Conclusion.....	277
Appendix: Reevaluating Criticisms of the DTC.....	278
Miller & McKean 1964.....	278
Passivization and Reversibility (Slobin 1966).....	283
Transformations and Memory Capacity.....	285
Studies of other transformations.....	287
References.....	291

Chapter 1

Introduction

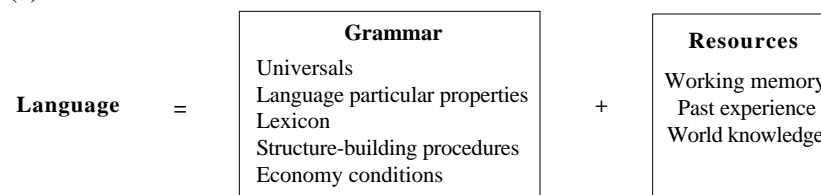
The main aim of this thesis is to argue for the following two claims. First, natural language grammars construct sentences in a strictly left-to-right fashion, i.e. starting at the beginning of the sentence and ending at the end. Second, there is no distinction between the grammar and the parser. In other words, we perceive sentences by generating them for ourselves.

These claims are both rather mundane, but they are anything but standard assumptions about grammar and parsing. Most work on syntax does not assume that sentences are constructed from left-to-right, and most work on parsing assumes a parser–grammar distinction. Since most of the thesis will be taken up with discussion of specific issues internal to the traditional areas of syntax and parsing, the role of this introductory chapter is to explain the overall plot of the thesis, so that the reader may know where I am headed. The main arguments about syntax and parsing will just be presented in outline here, and discussion of general issues of the architecture of language will be kept brief, as I will return to a more detailed discussion of these issues in Chapter 5.

1.1 General Architecture

The general view of the architecture of language that I will be arguing for in the chapters that follow is sketched in (1). It has two components: a grammar and a finite set of resources which the grammar uses. This view is very similar to the model proposed by Miller & Chomsky in 1963, in which there was little or no role for a parsing system distinct from the grammar. For want of a better name, I will refer to this as the *PIG model* of language (Parser Is Grammar).

(1)



The PIG Model

Introduction

I should clarify at the outset what I mean by the term *parser*. I take this to refer specifically to the structure building system that it used in sentence recognition, and not to the many other psychological processes involved in understanding sentences. Sentence *parsing* should not be confused with sentence *comprehension*, which is a complex cognitive act involving the integration of many different sources of information (language, world-knowledge, expectations, attentional state etc.). The parser is just one of the systems involved in language comprehension, and in fact might not always be involved in comprehension of linguistic acts.

In the PIG model the distinction between sentences which are *grammatical* and sentences which are *parsable* is just the distinction between those sentences which the grammar could generate given potentially unbounded resources and those sentences which the grammar can generate given a certain limited set of resources. In other words, grammaticality is parsability in the limit.¹

The steps of parsing a sentence can be seen as proceeding as follows. Parsing is an *active* process, in which the grammar tries to generate a sentence whose phonetic form matches the incoming sentence, using the normal structures and operations of the grammar. If the grammar can find a structural description and meaning to pair with the sound input, then the incoming sentence is successfully recognized. If, on the other hand, the grammar fails to generate a matching sentence, either because it does not generate a match in principle, or because generating a match exceeds the available resources, then recognition fails. This is what is known as an *analysis-by-synthesis* model of sentence recognition.²

Therefore, sentences are not inherently parsable or unparsable, rather they are parsable (or not) given a certain set of resources, where the resources can include short and long-term memory, expectations among other things. Meanwhile grammaticality is just the name given to the special case of parsability in which resources are unbounded. It is in this sense and only this sense that grammaticality represents an abstraction from the steps of generating and comprehending sentences. Apart from the idealization of unbounded resources, grammaticality and parsability are just the same.

¹ It is well-known that there are sentences which are judged to be ungrammatical but which are quite easy to understand, such as violations of the *that-trace* filter (**who do you think that left?*) or violations of restrictions on double object constructions (**The collector donated the museum a painting*). The existence of such cases is sometimes taken to show that grammaticality and parsability are necessarily independent properties. However, this conclusion is the result of identifying parsing with comprehension. It is clear that such sentences are understood, but it is not clear that the parser generates a complete structure for them. This issue is taken up in more detail in Chapter 5.

² Analysis-by-synthesis models of parsing with the grammar were briefly explored in the early 1960s (cf. Matthews 1962), but mostly ignored after that. However, analysis-by-synthesis has remained the basis of a respectable (albeit highly controversial) approach to speech perception, often under the rubric of the *Motor Theory of Speech Perception* (cf. Liberman 1957, Halle & Stevens 1964, Liberman et al. 1967, Liberman & Mattingly 1985, Fowler & Rosenblum 1991: for a brief summary of this research see the relevant section of Remez 1994).

Introduction

Nevertheless, the view that the parser is the grammar conflicts with a broad consensus of opinion since the late 1960s, which holds that the parser is not the same as the grammar. Instead there are two distinct but related structure building systems in the language faculty. The main reasons for this view are the following, all of which are discussed in greater detail in Chapter 5, and which received their classic formulation in Fodor, Bever & Garrett 1974:

- Standard models of grammar cannot be implemented as a sentence recognition device which can successfully recognize sentences in finite time.

This is because the grammar is standardly viewed either as a mapping from underlying structures to surface (phonetic) representations (transformational theories), or as a constraint set which applies to fully-built representations (non-transformational theories). Given these models of grammar, the steps of an incremental parser cannot correspond exactly to the steps of a grammatical derivation, since the grammar either defines different steps or the grammar does not define derivational steps at all. An analysis-by-synthesis implementation of such grammars can therefore only recognize sentences by randomly generating enormous sets of generated sentences in search of a match—which is obviously quite inefficient and unrealistic. The only way of narrowing down the search is to add an extra ‘preprocessor’ to the model in (1), which performs preliminary analysis of incoming sentences. Once spelled out, though, this preprocessor tends to take over most of the work of processing the sentence, and effectively reduces the role of the grammar in sentence recognition.

Note that while this argument was initially formulated in the context of an *Aspects* style transformational grammar, it applies to the vast majority of other theories of grammar that have been proposed, whether or not they assume transformations, or even phrase structure. Any grammar that does not specify an incremental left-right mapping from surface strings to structural descriptions will face the same difficulties if used as an analysis-by-synthesis sentence recognition device.³

- Something very similar to the PIG model is widely considered to have been experimentally disconfirmed in the 1960s.

From the early 1960s onwards a number of experimental studies were undertaken to test whether the operations proposed in transformational grammars of the time had a measurable effect on sentence comprehension or recall. The received view of the outcome of these studies is that they disconfirmed the view that the operations of the parsing device and transformational grammars were the same (a.k.a. *Derivational Theory of Complexity*: cf. Fillenbaum 1971, Fodor et al. 1974, Levelt 1974, vol. III, Berwick & Weinberg 1983, Bever 1988, Wanner 1988 for reviews).

³ Two exceptions to this are augmented transition network (ATN) grammars (e.g. Woods 1970) and certain versions of Categorical Grammar which include type-raising rules which make left-to-right assembly of structures a possibility, though not a necessity (e.g. Ades & Steedman 1982).

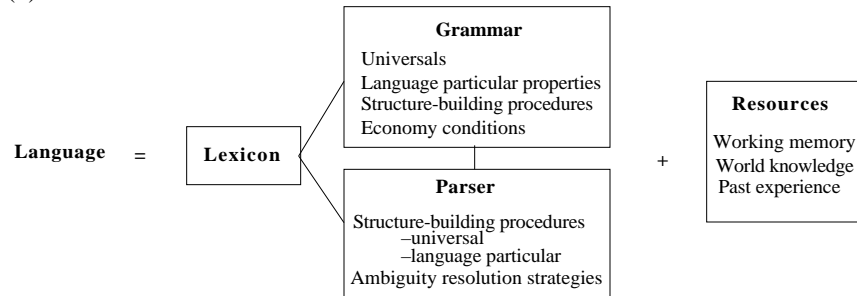
Introduction

- Properties of the parser can be observed which cannot be reduced to properties of the grammar.

These first two considerations made it seem necessary that there be a sentence processor of some kind in addition to the grammar. This supposition received further support from investigations of sentence processing which pointed to the existence of a number of sentence processing principles which are not obviously related to the grammar, such as ambiguity resolution strategies and phrase-boundary location heuristics.

Considerations such as these,⁴ which will be discussed in rather more detail in Chapter 5, led to a different conception of human linguistic capacities, which included at least the components in (2).⁵

(2)



The relevance of the grammar to the operations of the parser varies greatly from theory to theory, and the amount of internal structure that is attributed to the parser varies greatly, as does the way in which the parser accesses grammatical knowledge, but the basic picture appears to have been generally agreed upon since the mid-1960s. One effect of this has been that the study of the grammar and the study of the parser have fractionated into separate disciplines, with the result that the issue of whether the simpler model in (1) is

⁴ In addition to these empirical reasons for distinguishing the parser and the grammar, it is sometimes claimed that there exists an *a priori* distinction between the parser and the grammar. Consider, for example, the claim in a recent book on sentence processing (Crocker 1996) that ‘This *grammar as parser* approach is not a rational position given the competence–performance division [...] which clearly separates the declarative properties of the syntactic theory from any procedural notions’ (p.49), and that ‘The suggestion that the grammar *is* the parser is simply not well-conceived’ (p.51).

However, such claims that the parser and the grammar are necessarily distinct are based on *a priori* distinctions made in theories of formal grammars, which do not automatically apply to the study of human linguistic abilities, where distinctions between data structures and procedures require empirical justification.

⁵ The model in (4) separates the lexicon from both the grammar and the parser, in order that it may be independently accessed by both the grammar and the parser. It could also be treated as a subcomponent of the grammar.

Introduction

possible has effectively been closed, more by boundaries between disciplines than by actual argumentation.

I will not take up any further space at this point to discuss reasons for adopting the view of language in (1). Instead I will first try to demonstrate that the parsing-compatible features of the grammar that I adopt are well-motivated based on considerations of grammar alone, and that the grammar-compatible features of the parser that I adopt are well-motivated based on considerations of parsing alone. Having shown that the parser and the grammar look very much alike even in advance of considerations of how they interact, it will make much more sense to return to the issues raised here about the parser–grammar relation.

There are two main components to this argument, both of which involve making non-standard assumptions about the form of the grammar and the parser, and both of which take away the force of the classic arguments against viewing the parser and the grammar as the same.

First, I reexamine standard views of how the grammar builds syntactic structures (either not at all, or from bottom-to-top), and argue that structures are built from left-to-right, based on evidence from constituency tests and ordering asymmetries. If this conclusion is correct, then the main argument against using the grammar as an analysis-by-synthesis recognition device goes away, because the grammar specifies an incremental left-to-right mapping from surface strings to structural descriptions.

Second, I look at ambiguity resolution strategies, an area which is generally viewed in terms of an independent parser. I argue that structural complexity metrics in parsing, which contribute to what is easy to understand and what leads to garden paths, reduce to an independently motivated economy principle of the grammar, which favors the building of right-branching structures where possible.

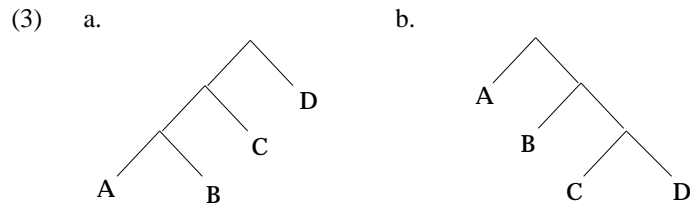
These arguments address the first and the third objections to the PIG model given above, and discussing them takes up the greater part of the thesis. However, I also address the second argument against the PIG model, surrounding the so-called ‘psychological reality’ of transformational grammars, and the claim that there is no evidence for their operations in parsing.⁶ I show that the evidence for this argument was never particularly strong, and is even weaker now than it was in the 1960s.⁷ Moreover, it was never the most important argument against the pig model—the other two arguments were always more important, although they received less attention.

⁶ The term ‘psychological reality’ is an unfortunately loaded and misleading expression. It generally refers to whether a given theoretical construct can be shown to have a measurable effect given the psychologist’s standard battery of tools, typically involving reaction time measurements. I use it here only as a convenient cover term for a certain line of research, and do not intend any further endorsement of what the term implies, which is that some kinds of evidence are inherently more privileged than others.

⁷ In any case, a number of different ways of circumventing this argument have been given before in the literature, mostly involving changes to the form of the parser or the grammar or both (cf. Bresnan 1978, Berwick & Weinberg 1983, 1984, Pritchett & Whitman 1993, 1995).

1.2 Outline

The thesis is organized as follows. Chapter 2 argues that syntactic structures are built in a strictly left-to-right fashion. The evidence for this comes from a study of apparent contradictions between the results of different constituency tests. These diagnostic tests include coordination, movement, ellipsis, coreference, disjoint reference and the licensing of bound variables and polarity items. Some of these diagnostics make sentences (or parts of sentences, eg. VP) appear to have a left-branching structure (3a), others make sentences appear to have a right-branching structure (3b), and yet others yield both results.



The only existing approaches to these problems have required the adoption of either dual representations for all sentences (Pesetsky 1995⁸, Brody 1994⁹) or the flexible constituency of some versions of categorial grammar which effectively allows multiple constituent structures for any sentence (cf. Ades & Steedman 1982, Steedman 1996, Dowty 1988, Pickering & Barry 1993 for approaches and applications). I show that the problem of contradictory constituency does not arise if it is assumed that syntactic structures are constructed incrementally from left-to-right, as dictated by the condition *Merge Right*, and if structure building is subject to the economy condition *Branch Right*.

- (4) MERGE RIGHT
New items must be introduced at the right edge of a structure.

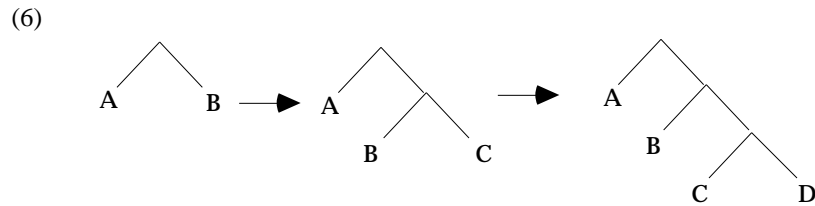
⁸ Pesetsky (1995) in fact suggests that the right-branching *Cascade* structures of his dual structure theory might be transient structures, resulting from the preliminary parse of a sentence, and that the relatively left-branching *Layered* structures are the only ‘real’ structure for a sentence. Thus, sentences have two structures, but one of them is only tentative. The theory of Chapter 2 goes a step further than this, in that only one structure is built for any sentence. See Chapter 2, Section 8 for further comparison of the two approaches.

⁹ Brody’s theory in fact only assigns a single tree structure representation to any sentence, but a second representation is given for each sentence in terms of ‘dependency’ relations, which are independent of phrase structure relations. In this way something very similar to Pesetsky’s distinction between Cascade and Layered structures is recreated.

Introduction

- (5) **BRANCH RIGHT**¹⁰
Metric: select the attachment that uses the shortest path(s) from the last item in the input to the current input item.
Reference set: all attachments of a new item that are compatible with a given interpretation.

The effect of Merge Right is that a structure like (3b) has the derivational steps in (6).



All of the strings listed in (7) are constituents at some point in the derivation in (6). Notice that although (6) builds a right-branching structure, all of the strings that are constituents in the left-branching structure in (3a) are constituents at some point in the derivation in (6).

- (7)
- | | |
|------|-----|
| AB | BC |
| ABC | BCD |
| ABCD | CD |

Motivation for Merge Right and Branch Right is drawn from evidence for intermediate constituents of derivations like (6), and from evidence for the restricted distribution of contradictory constituency effects, which is predicted by the left-to-right theory, but not by theories which invoke multiple parallel representations.

Furthermore, I show that the Merge Right/Branch Right approach to structure building provides more than a restatement of the effects of multiple structure theories, because it explains why different constituency tests yield the results they do, and where constituency contradictions should and should not be found. The arguments in this chapter are drawn primarily from English.

In Chapter 3 I focus on parsing, in particular on the topic of ambiguity resolution, which has been the focus of most work in sentence processing over the last 20 years. I argue that the syntactic component of ambiguity resolution can be reduced to the principle *Branch Right*, which favors the construction of maximally right-branching structures, all other things being equal. Branch Right is closely related to the local attachment preference that almost all models of parsing incorporate (e.g. *Right Association*, Kimball 1973; *Late Closure*, Frazier 1978). My claim is therefore that all structural biases in parsing can be

¹⁰ The effect of shortening the paths between adjacent terminal elements in a phrase marker is to create a more right-branching structure. Why this is so, and the details of how Branch Right is applied is discussed in more detail in Chapter 3.

Introduction

reduced to the same principle that causes a bias to interpret the adverbial *yesterday* inside the lower clause in well-known ambiguities like (8).

(8) John said Bill left yesterday.

I show that Branch Right can account for a wide range of parsing biases which are normally attributed to other strategies, and present results from an experiment on a novel structural ambiguity, which pits the choices of Branch Right against the choices of well-known parsing strategies such as *Minimal Attachment* (Frazier & Fodor 1978, Gorrell 1995), and *Attach Arguments* (Ford et al. 1982, Abney 1989, Crocker 1996, Schütze & Gibson 1996). The experimental results show that given the choice between a local attachment which is structurally more complex, not supported by discourse and not required by syntax or semantics, and a non-local attachment which is structurally 'simpler' and involves an obligatory syntactic constituent, the parser opts for the more local attachment, as predicted by Branch Right. The evidence in this chapter is again drawn primarily from English, but includes some discussion of ambiguity resolution in German and Japanese.

Chapter 4 returns to issues in the traditional domain of grammar, and extends the discussion of the left-right grammar to issues at the interface of syntax with morphology and phonology. In this model morphophonological representations are mapped onto surface syntactic structures, which in turn are mapped onto underlying syntactic structures. This ordering is the opposite of what is assumed in most theories, and is forced because of the fact that surface positions of words are generally to the left of their underlying positions, and are therefore built earlier in a left-to-right derivation. I show how it is possible in this approach to give a uniform treatment of local and non-local head movement and clitic placement operations. I show how certain left-right asymmetries in both head movement and clitic placement are predicted by the theory. I also discuss some issues involving the relationship between phonological phrasing and syntactic constituency, and show how it might be possible to draw a closer connection between phonological and syntactic constituency than is commonly assumed. This chapter draws more on cross-linguistic evidence than the earlier chapters. Topics covered include 'long head movement' in Slavic, Romance and Breton, clitic placement in Serbo-Croatian and phonological phrasing in Tohono O'odham (Papago), among others.

Chapter 5 draws together the arguments developed independently in Chapters 2–4 and returns to the issues raised in this chapter involving the general architecture of language. I show that the best objections to the PIG model of language no longer apply, given the view of the parser and the grammar developed in Chapters 2–4. This chapter also discusses other issues concerning the parser–grammar relation, including the competence–performance distinction, and some further arguments against the PIG model.

1.3 Some reminders

Before proceeding, I should emphasize at the outset a couple of things that I am *not* trying to do here, and one thing that I am trying to do.

Introduction

First, my agenda here should not be mistaken for an attempt to give a ‘parsing explanation’ for grammatical phenomena. The literature contains a number of arguments which run something like: ‘phenomenon X is generally considered to be a property of grammar, but it in fact is better explained in terms of properties of the parser’ (for examples, see Fodor 1978, Hankamer 1973, Kuno 1973, Dryer 1980, Hawkins 1995, Berwick & Weinberg 1984, Alphonse & Davis 1992, 1996, Pritchett 1991, Fox 1996). What I am trying to do here is quite different. Given that I am claiming that the parser and the grammar are the same system, the traditional kind of reassignment of work from the grammar to the parser is mostly unavailable to me. Of course, given the distinction between language and resources it is still possible to distinguish between unacceptability of a sentence due to the grammar or due to resource limitations.

Second, readers who are at all familiar with the recent experimental literature on sentence processing will be aware of the fact that an enormous amount of the research done on parsing is directed to exploring questions of *modularity*; in other words, to what extent are different sources of linguistic and non-linguistic information used in sentence comprehension, and what is their relative importance. Note that my focus in this thesis on the direct implementation of the grammar as a parsing device entails no commitment whatsoever regarding the modularity issue. Adopting the PIG model commits me to the claim that sentence parsing involves building representations that are sanctioned by the grammar, but this says nothing about how ambiguities are resolved in situations which are lexically or pragmatically biased—which is where much of the action has been in the modularity debate. In Chapter 3 I discuss briefly the current status of the debate on the informational encapsulation of the parser, but this question is logically quite independent of the issue of whether there are distinct syntactic systems used in parsing and grammar.

Thirdly, I am trying to do more than show that an incremental parser can be built whose operations are more or less transparently related to the operations of the grammar. This possibility has been amply demonstrated for a variety of grammatical formalisms in the computational literature on parsing.¹¹ In incremental parsers based on standard grammatical models, the intermediate stages of a parse are not grammatically defined objects, and are not expected to play any role in grammatical phenomena. One of my main aims here is to show that the intermediate structures built by a left-to-right grammar play a crucial role in certain grammatical phenomena, and that grammatical derivations can therefore *only* proceed from left-to-right.

Finally, this should go without saying, but the overall aim of this thesis is to address issues in parsing and grammaticality together. The chapters on syntax and parsing are mostly written so that they may be read independently, but to overlook the similarity between the issues that arise in parsing and grammar would be to miss the main point, which is that these are not separate lines of inquiry.

¹¹ For parsing models based on government-binding theory see Wehrli 1988, Stabler 1992, Johnson 1989, Berwick & Fong 1995, Berwick & Epstein 1995, Crocker 1996, among others.

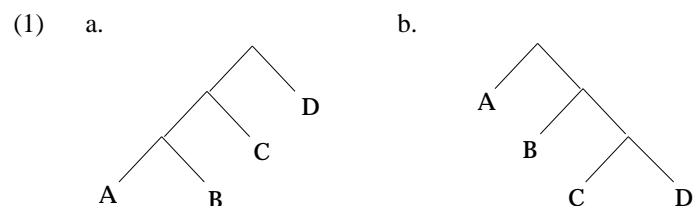
Introduction

Chapter 2

Constituency

2.1 Introduction

Diagnostics of constituency typically test for what strings of words can be coordinated, moved or elided, and which pairs of phrases can enter into relationships of binding, disjoint reference or other dependencies. A problem that often arises in syntactic research is that faithful application of the constituency tests leads to situations where the results of one test contradict the results of another. Some diagnostics make sentences appear to have a relatively left-branching structure of the kind shown schematically in (1a). Other diagnostics, meanwhile, make the same sentences appear to have a much more right-branching structure, as in (1b).



The problem of contradictory constituency poses a serious problem for one of the leading ideas of phrase structure grammar which I will call the *Single Structure Hypothesis*. This is the hypothesis that a wide range of otherwise unrelated syntactic processes all refer to the pieces of a single constituent structure or derivation for any sentence.

Existing approaches to the problem of contradictory constituency have adopted one of two approaches. On the one hand, some have attempted to dismiss the problem by arguing that some of the diagnostics of constituency are either misleading or have been misinterpreted. On the other hand, a number of people have recently argued that the conflicts between different constituency tests provide evidence for multiple parallel phrase structures (Pesetsky 1995, Brody 1994) or for the flexible constituency allowed by enriched categorial grammars (Steedman 1985, 1988, in press; Dowty 1988; Pickering & Barry 1993).

The aim of this chapter is to suggest a different kind of solution to the contradictory constituency problem. I argue that the problem of contradictory

Constituency

constituency does not arise in a system in which phrase markers are derived by building from left-to-right, i.e. starting at the beginning and ending at the end. The requirement that new material is always added at the right of the phrase marker is imposed by a principle which I call *Merge Right* (2).

- (2) MERGE RIGHT
 New items must be attached at the right edge of a structure.

Although the theory of phrase structure which I assume is in other respects rather standard, the effects of Merge Right are far reaching. The main consequence of left-to-right derivations, which I focus on in this chapter, is that the strings that form constituents at intermediate stages in the derivation are different from the constituents of more orthodox bottom-to-top derivations. This fact makes it possible to derive the effects of contradictory constituency without assuming parallel representations or flexible constituency. As I show in §2.2 and §2.3, the appearance of constituency conflicts is just the consequence of how structure changes over the course of a left-to-right derivation of a single right-branching phrase marker.

More important than just being able to describe in a single derivation what had previously appeared to be contradictions between the results of different constituency tests, the account based on left-to-right derivations begins to provide an account of *why* each constituency test yields the results it does, and makes novel predictions about which kinds of tests will be able to diagnose which kinds of constituents. §2.4–§2.6 test these predictions and demonstrate a number of correlations between how a constituency test probes for structure and the kinds of results it produces.

In §2.7–§2.8 I compare the results of the Merge Right approach to other existing approaches to the problem of contradictory constituency, including attempts to deny that there really is a problem at all.

A further consequence of Merge Right is that in almost reversing the order in which syntactic derivations are standardly assumed to occur—because of its left-to-right nature—the grammar proposed here effectively computes from relations which traditionally hold at S-structure to relations which traditionally hold at D-structure, rather than vice versa. This means that movement operations are generally rightward and downward rather than leftward and upward, as they are in most transformational grammars. This ordering of derivations also opens the possibility of far greater similarity between the operations of the parser and the grammar.

2.2 The Problem of Contradictory Constituency

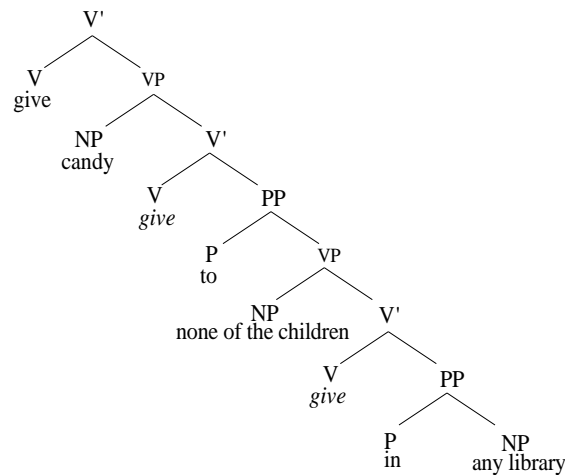
As an illustration of the problem of Contradictory Constituency, consider first the sentence in (3), and the constituency tests that have been applied to it in (4) and (5). The tests of negative polarity item licensing and coordination in (4–5) point to an extremely *right-branching* VP-structure, such as in (6), which corresponds to the *Cascade* structures proposed in Pesetsky 1995.

- (3) John gives candy to children in libraries on weekends

Constituency

- (4) a. John gave **nothing** to **any** of my children in the library on his birthday.
 b. John gave candy to **none** of my children in **any** library on his birthday.
 c. John gave candy to children in **no** library on **any** public holiday.
 d. * John gave **anything** to **none** of my children in the library on his birthday.
 e. * John gave candy to **any** of my children in **no** library on his birthday.
- (5) a. John gives [candy to children on weekends] and [money to homeless people on weekdays.]
 b. John gives money [to children on weekends] and [to homeless people on weekdays.]
 c. John gives candy to [children on weekends] and [homeless people on weekdays.]

(6)



The facts in (4–5) motivate the structure in (6) based upon the assumption that negative polarity item licensing requires c-command or m-command,¹² and that coordinability is an indicator of constituenthood (these assumptions are standard, though by no means necessary: see §2.7 for further discussion). Using this reasoning we are led to the conclusion that the complex VP in (3) is right-branching to such an extent that the complement of a preposition forms a constituent with the *following* PP, to the exclusion of the preposition that selects it. The evidence for this is that an NP can c-command an element outside of the PP that it is generally thought to be inside, such that it can license a polarity item in the immediately following PP, as in (4b–c). Similarly, an NP

¹² X c-commands Y iff all nodes that dominate X also dominate Y, and X does not dominate Y. X m-commands Y if all maximal projections that dominate X also dominate Y, and X does not dominate Y.

Constituency

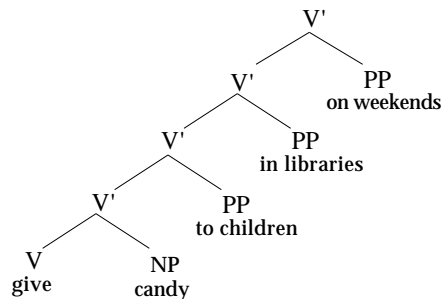
can form a conjunct which includes the PP that follows it, but excludes the preposition that selects it, as in (5c).

Based on these tests, then, rightwards roughly equals downwards in the phrase structure tree. A number of other structural diagnostics yield the same pattern of results, including anaphor binding, disjointness (Condition C effects), weak crossover and bound variable anaphora (cf. Barss & Lasnik 1986, Stroik 1990, 1996, Pesetsky 1995).

Contrasting with the evidence for right-branching structures, meanwhile, certain kinds of movement tests point to a left-branching structure for the very same VP, as can be seen from the examples of VP-fronting in (7). The basic generalization in this case is that any string of phrases starting from the *left* edge of VP can be fronted (7a–d). Strings of phrases that do not include the left edge of VP cannot be fronted (7e–f). If we assume that the strings that can front are constituents, then the results of this test point to a left-branching structure like (8), which is the kind of structure traditionally assumed for VPs containing multiple modifiers.

- (7)
- a. John intended to give candy to children in libraries on weekends,
... and [give candy to children in libraries on weekends] he did ____.
 - b. John intended to give candy to children in libraries,
... and [give candy to children in libraries] he did ____ on weekends.
 - c. John intended to give candy to children,
... and [give candy to children] he did ____ in libraries on weekends.
 - d. ... and [give candy] he did ____ to children in libraries on weekends.
 - e. * ...and [to children in libraries] he did ____ give candy on weekends.
 - f. * ...and [in libraries on weekends] he did ____ give candy to children.

(8)



There therefore appears to be a conflict between the results of the polarity item licensing and coordination tests in (4–5) and the results of the movement test in (7). This kind of conflict is the basis of the contradictory constituency problem.¹³ In fact, this conflict is sharpened by the fact that we find diagnostics

¹³ There are, in fact, some fairly standard ways of avoiding the contradiction between (4–5) and (7), typically based on a reinterpretation of the binding and coordination results as involving relations other than c-command and constituency, e.g., binding is taken to require precedence and c-command (Barss & Lasnik 1986,

Constituency

for both left- and right-branching structures satisfied in a single sentence, as in the sentences in (9), taken from Pesetsky 1995 (p.230), in which sequences of phrases starting at the left-edge of VP have been fronted, implying the kind of structure in (8), but the fronted portion of VP contains an NP which binds a reciprocal in the stranded portion of VP, implying a right-branching VP structure more along the lines of (6).

- (9) a. ...and [give the books to **them** in the garden] he did ___ on **each other's** birthdays.
b. ...and [give the books to **them**] he did ___ in the garden on **each other's** birthdays.

Notice, however, an important step in the reasoning that leads to the constituency conflict. What the results of the movement test in (7) show is that *give candy* is a constituent, that *give candy to children* is a constituent, that *give candy to children in libraries* is a constituent, and so on. The standard way of representing the fact that each of these strings is a constituent is to assign them the nested, left-branching structure in (8). But this inference is by no means necessary, particularly if we assume left-to-right structure building, as the next section shows.

On the other hand, the binding and coordination tests in (3–5) provide convergent evidence for the right-branching structure in (6). The right-branching structure is motivated by evidence from both constituency tests (i.e. tests that ask: ‘is this string a unit?’) and c-command tests (i.e. tests which ask what the relative hierarchical relation of two units is).

2.3 Constituency in Structure Building

2.3.1 A left-to-right derivation

(10) shows a very much simplified version of how the sentence *The man saw Mary* is built up in the theory outlined in Chomsky 1995a. The relevant property of this kind of derivation is that it proceeds largely from bottom-to-top, as dictated by the *Strict Cycle Condition* (Chomsky 1993).¹⁴ When new items are added at the top of the tree new constituents are created, but existing constituents are preserved from each step of the derivation to the next. Inflectional material and functional projections have been omitted from this

Jackendoff 1990, Ernst 1994), coordination is replaced by NP or S coordination followed by ‘conjunction reduction’ (eg. Hudson 1976, Wilder 1994).

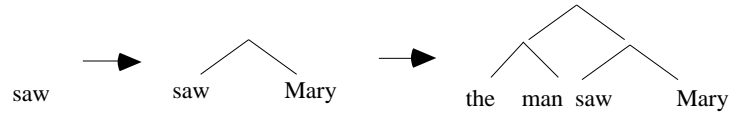
Traditionally, the results of movement tests have tended to be taken most seriously, and the results of other tests have been made to fit with these. In §2.7–§2.8 I discuss alternative approaches to constituency conflicts and alternatives to right-branching VP-structures; but until that point I will continue to just assume the interpretation of the binding and coordination tests given in the text. The reader who is reluctant to grant me this liberty is encouraged to look ahead to §2.7.

¹⁴ The *Strict Cycle Condition* is referred to in some of the literature as the *Extension Condition* (cf. Chomsky 1993).

Constituency

derivation for the sake of simplicity, but I assume that they are added to the structure in the same way as lexical material.¹⁵

(10)



The strings that are constituents at some point in the derivation in (10) are listed in (11). Note that these are exactly the constituents of the final structure in (10).

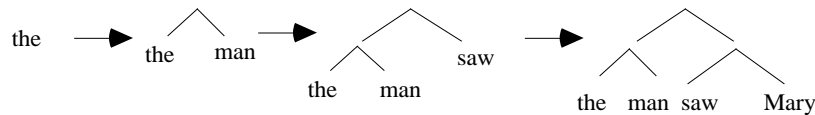
- (11)
- | | |
|------|------------------|
| the | the man |
| man | saw Mary |
| saw | the man saw Mary |
| Mary | |

But now consider what happens if instead of always adding new material at the top of the tree, structures are built in a strictly left-to-right fashion, so that new material is always added at the right-hand edge of a tree. Let us assume that this requirement is imposed by the condition *Merge Right*, given in (12).

- (12) **MERGE RIGHT**
 New items must be attached at the right edge of a structure.

A simplified derivation of *the man saw Mary* in this left-to-right manner is shown in (13). As in (10) inflectional material has been omitted for ease of exposition. The important thing to notice here is the difference between the third and fourth steps in the derivation. In the second step, at which point the verb is the rightmost element in the structure, the subject and the verb form a constituent. But once the object is added to the structure, the subject and the verb no longer form a constituent. At this point the verb and the object form a constituent, as in the structure traditionally assumed for English SVO sentences.

(13)



(14) lists the strings that are constituents at some point in the derivation in (13). The final structure is identical to the one built in (10), but the list in (14) includes one string which is not a constituent in the final structure in (13), namely *the man saw*.

¹⁵ The notation X(P) denotes a node which is of category X and is both the maximal and the minimal projection of that category.

Constituency

- (14) the the man
 man **the man saw**
 saw saw Mary
 Mary the man saw Mary

Therefore, two unusual properties of derivations that respect Merge Right are the following. First, in the construction of a right-branching structure some constituents are created during the derivation which are not constituents in the completed (final) phrase marker. This fact is the key to being able to describe contradictory constituency effects without recourse to multiple parallel structures or flexible constituency.

Second, the creation of new constituents in left-to-right derivations sometimes has the effect of destroying existing constituents, such as when the addition of the direct object *Mary* to the derivation in (13) created the new constituent *saw Mary*, but destroyed the existing constituent *the man saw*. This property of left-to-right derivations plays an important role in the explanation of *why* different structural diagnostics yield different results.

Before running through the effects of Merge Right for some more involved examples, I should first spell out some additional assumptions that I will be making.

First, I assume that structure building is constrained by the condition *Branch Right*, which forces structures to be as right-branching as possible.¹⁶

- (15) BRANCH RIGHT
 Metric: select the most right-branching available attachment of an incoming item.
 Reference set: all attachments of a new item that are compatible with a given interpretation.

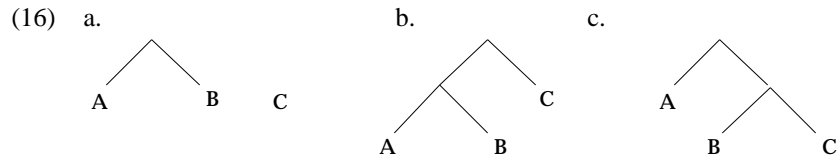
I assume that a structure is ‘right-branching’ to the extent that there is a match between precedence relations among terminal elements and c-command relations among terminal elements. While complete correspondence between precedence and c-command relations is the extreme situation, we can talk about one structuring of a given set of terminals as being *more right-branching* than another structuring of the same set of elements if there is *greater* correspondence between precedence and c-command relations among terminals.¹⁷

¹⁶ The fact that the reference set for Branch Right refers to interpretations commits me to the assumption that the interpretation of a structure is built up incrementally. In what follows I will assume that this is feasible, but I will have little to say about exactly how this is achieved. See Steedman 1996 for an example of a system which could allow incremental left-to-right interpretation.

¹⁷ See Chapter 3 for evidence that Branch Right can account for a wide variety of structural biases in sentence parsing, based on both existing and novel experimental findings. Branch Right is closely related to the principle of *Right Association* proposed by Kimball (1973) to account for the preferred low attachment resolution of ambiguities like *John said that Bill left yesterday*, among other things. It is also closely related to the principle of *Late Closure* (Frazier 1978), and other locality principles in a similar vein.

Constituency

As an illustration, imagine a derivation that has reached the point in (16a), where A and B form a constituent, and C is yet to be added to the structure. Let us assume that C could be attached at the right of the existing structure in two ways without affecting the interpretation of the structure. The two alternatives are shown in (16b) and (16c).



Given the alternatives in (16b) and (16c), Branch Right chooses (16c), because B c-commands C in the more right-branching (16c), but not in the more left-branching (16b). I assume that Branch Right locally determines what is the most right-branching attachment of an incoming item by choosing the attachment that creates the shortest path through the phrase marker from the preceding item to the incoming item.¹⁸ The details of this local way of finding the most right-branching structure will not be important in this chapter, but they are discussed at length in the treatment of parsing in Chapter 3.

I assume that the condition in (17) applies to arguments and predicates. (17) requires that thematic relations be satisfied under sisterhood. It does not require that the thematic relations be satisfied at any specific point in a derivation, and it also does not require that the sisterhood relation be preserved once established.¹⁹

(17) *Configuration for Arguments and Predication*

A head X may discharge a thematic role to a position Y or take position Y as a predicate iff Y is the sister of a head containing X or the sister of a projection of a head containing X.

Finally, I follow Chomsky 1995a,b in assuming that all non-terminal nodes in a phrase marker are branching nodes. In other words, there is no vacuous projection of phrase structure nodes in order to conform to an X-bar template.

With these preliminaries in mind, we are now in a position to see how Merge Right accounts for the appearance of contradictory constituency in complex VPs.

¹⁸ The reason for this is the following. Any path through a phrase marker between two adjacent heads X Y contains an ‘upward path’ to the first node that dominates both X and Y and a ‘downward path’. The more nodes there are in the upward path from X, the more heads there are that precede but do not c-command Y. The more nodes there are in the downward path to Y, the more heads there are that Y precedes but does not c-command. See chapter 3 for further illustration of this point.

¹⁹ The claim that theta-role assigning relations need not be preserved throughout the derivation might seem strange, given the standard assumption that theta-role assignments only matter at the interpretive interface with the syntax. However, I assume here that interpretation is incremental, rather than applying to a ‘complete’ final representation for a sentence. Given this, the transience of theta-role assigning configurations is no longer anomalous.

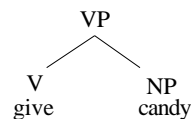
Constituency

(18) shows the steps involved in building the complex VP from the sentence in (3) from left-to-right. The derivation of the VP begins with the verb *give* in (18a). The verb does not project until the noun phrase *candy* is merged to the right of the verb as its sister. At this point in the derivation the verb may discharge one of its theta roles to the NP.

(18) a.

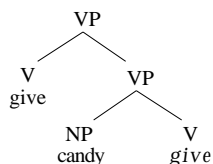
V(P)
give

b.

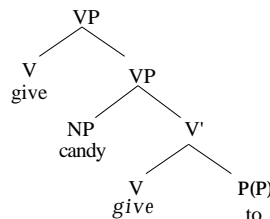


The next step in the derivation involves the addition of the PP *to children*, and is shown in (18c–e). The PP could in principle be merged with the constituent *give candy* in (18b) to form the structure $[[\text{give candy}] \text{to children}]$. However, there is an alternative way of adding the PP to the structure which receives the same interpretation and satisfies the condition in (17) above, and is more right-branching and therefore preferred by Branch Right. First a copy of the verb *give* is generated, which merges at the right of the phrase marker as the sister of the NP *candy* (18c). This copy of *give* is then projected to create an attachment site for the preposition *to* (18d). Then the preposition *to* is projected to allow attachment of the NP *children* as its sister (18e). At this point in the derivation merger satisfies the thematic relation between the P and the NP.

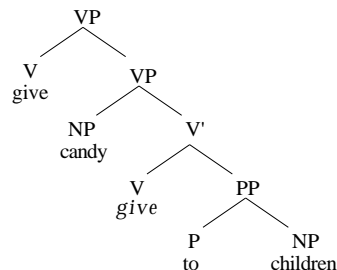
(18) c.



d.



e.



Notice that the structure in (18e) is a right-branching structure very much like the *VP Shell* structures proposed by Larson (1988, 1990). Therefore, right-branching VP-structures have a different motivation here from in other theories.

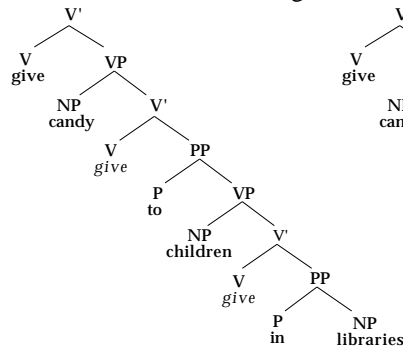
Constituency

Here they are just a consequence of the economy condition Branch Right. In the current theory, left-branching VPs are syntactically well-formed, except when they are blocked by a more economical right-branching VP-structure.

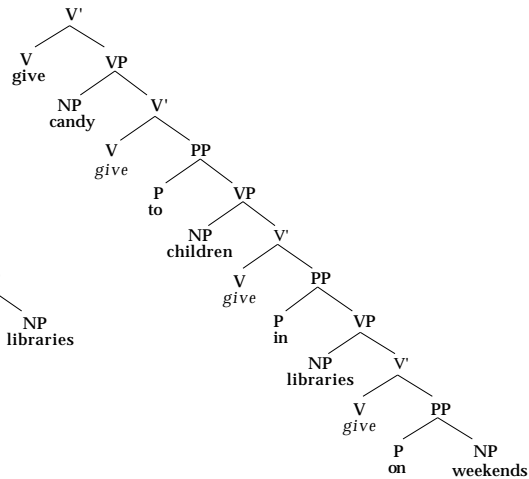
The steps involved in the addition of additional adverbial PPs are shown in abbreviated form in (18f–g). As in (18c–e) the adverbial PPs are merged with the existing phrase marker in such a way as to maximize precedence/c-command correspondences, in accordance with Branch Right.²⁰

(18)

f.



g.



An important property of (18f–g) to notice is what happens to PPs when additional PPs are added to their right in the phrase marker. For example, when the PP *in libraries* is added in (18f), the existing PP *to children* is split up, such that the NP *children* forms a constituent with the following PP, to the exclusion of the preposition that selects it.

The structure that is ultimately built is very similar to the radically right-branching Cascade structures proposed by Pesetsky 1995. However, by building right-branching phrase markers from left-to-right, the system proposed here differs from Pesetsky's system in two important respects.

First, derivations like (18) combine properties of traditional phrase structure theories and Pesetsky's Cascade strictures. Complements of prepositions, for example, enter the derivation as the sister of the preposition, as in traditional

²⁰ The steps of V-lowering that create attachment sites for the adverbials in (18f–g) involve crossing an intervening P head, in violation of the strict locality requirement imposed by the Head Movement Constraint (Travis 1984, Baker 1988). See §4.3.1 for a detailed discussion of where locality conditions do and do not apply to head movement. There I argue that strict locality only applies to operations which (dis)assemble complex heads, but movement of entire heads need not be strictly local, and therefore non-local V-lowering in the examples in (18) is possible.

Constituency

theories, but wind up in the specifier position of a lower projection, as in Pesetsky's Cascade structures.²¹

The second, more interesting difference between this system and Pesetsky's system is that there is no need under the current system to represent left-branching and right-branching structures in parallel. The reason for this is that all the strings that are constituents in Layered Syntax structures are also constituents at some point in the construction of the right-branching structure, although these strings are not always constituents in the final structure. For example, *give candy* is a constituent in (18b), but is no longer a constituent from (18c) onwards.²²

In the light of this derivation, it is useful to reconsider the evidence for left-branching constituency presented in (7) above. The VP-fronting test showed that *give candy* is a constituent, that *give candy to children* is a constituent, that *give candy to children in libraries* is a constituent, and so on. It would be normal to infer from these facts that the VP must have the left-branching structure in (8), but the derivation in (18) shows that this conclusion is not necessary, because all of the strings that can undergo VP-fronting are constituents at some point in the derivation of the right-branching VP.²³ Therefore, the existence of contradictory constituency effects as described in §2.2 does not force us to assume any kind of parallel structure or flexible constituency theory. These effects may therefore be explainable in terms of the derivation of a single structure for any sentence, as we shall see in what follows.

More interestingly, what have traditionally been taken to be the constituents of left-branching structures are in this theory transient stages in the construction of right-branching structures. This generates a series of novel predictions about the relation between structural diagnostics and their results, which are tested in the sections that follow.

2.3.2 Prediction I: Uniform C-command

The first prediction derives from the fact that although constituency can change over the course of a derivation, asymmetric c-command relations are never destroyed once they have been created. This means that we should not expect to find conflicts among structural diagnostics which probe for c-command

²¹ Another difference between theta role assignment in standard theories and this system results from the fact that movement operations are forced to be rightwards and (typically) downwards in the *Merge Right* system, whereas they are generally leftwards and upwards in more traditional approaches. This means that movements that are assumed to originate in argument positions and proceed leftwards into case or operator positions ('scope positions') in standard theories translate in the left-to-right derivations into rightwards movements from case or operator positions into argument positions.

²² Also, there are strings that are constituents in derivations like (18), but which are not constituents in Pesetsky's Layered structures, e.g. *the man saw* in *the man saw Mary*.

²³ There are some strings that are constituents in the derivation of VP in (18) which cannot undergo VP-fronting, e.g., *give candy to*. See §2.5 below for further discussion of these additional restrictions.

Constituency

relations. The only conflicts should be between c-command tests and constituency tests, and among different constituency tests. Also, given the effects of Branch Right, we expect that c-command tests will predominantly diagnose right-branching structures.

- (19) *Prediction I*
Constituency changes during the course of a derivation, asymmetric c-command relations do not. Therefore, tests involving c-command relations should not conflict with one another.

The only exceptions to this generalization should be situations in which a less right-branching arrangement of a set of terminals is permitted because it receives a different interpretation from a more right-branching arrangement of the same set of terminals. These predictions about c-command diagnostics are tested in §2.4.

2.3.3 Prediction II: Left-edge constituency

The second prediction relates to the fact that although the constituents of a left-branching structure are also constituents during the derivation of a right-branching structure, these constituents are often destroyed once material is added on their right. The prediction is quite straightforward: once a constituent has been destroyed, it should be impossible to refer to it at any subsequent point in the derivation. Put another way, the only structural diagnostics that should be able to pick out the constituents of left-branching structures—which I shall refer to as *left-edge constituents*—are those diagnostics based on syntactic relations established *before* the constituency-destroying material is added on the right. For examples, diagnostics of the constituenthood of *the man saw* should involve syntactic relations established prior to the addition of the object NP *Mary* to the structure. Meanwhile, tests that diagnose right-branching structures should not be subject to the same restriction. This prediction is verified in §2.5.

- (20) *Prediction II*
Left-edge constituents are destroyed when material is added on their right. Therefore, evidence for left-edge constituents should be restricted to relations established before their constituenthood is destroyed by the addition of new material to their right.

2.3.4 Prediction III: Parallelism

The third prediction is an extension of the prediction that once a constituent has been destroyed it cannot be referred to later in the derivation. Consider what this means for constructions which require parallelism between two conjuncts. If we assume strict left-to-right structure building, this ensures that the first conjunct will be entirely built before the second is begun. Therefore, any intermediate properties of the first conjunct which might give rise to contradictory constituency effects will no longer be available when the second conjunct is being constructed. Parallelism requirements should therefore only be able to apply to the *final* properties of the first conjunct, and should not be able

Constituency

to access any properties of the first conjunct which were destroyed in the course of its derivation. A consequence of this is that contradictory constituency effects should be blocked in constructions requiring parallelism across two conjuncts.

- (21) *Prediction III*
- a. Parallelism requirements across two conjuncts should only be able to refer to properties of the *final* structure of the first conjunct.
 - b. Parallelism requirements between conjuncts should block contradictory constituency effects which would be possible in either of the conjuncts individually.

§2.6 argues that this prediction is correct, based on some differences in the distribution of contradictory constituency effects between movement and ellipsis constructions.

2.4 C-Command Tests

This section tests the first prediction, that different c-command tests should not contradict one another's results, and should diagnose right-branching structures, except where an alternative structure is forced by interpretive requirements.

- (22) *Prediction I*
- Constituency changes during the course of a derivation, c-command relations do not. Therefore, tests involving c-command relations should not conflict with one another.

2.4.1 Binding

(23–27) are familiar examples from the literature on double object and complex VP constructions (cf. Barss & Lasnik 1986) which show that c-command tests like anaphor binding, negative polarity item licensing and weak crossover all diagnose right-branching structures in double object and dative constructions, as we would expect. In all of the examples an element towards the left of VP behaves as if it c-commands an element on its right, and not vice versa.²⁴

- (23) *Reflexive Binding*
- a. I showed John himself in the mirror.
 - b. * I showed himself John in the mirror.
 - c. I showed the children_i to each other_i in the mirror.
 - d. * I showed each other_i to the children_i in the mirror.

²⁴ See Jackendoff 1990 for extensive documentation of the fact that the paradigms in (23–27) are also found with a wide range of other double complement constructions in English.

Constituency

- (24) *Bound Variable Anaphora*
- a. I denied each worker_i his_i paycheck.
 - b. * I denied it_is owner every paycheck_i.
 - c. I gave every paycheck_i to its_i owner.
 - d. * I gave his_i paycheck to every worker_i.
- (25) *Negative Polarity Item Licensing* (Klima 1964)
- a. I gave no one anything.
 - b. * I gave anyone nothing.
 - c. I gave nothing to anyone.
 - d. * I gave anything to nobody.
- (26) *Weak Crossover* (Postal 1971; Wasow 1972)
- a. Who_i did you show his_i reflection in the mirror?
 - b. * Which lion_i did you show it_is trainer?
- (27) *Superiority* (Chomsky 1973)
- a. Who did you give which book?
 - b. * Which book did you give who?

Therefore, these diagnostics provide promising initial support for the part of Prediction I which states that c-command tests should uniformly point to right-branching structures.

2.4.2 Scope

There is, however, one case of a c-command test which appears to contradict both parts of Prediction I. This test uses scope relations as a probe for c-command relations—wide scope is assumed to imply c-command—and the relevant cases involve the relative scope of sequences of postverbal adverbial modifiers. These phrases have been claimed to motivate a left-branching structure, based on the scope relations they exhibit (Andrews 1983, Ernst 1994, Pesetsky 1995), in violation of the prediction that c-command tests should diagnose right-branching structures except in cases of ambiguity.

The evidence comes from pairs of sentences like those in (28–30), in which the first adverbial and the rest of the VP is preferentially interpreted as taking narrow scope with respect to the second adverbial. Also, reversing the order of the modifiers reverses the scope relations. For example, (28a) is most naturally understood as meaning that the frequency of the hitting was purposeful, whereas (28b) is most naturally understood as meaning that what was purposeful was the hitting, but we don't know whether the frequency of the hitting was purposeful. Similarly, (30a) is most naturally understood as restricting concerto playing in foreign countries to weekends, whereas (30b) restricts concerto playing on weekends to foreign countries. Facts like this, then, are taken to motivate left-branching VP structures like (31).

- (28) a. Joe hit him frequently on purpose.
b. Joe hit him on purpose frequently.

(Ernst 1994)

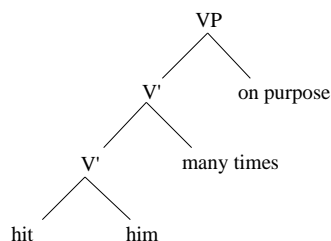
Constituency

- (29) a. She kissed him many times willingly.
 b. She kissed him willingly many times.

- (30) a. Kremer plays concertos in foreign countries on weekends.
 b. Kremer plays concertos on weekends in foreign countries.

(Pesetsky 1995)

(31)



If the argument for the structure in (31) based on the examples in (28–30) goes through, then Prediction I clearly faces a problem. I should emphasize that it is not the mere existence of a left-branching structure that poses a problem for the Merge Right/Branch Right system I am proposing: I assume that left-branching structures are tolerated where they are necessary. Nor is it problematic that evidence for a left-branching structure should come from a c-command test: I predicted that c-command tests should not conflict in their results, not that they should always diagnose right-branching structures. What *is* problematic is the claim, if true, that in sequences of postverbal modifiers the rightmost modifiers *must* take widest scope, and that therefore this *must* be represented as a left-branching structure. This is unexpected in the current system, first because there should be nothing to block phrases on the left taking widest scope, as in a right-branching structure; second, because deviations from right-branching structures are predicted to be possible only when it makes a difference to interpretation, precisely what cannot be the case if (28–30) are unambiguous.

In addition, the kinds of scope readings among adverbials which are used to argue for left-branching structures are available even when there is also a variable binding dependency between the adverbials of the kind that has been used to motivate *right*-branching structures (Ernst 1994, Phillips 1995), in apparent violation of the prediction that there should be no conflicts between the results of different c-command tests. The examples in (32–33) are based on the examples in (28–30), except that a left-to-right quantifier-variable dependency has been added. Adding the quantifier-variable dependency seems to make no difference to the relative scope of the two adverbials, which is the same as in (28–30).

- (32) a. I misled everyone_i on purpose the day before his_i briefing.
 b. She kissed everyone_i willingly on his_i cheek.

(Ernst 1994)

- (33) a. Kremer plays quartets in foreign countries_i on their_i national holidays.

Constituency

- b. Kremer plays quartets on new federal holidays_i in their_i first 5 years of existence.

(Phillips 1995)

At this point it seems that scope facts both contradict the results of other c-command tests and motivate a left-branching VP-structure. However, these facts do not pose problems for Prediction I, because the scope generalization breaks down under closer scrutiny. We must control for the fact that sentence final focal stress has an independent effect on interpretation, which makes it tend to be associated with widest scope. This can be controlled for by adding a third adverbial, as in the examples in (34). While ensuring that the third adverbial is receiving focal stress, we can ask whether the first two adverbials show the same scopal biases that they showed when they were the only two adverbials. My informants share the intuition that any forced scope nesting among the first two adverbials that might have been present in (28–30) goes away when an extra phrase is added that takes away the focal stress.

- (34) a. Sue kissed him willingly many times in front of the boss.
b. Kremer plays concertos in foreign countries on weekends at the height of the season.

In (34a) it is much easier than it was in (28b) to obtain a reading in which it is kissing many times that Sue did willingly (left-to-right scope), although the reading in which there were many individual willing kisses (right-to-left scope) is also still available. The loss of the requirement for right-to-left scope readings is even clearer in (34b). Recall that (30b) was most naturally taken to mean that *it is on weekends that Kremer plays concertos in foreign countries*. If this is the result of obligatory right-to-left scope then (34b) should be interpreted as *it is at the height of the season that it is on weekends that Kremer plays concertos in foreign countries*. This double restriction implies that when it is not the height of the season Kremer plays concertos in foreign countries at times other than on weekends. This reading is certainly not the required reading for (34b), and for many speakers it is not even available.

The fact that the scopal relations among adverbials are not fixed by their linear order, as the examples in (34) seem to indicate, is more consistent with the system proposed here. It suggests that the facts in (28–30) probably do not reflect obligatory right-to-left c-command among multiple adverbial phrases, but instead reflect some independent property of focal stress assignment. Furthermore, if the scope readings in (28–30) are not indicative of c-command relations, then the examples in (32–33) also should not be taken to show a contradiction between the results of two different c-command tests.

I should stress again that I am not trying to claim that scope relations among adverbials are never structurally represented. This will become evident when we consider the interaction of adverb scope with VP-ellipsis in §2.6 below. What I am challenging is the claim that the scope readings in sequences of adverbials motivates *obligatory* right-to-left scope and hence *obligatory* right-to-left c-command relations.

Constituency

Therefore the first prediction holds up: that c-command tests should not conflict, and should uniformly diagnose right-branching structures except where forced by interpretation.

2.5 Linear Order and Left-edge Constituency

Prediction II from §2.3 is repeated below as (35).

- (35) *Prediction II*
Left-edge constituents are destroyed when material is added on their right. Therefore, evidence for left-edge constituents should be restricted to relations established before their constituenthood is destroyed by the addition of new material to their right.

Prediction II points out a key prediction of the Merge Right approach to constituency. If apparent contradictions between constituency tests are a reflection of the stages of left-to-right derivations, in which certain constituents are destroyed when other constituents are created, then we expect some constituents to be available to syntactic processes for only part of a derivation.

This section focuses on one aspect of this prediction. I show that when there is evidence for both left- *and* right-edge constituency in a given sentence, those syntactic relations which motivate the existence of left-edge constituents are always established *before* the addition of the material that motivates the existence of right-edge constituency. The constructions discussed here involve leftward and rightward movement and Right Node Raising.

2.5.1 VP-Fronting

VP-fronting constructions appear to support the existence of a left-branching structure for VP, because strings starting at the left-edge of VP can be fronted, stranding material on the right-hand side of VP. The relevant examples were already presented in §2.2, and are repeated below as (36).

- (36) a. ... and [give candy to children in libraries on weekends] he did.
b. ... and [give candy to children in libraries] he did on weekends.
c. ... and [give candy to children] he did in libraries on weekends.
d. ... and [give candy] he did to children in libraries on weekends.
e. * ... and [to children in libraries] he did give candy on weekends.
f. * ... and [in libraries on weekends] he did give candy to children.

The examples in (37) are similar to the examples in (36), except that they contain a reciprocal binding relation between a pronoun in the fronted portion of VP and a reciprocal in the portion of VP that is stranded. This kind of binding relation is what we would expect to find if the fronted portion of VP were in its *unfronted* position, and if the entire VP were right-branching. However, the highlighted pronouns in (37a–b) do not c-command the reciprocal from their fronted position, nor would they even c-command the reciprocal if they were in their unfronted position in a left-branching VP structure. In other words,

Constituency

assuming ‘reconstruction’ of the fronted phrases into a left-branching VP is insufficient to account for the binding facts.

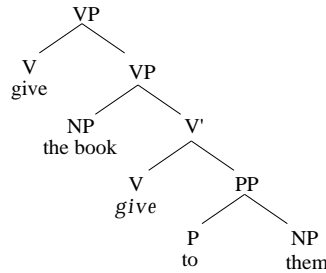
- (37) a. ...and [give the book to **them** in the garden] he did ___ on **each other’s** birthdays.
 b. ...and [give the book to **them**] he did ___ in the garden on **each other’s** birthdays.

(Pesetsky 1995, p.230)

The fact that the material at the left-edge of VP can form a constituent to the exclusion of material at the right of VP would normally provide a straightforward argument for a left-branching VP structure. In combination with the binding evidence motivating a right-branching structure in (37), then, the VP-fronting facts appear to contradict one another. However, since in a left-to-right derivation the movement chain is completed before the anaphor is added, the system proposed here allows for the fronted portion of VP to be an incomplete right-branching VP-structure, rather than a piece of a left-branching VP structure.

(38) shows how the facts in (36–37) are expected in a left-to-right derivation of a right-branching structure. In (38a) the fronted portion of VP is first built, in its fronted position. This partial VP is internally right-branching, and is the result of a derivation like (38a–e) above.

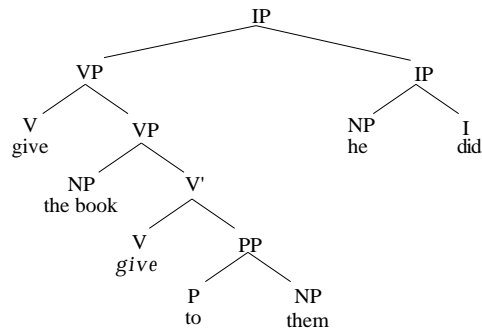
- (38) a. ... and [give [the book [to them]]]



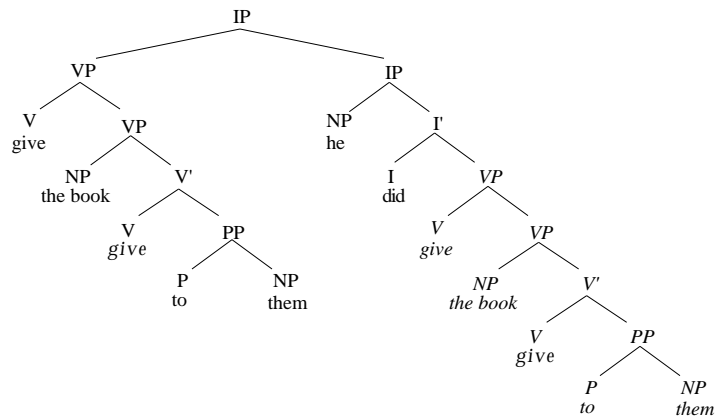
Then in (38b) and (38c) the subject, and *do* are added to the structure, and then a copy of the fronted VP is inserted as the complement of Infl. I assume that the movement chain is licensed at this point in the derivation, and not later.

Constituency

(38) b. ... and [give [the book [to them]]] he did



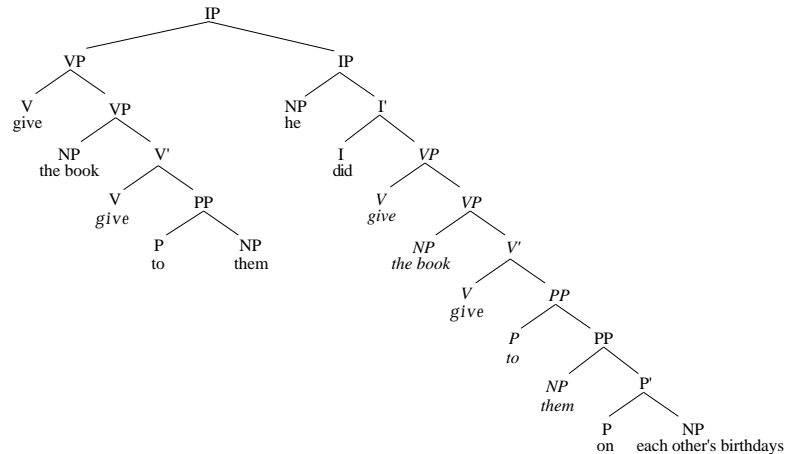
c. ... and [give [the book [to them]]] he did [give [the book [to them]]]



Subsequently, in (38d) extra material is added to the right of VP, inserting the temporal modifier *on each others birthdays*. This adverbial is inserted at the bottom of a right-branching VP, which (i) creates the c-command relation necessary to license the reciprocal binding relationship, and (ii) destroys the constituency of the string *give the book to them*. The loss of this constituent does not matter, though, because the movement chain involving this constituent had already been established and licensed before the modifier was added.

Constituency

- (38) d. ... and [give [the book [to them]]] he did [give [the book [to [them
[on each other's birthdays]]]]]



In this way the apparent constituency conflict can be captured in the derivation of a right-branching structure. [See §2.6 for discussion of a contrast between the VP-fronting facts presented here and related facts involving VP-ellipsis.]

It is important to note that this way of achieving the effects of contradictory constituency from a single derivation depends on the left-to-right ordering of the derivations that I am assuming here. To see this, imagine what would happen if we tried to capture the same effects in a bottom-to-top derivation. In this kind of a derivation, the entire right-branching VP would be built before the VP-fronting operation could apply. But this would mean that by the time the VP-fronting operation could apply, the portion of VP that is fronted in (38) would no longer be a constituent, and therefore could not be a candidate for movement.²⁵

Alternatively, it might be objected that the constituency conflict shown by (36–37) is only apparent, because the binding relations are entirely consistent with a left-branching VP-structure, and that the problem is an artifact of assuming that the binding relations motivate a right-branching structure. This analysis could account for the facts in this subsection, but arguments to be presented in §2.6–7 involving contrasts between VP-fronting and VP-ellipsis show that this analysis leads to the loss of important generalizations about the distribution of constituency conflicts.

While the account given here makes it possible to account for the apparent conflict between binding and movement diagnostics, there is an additional restriction on VP-fronting which does not follow automatically from the phrase

²⁵ If the derivation was *strictly* bottom-to-top then the portion of VP that is fronted in (38) would not in fact ever be a constituent. Only in a less strictly bottom-to-top derivation would it ever be a constituent. Nevertheless, in any theory in which construction of the entire VP precedes VP-fronting, the kind of fronting shown in (38) is predicted to be impossible.

Constituency

structure theory presented here. VP-fronting does not allow the two complements of a double object construction to be separated by movement (39), and neither can an argument or adverbial phrase be split up by movement (40).

(39) * ...and [give the children] he did candy in libraries on weekends.

(40) a. * ...and [give candy to] he did the children in libraries on weekends.

b. * ...and [give candy to children in] he did libraries on weekends.

In Pesetsky's left-branching Layered structures this restriction follows immediately, because the bracketed strings in (39–40) are not constituents. Under the Merge Right approach, on the other hand, in which partial VP-fronting is just fronting of intermediate stages in the construction of right-branching VP structures, the bracketed strings in (39–40) *are* constituents, and so we might expect them to be allowed to front.

The additional requirement seems to be that the fronted portion of VP be a *potential complete VP*. None of the fronted strings in (39–40) are possible as complete VPs. I assume that the requirement that a potential complete VP fronts is a construction-specific semantic requirement, and that the restriction does not undermine the claim that the bracketed strings in (39–40) are potential constituents. As we will see later in this section, Right Node Raising allows coordination of many of the constituents that cannot undergo VP-fronting, because it is not subject to the same semantic requirement.

Note that the restriction cannot be that only adverbial phrases can be stranded by VP-fronting. The two complements of the verb in a dative construction may sometimes be separated by VP-fronting. When the goal argument is optional, as with the verb *give*, the verb and the theme may be fronted (41a); when the goal argument is obligatory, as with the verb *hand*, the verb and the theme cannot be fronted (41b).

(41) a. (?) ...and [give candy] he did to the children in libraries on weekends.

b. * ...and [hand candy] he did to the children in libraries on weekends.

The contrast between (41a–b) is consistent with the generalization that only potential complete VPs may be fronted, and shows that this notion must be relativized to individual lexical items.²⁶

I leave the reason for the semantic restriction on VP-fronting as an open question for the time being. However, I note one reason why the restriction is not surprising. The fronted portion of VP in VP-fronting constructions is the *entire* VP in the first conjunct of these constructions. The initial conjunct for sentences like (39–40) would have to be like (42), which are clearly impossible.

²⁶ The contrast between (41a) and (41b) also suggests that the semantic account of restrictions on VPF fares better than Pesetsky's account in terms of constituents in Layered structures. On the assumption that dative constructions involving *give* and *hand* have an identical Layered phrase structure, the contrast in (41) is not expected.

Constituency

- (42) a. * John intended to give the children, and ...
b. * John intended to give candy to, and ...
c. * John intended to give candy to children in, and ...

Next we consider constructions which constituency conflicts in only a limited range of situations, in a manner predicted by their ordering properties.

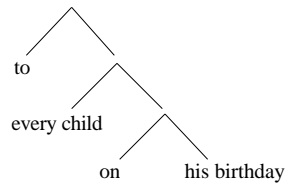
2.5.2 PP-movement

The interaction of binding and movement processes involving PPs shows a constituency conflict similar to what we have seen with VP-fronting in (36–38), but with an additional twist which enables us to test the predictions of Merge Right more closely.

As we have already seen above, the binding and coordination properties of noun phrases inside PPs motivate right-branching structures in which the NP is not the sister of the preposition that selects it, and instead forms a constituent with the category that follows it. (43) gives some examples of the kinds of binding phenomena which have led to this conclusion, and (44) shows the ‘split’ PP structure that these motivate.

- (43) a. Mrs. McGarrick sent a card to **every child_i** on **his_i** birthday.
b. The urban-hygiene inspectors departed from **every city_i** during **its_i** rush hour.
c. The chef told the guests about **every dish_i** as **it_i** was served.
d. Mrs. McGarrick gave a card to **none** of the children on **any** of their birthdays.
e. Mrs. Murray gave money to **her children** on **each other’s** birthdays.

(44)



A P-NP combination that has been split up in the manner shown in (44) is not a constituent, and therefore should not be able to undergo movement. Clearly, though, leftward movement of PPs presents no problems, as the examples in (45) indicate. This implies that the P-NP combination is a constituent after all.

- (45) a. To each of the girls John gave a package ___ wrapped in brown paper.
b. To which city in Connecticut did Mary take the train ___ every day of the week?

Constituency

Moreover, the kind of binding out of a PP which motivated the PP-splitting structures is still possible when the PP containing the binder is fronted, as the examples in (46–47) show. The examples in (47a–b) are taken from Pesetsky 1995 (p.228).²⁷

- (46) a. To each of the girls John gave money for her college fees.
b. To which pair of boys did John accidentally give money on each other's birthdays?
- (47) a. To none of the officials did Sue send her money ___ on any of these days.
b. On which table did Tom put the book ___ during its construction?

The interaction of movement and binding with PPs thus gives rise to a constituency contradiction similar to the one we saw with VP-fronting, since the movement properties support a structure in which the PP is a constituent, whereas the binding properties support a structure in which the PP is not a constituent.

The contradiction can be accounted for in exactly the same way that the VP-fronting facts were explained, because both links of the movement chain are built prior to the addition of the adverbial phrase that creates the c-command relation required for binding and destroys the constituency of the PP. The relevant steps of the derivation of sentence (46a) are given in (48).

First the fronted PP is built sentence initially (48a). At this point the PP *to each of the girls* is a constituent. Next the material intervening between the head and the tail of the PP-movement chain is added to the structure (48b), and then a copy of the fronted PP is inserted at the appropriate position in VP for the goal argument of *give* (48c). At this point both ends of the PP-movement chain are constituents. It is only when the additional phrase *for her college fees* is added on the right that the PP *to each of the girls* is split, such that the NP *each of the girls* forms a constituent with the following PP and is able to license the bound variable pronoun *her* (48d).

²⁷ Note that not all instances of PP-fronting allow binding out of the fronted PP as (46–47) do. As van Riemsdijk & Williams 1986 observe (cf. also Baltin & Postal 1996), prepositions that can be stranded by pseudopassivization act as blockers to binding when the entire PP is fronted, as the examples in (i–iii) show.

- (i) Louise was talked to.
(ii) Which girl did Ernest talk to ___ about herself?
(iii) * To which girl did Ernest talk ___ about herself?

The extraction examples improve in appositive relatives, e.g. *?Mary, to whom I talked about herself, was in a good mood*, but I have no account of why this should make a difference.

Because of the effects of pseudopassivizable verbs, the current discussion focusses on PPs whose head cannot be stranded by pseudopassivization. I leave the question of why pseudopassivizability affects binding possibilities as an open question at this point.

Constituency

- (48) a. [to [each of the girls]]
b. [to [each of the girls]] John gave money
c. [to [each of the girls]] John gave money [to [each of the girls]]
d. [to [each of the girls]] John gave money [to [[each of the girls] for her college fees]]

Thus far the PP movement facts are entirely parallel to the VP-fronting facts in (36–38). In both cases we have observed what under standard assumptions would be a straightforward constituency contradiction, and shown that the Merge Right approach makes it possible to account for such facts in terms of how constituency changes over the course of a derivation. Until now, though, we have not directly tested the prediction that once a constituent is destroyed it cannot be referred to again later in the derivation. This prediction can be tested with PPs, since PPs can be moved both leftwards and rightwards.

If the Merge Right approach to contradictory constituency involving PPs is correct, then we should expect to find differences between leftward and rightward movement of PPs with respect to how they interact with binding. I assume here that rightward movement is identical to leftward movement insofar as it involves a series of copies of a given phrase, just one of which is overtly realized. The only difference between leftward and rightward movement, therefore, will be in whether the overt copy is on the left or on the right of the unpronounced copies. I assume in addition that heavy shift operations involve a lowering operation which copies a phrase in its base position inside VP to a position lower in a right-branching VP structure.²⁸

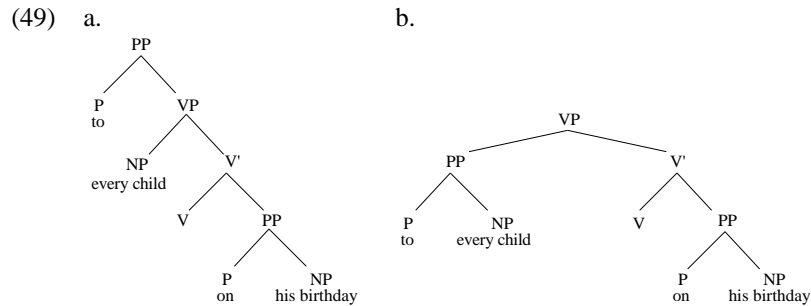
Now consider the structure in (44), repeated below with category labels as (49a).

²⁸ The claim that heavy shift involves a lowering operation diverges from a body of literature which assumes that heavy shift involves upward movement (cf. Pesetsky 1995 and references cited therein). One of the main pieces of evidence in support of this view is the claim that heavy shift licenses parasitic gaps, based on examples like (i) (Engdahl 1983: observation attributed to Tom Wasow). Parasitic gaps are standardly taken to need to be *at least* c-commanded by the head of a well-formed *wh*-chain.

(i) Sue offended ____t by not recognizing ____{pg} immediately her favorite uncle from Cleveland.

See Appendix 2 of this chapter and Postal 1994 for arguments that this construction does not involve a parasitic gap and is more similar to right node raising constructions.

Constituency



As we have already seen, structure (49a) is consistent with *leftward* movement of the PP *to every child*, because PP-splitting occurs only after the movement chain has been completed. On the other hand the structure in (49a) should be incompatible with *rightward* movement (i.e. heavy shift) of the PP *to every child* across the PP *on his birthday*. This is because a left-to-right derivation does not allow the rightward movement chain to be completed before the addition of the following PP, which would normally be the trigger for PP-splitting. If, on the other hand, the PP fails to undergo PP-splitting and remains a constituent when a subsequent PP enters the derivation, yielding the slightly less right-branching VP-structure in (49b), then the PP should be fully capable of participating in a rightward movement chain. The price of failing to undergo PP-splitting, though, is that the NP *every child* should no longer be able to act as a binder, because it cannot c-command out of PP. This prediction appears to be correct, as the following examples show.

First we need to show that rightward movement does in principle allow ‘reconstruction’ effects for the purposes of binding. (50) demonstrates this for Heavy NP Shift using an example from Baltin & Postal 1996.

- (50) a. I described [the victim whose sight had been impaired by the explosion] to himself.
 b. I described ___ to himself [the victim whose sight had been impaired by the explosion].

(51) shows that Heavy PP Shift is a possible operation. The crucial examples in (52) and (53) show that when a PP that allows binding when *in-situ* (52a, 53a) undergoes Heavy PP Shift, the binding is no longer possible (52b, 53b).^{29,30}

²⁹ Some speakers have little difficulty in accepting (52b), but (53b) is more uniformly rejected across speakers.

³⁰ In (51–53) I deliberately focus on the binding properties under heavy shift of PPs that are *not* the first object of the verb. I focus on these PPs because of the fact immediately postverbal PPs lose their binding properties under *both* leftward and rightward movement, as the combined results of Reinhart 1981, van Riemsdijk & Williams 1986 and Postal 1986 show. For the purposes of testing the predictions of the Merge Right theory, though, we are only interested in the behavior under rightward movement of those PPs that *can* act as binders when they undergo leftward movement.

Constituency

- (51) I gave money in an envelope to every boy who had helped me clean the yard.
- (52) a. I gave money to the boys for themselves.
b. * I gave money ___ for themselves to the boys who had helped me clean the yard.
- (53) a. I gave money to every boy on his birthday.
b. * I gave money ___ on his birthday to every boy who had helped me clean the yard.

The contrast between the possibility of reconstruction and binding when a PP is moved leftwards and the impossibility of reconstruction and binding when the same PP is moved rightwards is a straightforward consequence of a theory which assumes left-to-right derivations and splitting of PPs. As far as I can tell the contrast is not expected under accounts in which structure is built from bottom-to-top or in which there are no derivations.

2.5.3 Right Node Raising

Right Node Raising gives rise to constituency puzzles similar to the ones discussed in §2.5.1 and §2.5.2, in that different properties of a single sentence appear to provide evidence for two different structural analyses of that sentence. But Right Node Raising provides the most extreme case yet. Whereas in §2.5.1 and §2.5.2 we were concerned with conflicting structural analyses for PPs or VPs, Right Node Raising creates conflicts in the analysis of entire sentences.

The classic form of Right Node Raising (RNR) involves coordination of subject-verb sequences, with the remaining clausal material effectively ‘shared’ between the two conjuncts, as in (54).³¹

³¹ It is tempting to assume that Right Node Raising is an exotic and stylistically marked quirk of English. However, the briefest of surveys shows this common conception to be false. The examples in (i–v) below represent a sample from my casual field work over a two week period.

- (i) The distance from the top to the bottom of the precipice is about 500 feet.
- (ii) “Stone also suggests that Nixon knew of, though he did not attempt to participate in, US attempts to assassinate Fidel Castro.”
(*Boston Sunday Globe* movie section, 12–17–95)
- (iii) “Textbook-classic homuncular maps can predict, but not guarantee structure function relationships even in normal subjects.”
(radiology journal article)
- (iv) “We’re sorry for the delay in this flight. The machine that rips the handles off of and punches holes in your baggage was broken, so it all had to be done manually.”
(in flight announcement, *United Airlines*)
- (v) “Receive \$5 off a child’s or receive \$10 off an adult lift ticket (when you present two special ‘Ski & Save’ side panels from Hood Milk).”
(milk carton, Boston, MA)

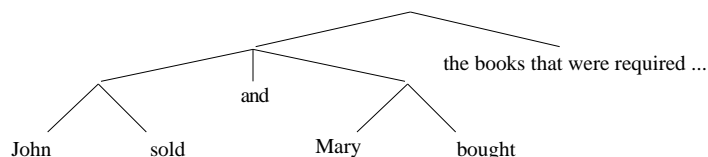
Constituency

- (54) [John sold] and [Mary bought] the stack of books that was required for linguistics 101.

If we adopt the logic standardly applied to coordination, that strings that can be coordinated are constituents, then the fact that the strings *John sold* and *Mary bought* can be coordinated in (54) provides evidence that the subject and the verb can form a constituent to the exclusion of the object.

The primary aim of this section is to show that RNR motivates the existence of non-standard constituents like [subject verb], which most accounts of RNR have attempted to deny. The secondary aim of the section is to show that the existence of constituents like [subject verb] does not entail the existence of structures like (55) in which the subject fails to c-command the object.

- (55)



The structure in (55) predicts that the subject should not be able to bind an object in RNR examples similar to (54). (56) shows that this prediction is clearly false. The shared object in an RNR sentence can be bound by the subject.³² Further evidence against the structure in (55) is given below.

- (56) a. John sold and Mary bought each other's textbooks.
b. Everyone_i suspected but nobody_i really believed that he_i was being investigated by the FBI.

I will show that the apparent conflict between the constituency motivated by the coordination and the binding in (56) can be resolved in a left-to-right approach to structure building.³³

2.5.3.1 Disguised Clausal Coordination

Since almost all phrase structure analyses have assumed that the subject and the verb in English uncontroversially *do not* form a constituent to the exclusion of the object, RNR has typically been analyzed as one form or another of

In addition, RNR is not particular to English. It is found in a wide variety of languages, including German, Georgian, Dutch, Polish, Russian, Spanish, French, Irish, Japanese, Basque, Hindi. At the time of writing I am unaware of any language that allows coordination but does not allow RNR.

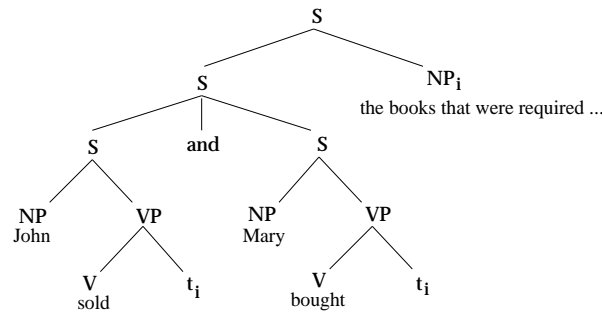
³² (56a) will be ruled out independent of RNR for speakers who do not accept split antecedents for reciprocals.

³³ See Appendix 1 of this chapter for more detailed presentation of the basic properties of Right Node Raising.

Constituency

disguised coordination of clauses. There have been two basic approaches to treating RNR as clausal coordination. The first approach, illustrated in (57), is to assume that RNR involves clausal coordination followed by *across-the-board* (ATB) rightward extraction of the shared material (cf. Ross 1967/1986, Maling 1972, Postal 1974 and many others). In other words, the shared material is part of both conjuncts, but it is not *in-situ* in either conjunct.

(57)



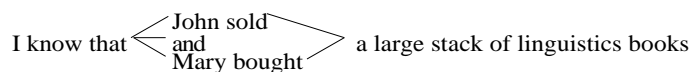
The second kind of clausal coordination approach to RNR encompasses a number of theories which modify standard phrase structure theories in such a way that the shared material in RNR can be *in-situ* and shared between both conjuncts without ATB extraction. Versions of this approach have been advanced by Williams 1978 and Ertshik-Shir 1987 under the heading of ‘clausal factorization’, by Goodall 1987 in terms of ‘phrase marker union’, and by Muadz 1991 and Moltmann 1992 under the heading of ‘three-dimensional phrase markers.’

What these approaches have in common is that they assume that RNR is the result of the superimposing of two partially identical sentences or *factors* upon one another. Where the two sentences are identical, there is just one representation for both occurrences. Only where the factors differ do the representations of the two factors diverge, as (58) shows. This separation of the two sentences is marked by a conjunction such as *and*, which is quite crucially not a part of either of the independent factors.³⁴

³⁴ McCawley (1982) has proposed a related account of RNR in terms of discontinuous constituents. McCawley’s theory is similar to factorization approaches insofar as it allows a phrase to be *in-situ* in two conjuncts simultaneously. However, McCawley’s theory is not committed to the idea that RNR derives from the superposition of two independent sentences.

Constituency

- (58) a. I know that John sold a large stack of linguistics books.
and
I know that Mary bought a large stack of linguistics books.
- b.



Both of the clausal coordination approaches to RNR manage to avoid positing non-standard constituents (e.g. subject-verb) by assuming that the shared material is somehow a part of each conjunct, either *in-situ* or extracted. However, I will argue that neither of these approaches can rescue a clausal coordination analysis of RNR (regardless of their other merits). Therefore, if it can be shown that the shared material in RNR sentences cannot have been moved out of both of the conjuncts, and cannot be *in-situ* in both conjuncts, then we must conclude that the shared material in RNR is not a part of both conjuncts, and therefore RNR must involve the coordination of units smaller than a clause. If this is the case, then the characterization of the puzzle in (54–56) stands.

2.5.3.2 Right Node Non-Raising

There are a number of arguments in the literature against the ATB extraction analysis of RNR. The logic of these arguments is typically to show that the shared material behaves as if it has not undergone movement based on some diagnostic or other. This could involve either evidence that the shared material in RNR fails to induce movement violations in situations in which the ATB extraction analysis would predict a movement violation, or evidence that binding relations are possible which are unexpected if the shared material has been displaced.

(59) shows that RNR does not induce *wh*-island violations (Wexler & Culicover 1980). (59a–b) shows that leftward movement across *who* leads to ungrammaticality; (59c) shows that no such violations are incurred in RNR, suggesting that movement has not occurred.

- (59) a. * What does Mary know a man who buys and Bill know a man who sells?
b. * It is pictures of Fred that Mary knows a man who buys and Bill knows a man who sells.
c. Mary knows a man who buys and Bill knows a man who sells pictures of Fred.

(60) shows that in languages in which preposition stranding is strongly ungrammatical the complement of a preposition can be shared in RNR (McCloskey 1986), leaving the preposition stranded at the right-hand edge of each conjunct. This suggests that extraction from PP has not occurred. Example (60) is taken from Irish, but identical arguments can be made with Spanish, French or Polish, as McCloskey shows.

Constituency

- (60) Níl sé in aghaidh an dlí a thuilleadh a bheith ag éisteacht le
is-not it against the law anymore be listen(prog)
nó ag breathnú ar ráidió agus teilifís an Iarthair.
with or look(prog) on radio and television the West(gen)
'It is no longer against the law to listen to, or to watch—Western
radio and television.'

Next, although the simplest cases of RNR involve the coordination of subject-verb sequences and the sharing of a direct object, the examples in (61) show that more than just subject-verb sequences can be coordinated and more than just direct objects can be shared in RNR constructions.

- (61) a. [John will] and [Mary already has] mailed the conference program to all of the presenters.
b. [John will post] and [Mary is about to e-mail] a copy of the conference program to all of the presenters.
c. [John will mail the abstracts] and [Mary is about to e-mail the program] to anybody who registered in advance.

The relevance of the examples in (61) to ATB accounts of RNR is that they show that a wide range of different categories can serve as the shared material, including categories for which there is no independent evidence that they can undergo movement. For example, neither the VP headed by a participial in (61a) nor the two objects of the double complement construction in (61b) can undergo leftward movement in English, as the examples in (62) show. Nor can they undergo rightward movement, as (63) shows.

- (62) a. * (and) [mailed the conference program to all of the presenters] Mary already has.
b. * [A copy of the conference program to all of the presenters] Mary is about to e-mail.
(63) a. * Mary already has ___ from her local post office [mailed the conference program to all of the presenters].
b. * Mary is about to email ___ from her company account [a copy of the conference program to all of the presenters].

I do not claim to have an explanation of why the movements shown in (62) are impossible. The relevance of the examples in (62–63) is just that since we know that the shared phrases in (61a–b) cannot be moved leftwards or rightwards, it would be surprising if these phrases are allowed to move only when the movement is string vacuous. But this assumption would be the only way of accommodating (61a–b) under an ATB analysis of RNR.

As a further argument against the ATB account of RNR, (64) shows that the shared material in RNR behaves as if it is *in-situ* for the purposes of a variety of tests of binding and coreference (Levine 1985). The subject can bind a variable inside the object in (64a), it can license a negative polarity item (64b), and it

Constituency

induces a Condition C violation in (64c). These facts again suggest that movement has not taken place.³⁵

- (64) a. [Everyone_i liked] and [at least one person_i loved] the paper he_j had been asked to review.
 b. [Nobody enjoyed] and [few people even liked] any of the talks on Right Node Raising.
 c. * [I know that she_i said] and [I think we all agree] that Mary_i needs a new car.

Finally, if RNR does not involve movement, then the shared constituent should always fill the final position of the coordinated constituents (cf. Oehrle 1991, Wilder 1994). If, on the other hand, RNR involves ATB extraction (Williams 1990; Postal 1994), then it ought to be possible to share a phrase that has been extracted from the middle of both conjuncts.

Distinguishing between these alternatives requires some care, because RNR may interact with heavy NP shift in such a way that it appears that the shared material has been extracted. For example, we could derive (65) either by directly moving the clause final NP out of each of the underlined gaps, or by first applying HNPS in each conjunct, and then sharing the final NP without movement, as in the derivation sketched in (67).

- (65) [Patty sent ___ to Greenland] and [Susie sent ___ to her rich Uncle Ben] a list of all the things she wanted for Christmas.
- (66) a. [Patty sent ___ to Greenland ___] and [Susie sent ___ to her rich Uncle Ben ___] a list of all the things she wanted for Christmas.
- (67) a. [... V NP PP] and [... V NP PP] basic order
 b. [... V ___ PP NP] and [... V ___ PP NP] heavy NP shift
 c. [... V ___ PP] and [... V ___ PP] NP right node raising

We can test for whether (65) is the result of ATB extraction from the middle of each conjunct or the result of heavy shift feeding RNR by constructing

³⁵ It might be objected that the facts shown in (64) are consistent with an ATB movement analysis of RNR, because the binding effects may be attributed to LF reconstruction, i.e. the ATB movement is ‘undone’ at LF. While this may be a viable approach to the variable binding in (64a) and the Condition C effect in (64c), given the existence of reconstruction effects with leftward movement shown in (i), it would not cover the polarity item licensing in (64b), since polarity item licensing seems not to show reconstruction effects (ii).

- (i) a. ? His mother, everyone likes.
 cf. Everyone likes his mother.
 b. * Which picture of Mary_i did she_i like?
 cf. * She liked a picture of Mary.
- (ii) * Whose theory about anything does John not like?
 cf. John doesn’t like Bill’s theory about anything.

Constituency

examples in which the shared material cannot undergo heavy NP shift. Once we do this, as (68–70) show, RNR becomes impossible.

(68) shows that stranding prepositions in RNR, where the stranded preposition is the *final* word of the first conjunct, is acceptable. (69) shows that P-stranding does however cause problems for heavy NP shift. (70) is like (68), except that it contains the impossible HNPS environment from (69).

- (68) Patty wrote to and Susie sent email to the person she hoped would bring her wonderful Christmas gifts.
- (69) * Patty wrote to after breakfast the person she hoped would bring her wonderful Christmas gifts.
- (70) * Patty wrote to after breakfast and Susie sent email to just before lunch the person she hoped would bring her wonderful Christmas gifts.

The fact that (70) is also bad therefore implies that RNR cannot share material from the middle of the conjuncts. Therefore, the impression that this *is* possible that we might draw from (65–66) is just due to the fact that heavy shift feeds RNR.

For reasons like these, it has generally been concluded that the ATB movement approach to Right Node Raising is not viable. But this does not necessarily entail that Right Node Raising cannot be clausal coordination, because all of the facts in (59–70) are consistent with the clausal factorization approach to RNR. This is because the shared material is *in-situ* in both conjuncts in clausal factorization theories.

2.5.3.3 Factorization and Ordering

In what follows I do not try to argue against three-dimensional or factorization approaches to coordination in general. In fact, I think that there are a number of good reasons to adopt such an approach. My criticism is targeted specifically at the use of these approaches to give a clausal coordination analysis of RNR, and thereby avoid the need to posit non-standard constituents like [subject verb].

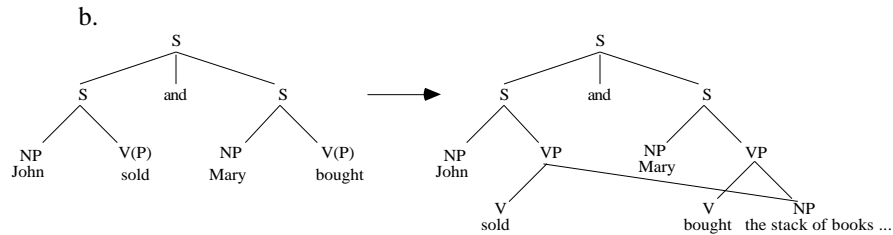
I assume that the final representations of RNR sentences involve *in-situ* shared phrases, as is the case in factorization theories, but I assume that coordination occurs at an earlier point in the left-to-right derivation, when the shared material has not yet been added. The example of RNR in (71a) is derived by first building a subject-verb constituent, at which point it can be coordinated with another subject-verb constituent.³⁶ I assume that this conjunction receives

³⁶ I assume that in order for coordination to take place the following two conditions must be satisfied: (i) each conjunct must be a constituent, (ii) any material that is semantically combined with one conjunct must combine identically with the other conjunct. Requirement (ii) rules out certain classes of sentences which are impossible but which satisfy requirement (i), such as: *John read and then Bill arrived a book* with the intended meaning ‘John read a book and then Bill arrived.’

Constituency

the kind of interpretation that Moltmann 1992 proposes for parallel structures in her theory of three-dimensional structures. It is only after this coordination has been licensed that the shared object is added on the right, destroying the constituency that had licensed the coordination, and creating a configuration in which the subject c-commands object.

- (71) a. John sold and Mary bought the stack of books that were required for linguistics 101.



Therefore, RNR only gives the appearance of coordinating the pieces of left-branching structures because the coordinated phrases are constructed prior to the addition of the shared material.

Given that the final representation in (71b) looks very much like what is assumed in factorization approaches, a natural question to ask is why there is any need to assume that non-standard subject-verb constituents are coordinated. The argument for this comes from some facts involving the relative ordering of the conjuncts and the shared material.

Any account of RNR must explain the impossibility of examples like (72), in which the shared material occurs at the end of the first rather than the second conjunct.

- (72) * John saw Mary and Bill likes.

According to the left-to-right theory, (72) is impossible because by the time in the derivation when the complete sentence *John saw Mary* has been built, the string *John saw* is no longer a constituent, and so there is no longer a subject-verb constituent available to coordinate with *Bill likes*.

In other approaches to RNR it is also rather straightforward to account for the ill-formedness of (72), by invoking some additional ordering requirement. This additional mechanism either deletes the copy of the shared material in the first conjunct (Wexler & Culicover 1980, Kayne 1994), or aligns the phrase that is *in-situ* in both conjuncts with the right-hand edge of the second conjunct (McCawley 1982, Moltmann 1992). However, we can show that such additional mechanisms fall short when faced with some additional ordering facts.

Thus far in this section we have only considered examples of RNR in which the two conjuncts are connected by a standard coordinator, such as *and*. However, it is possible to construct examples of 'non-coordinate' RNR, as pointed out by Richard Hudson in a 1976 paper and largely overlooked in most treatments of RNR since then. (73) repeats some of Hudson's examples. The examples in (74) are from Postal 1994: the highlighted strings take the role that

Constituency

coordinators take in simpler cases of RNR above. As the examples show, it is even possible (73a) for a subject and an object to behave as the two ‘conjuncts’ in this variety of RNR.

- (73) a. Of the people questioned, [those who liked] **outnumbered by two to one** [those who disliked] the way in which the devaluation of the pound had been handled.
b. I’d have said he was sitting [on the edge of] **rather than** [in the middle of] the puddle.
c. It’s interesting to compare [the people who like] **with** [the people who dislike] the power of the big unions.
- (74) a. [Politicians who have fought for] **may well snub** [those who have fought against] chimpanzee rights.
b. [People who are learning to speak (in)] **may hate** [those who already can speak (in)] that little-known language.
c. [People who believe there may soon be on Venus] **tend to distrust** [those who believe there already are on Mars] extraterrestrials capable of understanding parasitic gaps.
d. [Spies who learn *when*] **can be more valuable than** [those able to learn *where*] major troop movements are going to occur.

The examples in (75) extend Hudson’s examples and show that in non-coordinate RNR, where the Coordinate Structure Constraint presumably does not apply, one of the conjuncts can undergo movement, independently of the other. (75b–c) shows raising of the first conjunct, (75d) shows passivization, (75e) shows an instance of possible unaccusative raising, and (75f) shows *wh*-movement.

- (75) a. The people who liked **easily outnumbered** the people who disliked the movie.
b. The people who liked **must ___ have easily outnumbered** the people who disliked the movie.
c. The people who liked **seemed ___ to have easily outnumbered** the people who disliked the movie.
d. The people who like **are easily outnumbered ___ by** the people who dislike the movie.
e. The people who liked **arrived ___ much earlier than** the people who disliked the movie.
f. Which voter group that liked ___ **outnumbered** which voter group that disliked the info-mercial?

Now consider the examples in (76–77). In each of the examples the shared material appears to the right of both conjuncts. But whereas in the (a) examples the shared material also appears to the right of the underlying position of both

Constituency

formed instances of RNR. A theory which derived ordering restrictions on RNR from a surface linearization filter would therefore fail to exclude (76b) and (77b).

Meanwhile, a rule which simply required the underlying position of the shared material to be to the left of both conjuncts would fail to exclude situations like (77c), in which the surface position of the shared material precedes the second conjunct, but its underlying position does not. Such cases are clearly bad, as (79) shows.

- (79) a. * The people who liked the movie about Reagan's childhood seemed to the people who disliked ___ to be complete fools.
b. * Which voter group that disliked the info-mercial did which voter group that liked outnumber ___?

The examples in (79) are ruled out in the current approach to RNR because the combination of the first conjunct with the shared material prior to the building of the second conjunct rules out the possibility of coordinating the two conjuncts.

What I hope to have shown with this argument is that the conjuncts in RNR have properties before the shared material is added which they do not have after the shared material is added, e.g., they can move. This distinction is straightforwardly expressed in a left-to-right approach in which RNR involves coordination of *non-final* constituents, but it is not easily captured in a more standard version of factorization theories in which RNR involves coordination of units which include the shared material.

2.5.4 Summary

Briefly summarizing the results of this section. I have provided evidence for two aspects of the left-to-right approach to structure building. First, in both VP-fronting and Right Node Raising constructions I gave evidence for the participation in grammatical processes of the pieces of incomplete phrase markers.

Second, I have shown evidence for an account of constituency conflicts which attributes conflicting results of different constituency diagnostics to the different stages of a left-to-right derivation. We have observed a series of constructions in which evidence for right-branching structures appears to coexist with evidence for non-right-branching structures. In each case, though, the syntactic relations which motivated the non-right-branching structures were shown to be established to the left of the syntactic relations which motivated the right-branching structures. In the one case where this ordering generalization was violated (Heavy PP Shift) contradictory constituency effects were not observed.

This ordering generalization receives a straightforward explanation in the Merge Right approach to syntactic structure building, but is hard to capture otherwise.³⁸

³⁸ My main concern here has been to ask where constituency conflicts are and are not found. In doing so I have placed less emphasis on why some constructions allow the participation of more constituents than others. For example, we saw in §2.5.1

2.6 Constituency Conflicts and Parallelism

This section demonstrates a contrast in the distribution of contradictory constituency effects between VP-fronting constructions of the kind already discussed in §2.5 and VP-ellipsis constructions. The two constructions are superficially similar, in that they involve replacement of a VP by *do* and allow stranding of adverbial phrases.

- (80) a. Mary read the book on Monday and John did on Thursday. (VP ellipsis)
 b. John had to finish the paper, and finish the paper he did on Thursday. (VP fronting)

The contrast that I focus on here is that while the VP-fronting construction (VPF) exhibits contradictory constituency effects, as §2.5 showed, the VP-ellipsis (VPE) construction does not.³⁹ This contrast provides support for the left-to-right approach to structure building proposed here, as we shall see.

Both VPE and VPF involve coordination, and like all coordinate structures they are subject to parallelism requirements. (81) repeats the prediction from §2.3 above about the interaction of parallelism requirements and contradictory constituency effects.

- (81) *Prediction III*
 a. Parallelism requirements across two conjuncts should only be able to refer to properties of the *final* constituent structure of the first conjunct.
 b. Parallelism requirements between conjuncts should block contradictory constituency effects which would be possible in either of the conjuncts individually.

The reasoning behind this prediction is as follows. In a left-to-right derivation the first of a pair of conjuncts will be fully assembled before the second conjunct is built. Therefore, as the second conjunct is being constructed it should only be possible to access the properties of the *completed* first conjunct, and not properties of intermediate stages in the derivation of the first conjunct. Since contradictory constituency effects in this theory are explained with reference to intermediate stages in the derivation of clauses the conditions for contradictory constituency effects should not be available when parallelism constraints apply.

that VP-fronting is subject to a requirement that the fronted phrase be a potential complete VP. Right Node Raising, on the other hand, is not subject to the same restriction, and allows separation of the P and the NP in PPs, for example. I do not have a good account of this contrast at present.

³⁹ Comparative ellipsis shows constituency conflicts in some environments but not others, as §2.6.4 shows.

2.6.1 An Asymmetry between VP-Fronting and VP-Ellipsis

Both VPF and VPE allow fronting/ellipsis of strings of phrases starting at the left edge of VP, and stranding of material from the right edge of VP. Examples are shown in (82–83). These are the kinds of facts which in the past have led people to assume that complex VPs have a left-branching structure.

- (82) a. ... and [give candy to children in libraries on weekends] he did.
 b. ... and [give candy to children in libraries] he did on weekends.
 c. ... and [give candy to children] he did in libraries on weekends.
 d. ... and [give candy] he did to children in libraries on weekends.
 e. * ...and [to children in libraries] he did give candy on weekends.
 f. * ...and [in libraries on weekends] he did give candy to children.
- (83) a. John gives candy to children in libraries on weekends, and Mary does (too).
 b. John gives candy to children in libraries on weekends and Mary does on federal holidays.
 c. John gives candy to children in libraries on weekends and Mary does in urban parks on federal holidays.

Both VPF and VPE show evidence for right-branching structure *within* the fronted/elided portion of VP based on binding evidence, as shown in (84–85), in which no VP-material is stranded.

- (84) a. ... and [introduce **the children to each other**] the teacher proceeded to do.
 b. ... and [congratulate **everybody** on **his** birthday] he did.
- (85) a. The principal introduced **the children to each other**, and then the teacher did (too).
 b. The boss congratulated **everybody** on **his** birthday, and the receptionist did (too).

Up to this point VPF and VPE are entirely alike. However, a contrast emerges when we look at the relations that are possible between the fronted/elided portion of VP and the stranded portion of VP. The examples in (86) parallel examples from §2.5 above which show that the fronted portion of VP has the binding properties that it would have if it were *in-situ* in a right-branching VP. The evidence for this is that material in the fronted portion of VP is able to bind reciprocals or bound variable pronouns in the stranded portion of VP (86a–b), and a quantificational direct object is able to take wide scope with respect to a stranded adverbial, as demonstrated by the availability of a distributive reading for (86c), according to which each individual book-reading was fast.

Constituency

- (86) a. John said he would give books to **them**,
... and give books to them he did [on **each other's** birthdays].
b. Mary said she would congratulate **every boy**,
... and congratulate every boy she did [at **his** graduation].
c. John said he would read **every book**,
... and read every book he did [at **breakneck speed**].

In corresponding examples involving VPE, on the other hand, we do not find corresponding evidence for right-branching structure. Material inside the elided portion of VP is not able to license anaphors or bound variables in the stranded portion of VP, as (87a–b) show.^{40,41}

- (87) a. * John gave books to **them** on **each other's** birthdays, and Mary did [on **each other's** first day of school].
b. * Mary congratulated **every boy** at **his** graduation, and Sue did [at **his** 21st birthday party].

Stranded VP-material takes wide scope with respect to material in the elided portion of VP. This can be seen by comparing the possible interpretations of the single clause in (88) with corresponding examples involving VP-ellipsis.

- (88) Mary finished every book quickly. (ambiguous)

(88) allows both a collective reading in which it is the reading of all of the books which took place quickly, or a distributive reading, in which the reading of each individual book was fast. Speakers tend to report a preference for the distributive reading, which I take to be a reading in which the object NP has wide scope with respect to the adverbial, as in the tree in (89a). Both the collective reading and the distributive reading are available, however.

⁴⁰ Some speakers only accept these sentences with the addition of *too*, or pauses or commas before the stranded adverbials. This does not affect the analysis that follows, and the reader should feel free to add such embellishments wherever they help.

⁴¹ Given the failure of the anaphor and variable binding tests in (87) we might expect similar structures not to induce Condition C violations. The status of this prediction is unclear.

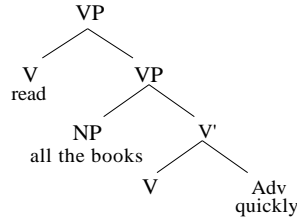
As an example, consider the following sentences (Uli Sauerland, pc):

- (i) *? John gave books to her on Mary's birthday, and Jill did at Christmas.
(ii) (?) John gave books to her at Christmas, and Jill did on Mary's birthday.

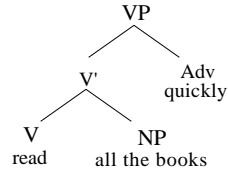
When both the pronoun and the name appear in the same conjunct (i), the Condition C violation is as strong as in non-ellipsis contexts, but when the pronoun and the name appear in differing conjuncts (ii) there is a noticeable improvement, particularly if focus in the second conjunct falls on the word *birthday* rather than the name *Mary*. However, it is hard to determine whether the improvement in (ii) is due to the ellipsis construction or to the kind of focus used.

Constituency

(89) a.



b.



When sentences like (88) are embedded in a VP-ellipsis context, though, ambiguities disappear. Given the two readings of the single conjunct in (88) there are up to four potential readings for the two conjuncts of a VP-ellipsis sentence (i.e. collective–collective, distributive–distributive, collective–distributive, distributive–collective). Only one of these four possibilities is actually available, namely the collective–collective reading in which what was quick (or slow) was the reading of the entire set of books, and not individual book-readings. The unavailability of the two readings in which the conjuncts have differing scopes may be ruled out by appeal to parallelism constraints, but we need an explanation for the absence of the distributive–distributive reading.

(90) Mary finished every book quickly, and John did slowly.
(collective reading only)

This loss of ambiguity is particularly striking because it involves the loss of the reading that is generally *preferred* in the simple sentence in (88), with the consequence that many speakers experience a ‘garden path’ kind of misanalysis when they first read through examples like (90).⁴² To my knowledge, all of the

⁴² The literature contains extensively discussed examples of the interaction of quantifier scope with VP-ellipsis, typically focusing on examples in which the *entire* VP is elided, as in (i) (cf. Sag 1976, Williams 1977, Tancredi 1992, Fox 1995a among others).

(i) Some student admires every teacher, and John does too.

(i) shows that the scopal ambiguity that is present when the first conjunct stands alone disappears in ellipsis contexts in which the second conjunct has a non-quantificational subject.

An important observation about examples in which the entire VP is elided is that the scope reading that went away in (i) becomes available again when the subject of the second conjunct is quantificational, as in (ii) (Hirschbühler 1982, Cormack 1984, Diesing 1992, Fox 1995a).

(ii) Some boy admires every teacher, and some girl does too.

According to Fox 1995a the reappearance in (ii) of the reading in which *every teacher* has wide scope is due to the fact that scopal ambiguities are possible in VP-ellipsis only when *both* conjuncts show a possible scopal ambiguity. This requirement is satisfied in (ii) but not in (i).

Constituency

examples in the literature showing loss of a scopal reading in ellipsis contexts involve the loss of the reading that is the *marked* or *dispreferred* scope reading in simple sentences. This makes the loss of the preferred reading in sentences like (90) all the more striking.

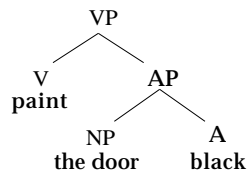
Consistent with the loss of the distributive reading in (90), if we replace the quantifier *every* in (88) and (90) with a quantifier like *each*, which only allows a distributive reading in the simple sentence (91a), we find that VPE becomes impossible (91b).

- (91) a. Mary finished each book quickly. (distributive reading only)
 b. * Mary finished each book quickly, and John did slowly.

Therefore the examples in (87–91) show the following contrast between VP-fronting and VP-ellipsis. When a partial VP is fronted, it has the binding properties that it would have if it was *in-situ* in its underlying position and formed part of a right-branching VP structure. When a partial VP is elided, on the other hand, has the binding properties that it would have if it was replaced in its underlying position and formed a part of a more left-branching VP.

Further confirmation of this contrast between VPF and VPE is provided by constructions which *require* a right-branching VP structure. By hypothesis, resultative constructions require a complement structure in which the object and the result-phrase form a single constituent, as they do in the right-branching structure in (92) (cf. Kayne 1985, Van Voorst 1986, Hoekstra 1988, 1992; but cf. Carrier & Randall 1992, Levin & Rapaport Hovav 1995 for dissenting opinion).

(92)



If VPF but not VPE allows ‘reconstruction’ into a right-branching VP, then we expect that VPF will allow fronting of the verb and the direct object,

In the light of Fox’s generalization, we might then expect that the distributive reading that is lost in (90) could be recoverable by changing the subject to a quantificational NP, such that the object NP *every book* can take wide scope. Surprisingly, this does not help, and still only the collective reading is available, as (iii) shows. The unavailable reading is one in which for each book there is some girl who finished it quickly and some boy who finished it slowly.

- (iii) Some girl finished every book quickly, and some boy did slowly.
 (collective reading only)

For the time being I leave it as an open question why the VP-ellipsis facts discussed here do not seem to interact with Fox’s generalization.

Constituency

stranding the result-phrase, but that VPE will not tolerate similar stranding of the result phrase. This prediction is correct, as the examples in (93–94) show.

- (93) On Saturday Mary resolved to paint her garage door,
... and paint her garage door she did all the colors of the rainbow.
- (94) * Mary painted her garage door black, and John did all the colors of
the rainbow.

Therefore, it seems to be a reliable fact that VPE does not allow reconstruction into a right-branching VP. We can also rule out the possibility that the absence of effects of reconstruction into a right-branching VP is an artifact of semantic or discourse properties of VPE.

This possibility is ruled out by the fact that VPE does allow the scope relations of right-branching VP-structures when the *entire* VP is elided (95), so it cannot be a property of ellipsis *per se* that blocks the distributive reading in (90) above.

- (95) Mary read all the books quickly, and John did too.
(collective & distributive readings both ok)

Example (96) shows that the VP-deaccenting construction (VPD), which has been shown to be very similar to VPE in a number of respects (cf. Tancredi 1992), does not show the loss of the distributive reading that we saw in (90). (96) is most felicitous when the adverbs are read with contrastive stress.

- (96) Mary read all the books quickly, and John read all the books slowly
(collective and distributive readings both ok)

Since VPE and VPD imply exactly the same kind of parallels and contrasts between the two conjuncts, we can rule out the possibility that the loss of the distributive reading in (90) is due to the *semantic* parallelism that has to hold between the two conjuncts in VPE, and we can therefore be confident that the loss of the distributive reading in (90) and the parallel unavailability of right-branching binding relations in (87) is due to some syntactic property of VPE.

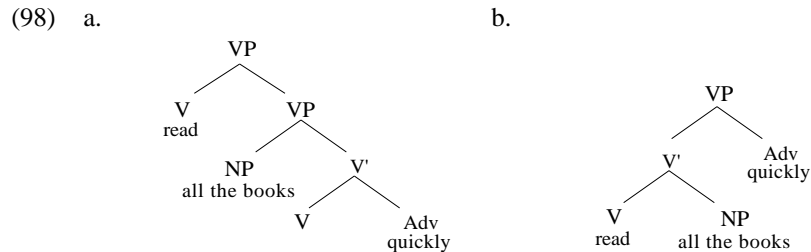
The Merge Right theory provides an account of the contrast between VPE and VPF as follows. In §2.5 I already showed how I assume that contradictory constituency effects are made possible in VPF constructions. This derivation is repeated in (97). Building as usual in a strictly left-to-right fashion, first the fronted portion of VP is built, presumably in a left-adjoined position (97a). The fronted portion of VP is internally right-branching. Next the subject and *do* are added (97b), and then a copy of the fronted VP is inserted in the normal position of VP (97c). At this point the movement chain can be licensed. Subsequent to this the stranded VP material is added at the right of VP, and the structure of the VP can be altered in the now familiar fashion to allow the continuation of a right-branching VP to be built (97d).

Constituency

- (97) a. ... and [give [the book [to them]]]
 b. ... and [give [the book [to them]]] he did
 c. ... and [give [the book [to them]]] he did [give [the book [to them]]]
 d. ... and [give [the book [to them]]] he did [give [the book [to [them [on each other's birthdays]]]]]

In this way we can resolve the apparent contradiction between the kinds of partial VPs that can be fronted, which lead to the appearance of a left-branching structure, but the possibility of the scope and binding relations of a right-branching structure.

Next consider what happens if we try to derive similar effects in a VP-ellipsis construction. I focus here on the loss of the distributive scope reading shown in (90), but the analysis applies equally to the impossibility of binding relations shown in (87). In a strictly left-to-right derivation the first conjunct of the VPE construction will be built in its entirety before the second conjunct is built. Let us suppose that there are two possible ways of deriving the first conjunct, one of which yields a left-branching VP structure, in which the adverbial takes wide scope with respect to the object NP (collective reading), and the other of which yields a right-branching structure (distributive reading). These are the alternatives shown in (89) above, and repeated in (98).



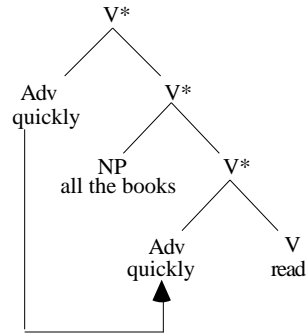
Just as I assumed that only *constituents* of VP may be fronted (although they need not be final constituents), I adopt the standard assumption that only constituents may undergo ellipsis, and that they must also be identical to a constituent of VP in the first conjunct.

If a left-branching VP like (98b) is formed in the first conjunct then the verb and the direct object form a constituent in the final structure for that conjunct. Therefore, ellipsis of the verb and the direct object is possible in the second conjunct when it is built, allowing for collective scope readings. Additionally, the semantic parallelism constraint that the two conjuncts in ellipsis are subject to forces the adverbial to stand in the same relation to the rest of the VP as the adverbial in the first conjunct, i.e. it must c-command the rest of the VP.⁴³

⁴³ Whether the parallelism requirement is strictly semantic or whether it holds at a syntactic level of Logical Form does not matter to this argument. In either case, the adverbial needs to c-command the rest of the VP in both conjuncts, in order for parallel scope readings to obtain.

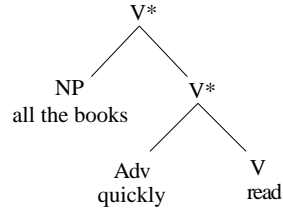
Constituency

(102) a.



collective or distributive

b.



distributive

Japanese has a construction in which a VP is replaced by *soo su*, roughly equivalent to English *do so*. This construction allows stranding of adverbials, just as in English. The one important contrast with English, not surprisingly, is that the pro-VP occurs clause finally, and therefore follows the stranded adverbial. As in English, the scope readings available in sentences like (101a–b) are not all available in the *soo su* construction (103): specifically, collective readings disappear. This makes the order *adverb object* unambiguous, and it makes the order *object adverb* ungrammatical.

(103) a. John-wa isoide dono hon-mo yonda, (sosite) Mary-wa yukkuri soosita.
 John-top quickly all books-acc read, (and) Mary-top slowly did-so
 ‘John read all the books quickly, and Mary did slowly.’
 (collective reading only)

b. * John-wa dono hon-mo isoide yonda, (sosite) Mary-wa yukkuri soosita.
 John-top all books-acc quickly read, (and) Mary-top slowly did-so
 ‘John read all the books quickly, and Mary did slowly.’
 (both scopes impossible)

The loss of the distributive reading in Japanese may be accounted for in the same manner as in English, in terms of the possibility of eliding an object-verb constituent.

The simpler case is (103b), in which the adverb intervenes between the object and the verb in the first conjunct. Assuming that the object cannot move to a lower position in which it forms a constituent with the verb, there is simply no OV constituent which can serve as an antecedent for ellipsis in this sentence. (103a) does allow ellipsis, though with the loss of the distributive reading. (102a) showed how both scopes are derived in non VP-ellipsis contexts. In the collective reading the adverb is interpreted *in situ*, and the distributive reading is obtained by lowering the adverb to its underlying position. I assume that the distributive reading is blocked in VP-ellipsis contexts because the

Constituency

lowering of the adverb which this requires would prevent the possibility of the verb and the object forming a constituent which could serve as an antecedent for ellipsis.

2.6.3 Is Branch Right Violable?

At this point a couple of comments are in order about how it is that non-right-branching VP structures can be built in the first conjunct of VPE constructions in order to license ellipsis of part of VP in the second conjunct. Why do the left-branching VP structures not violate Branch Right, repeated below as (104).

- (104) **BRANCH RIGHT**
Metric: select the most right-branching available attachment of an incoming item.
Reference set: all attachments of a new item that are compatible with a given interpretation.

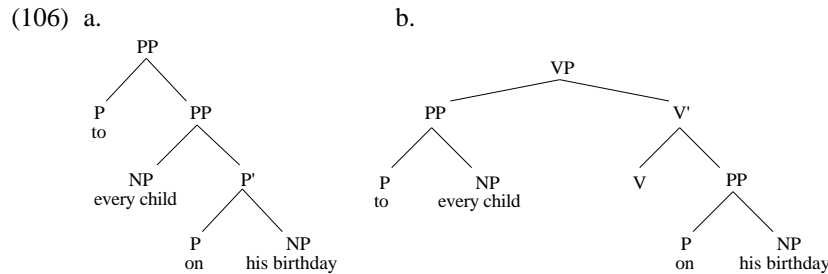
When the object and the adverbial are quantificational, as they are in sentences like (90), both left-branching and right-branching structures are allowed by Branch Right because they give rise to different interpretations. This is not so clear, however, in examples which are just like (90), except that the object and the adverbial show no scopal interaction.

- (105) John read *War and Peace* on Saturday and Mary did on Sunday.

Following the reasoning that was used to explain the unavailability of right-branching structures in (87) and (90), both of the VPs in (105) must have a *left*-branching structure, since this is the only way that the verb-object sequence in the first conjunct can act as an antecedent for ellipsis in the second conjunct. Given that the choice of left-branching versus right-branching VP-structure in the first conjunct of (105) has no effect on the interpretation of that clause, it appears that Branch Right must be violated in order to license ellipsis in the second conjunct. How is this possible?

Although I did not comment upon it at the time, a similar issue arises in connection with the discussion of Heavy PP Shift in §2.5. Recall that the absence of reconstruction effects for binding in Heavy PP Shift was attributed to the fact that rightward PP movement is incompatible with the PP-splitting structures that I assume to be responsible for the possibility of binding out of a PP (106a). Therefore rightward PP movement can only occur if PP-splitting does not occur, as in (106b).

Constituency



In the discussion of rightward PP movement I simply took it for granted that the choice between PP-splitting and PP non-splitting is free, and that the availability of PP non-splitting structures is what makes Heavy PP Shift an available operation. However, this assumption seems to be at odds with the claim that PP-splitting structures like (106a) exist because they are *forced* by Branch Right.

Notice that it is not sufficient to just say that matters of syntactic well-formedness override the choices made by Branch Right, because we have already seen in §2.3 above that a left-branching structure cannot be chosen over a right-branching structure simply to avoid a Condition C violation.

I contend that there is, in fact, a principled contrast between Binding Theory, which does not tolerate violations of Branch Right, and Heavy PP Shift and VP-ellipsis, both of which appear to tolerate ‘violations’ of Branch Right. I assume that the difference between a well-formed and an ill-formed binding relation is purely syntactic and entails no difference in logical form. In the absence of semantic consequences, then, Branch Right cannot be violated.⁴⁵

In the cases of Heavy PP Shift and VP-ellipsis, on the other hand, building a less right-branching structure plausibly has interpretive consequences. If heavy shift associates the shifted phrase with a special focus value, then there is an interpretive difference between movement and non-movement, and therefore Branch Right can be violated in order to make Heavy PP Shift possible. Similarly, if there is an interpretive value associated with identity in ellipsis constructions then this should suffice in order to allow a violation of Branch Right in order to license ellipsis.

Although these remarks are rather brief at this point, they hopefully serve to clarify where Branch Right does and does not apply.

2.6.4 Comparative Ellipsis

If characterization of the contrast between VP-fronting and VP-ellipsis given in §2.6.1 is correct, then it provides one of the strongest pieces of evidence in favor of the Merge Right approach to structure building. There are ways of accounting for the VPE facts *or* the VPF facts in a theory with bottom-to-top derivations or no derivations at all, but the contrast between VPE and VPF will be extremely difficult to capture.

⁴⁵ See Pesetsky 1995 for related observations about the blindness of his Cascade structures to Condition C violations.

Constituency

There is, however, one way of approaching the VPE/VPF contrast that I can see which avoids the need to adopt left-to-right derivations. I have not yet ruled out the possibility that there is some other (as yet unspecified) difference between fronting and ellipsis which happens to allow c-command from the null portion of VP into the stranded portion of VP in VPF but not in VPE. If this is what is responsible for the contrast, then it would just be an accident that the stranded portion of VP that we are interested in is present in the first conjunct in VPE but not in VPF.

In order to control for this possibility, then, we need to be able to manipulate the presence or absence of an adverbial in the first conjunct, without simultaneously switching between ellipsis and movement constructions at the same time. Unfortunately, this test cannot be run with the kinds of VPE and VPF constructions that we have been looking at so far, because these constructions require the strict presence or strict absence in the first conjunct of the stranded phrase in the second conjunct, as (107a–b) show.⁴⁶

- (107) a. * John read the books, and Mary did on Thursday.
b. *? John intended to read the books on Thursday, and read the books he did on Thursday.

A slightly different variety of VP-ellipsis provides the test case that we are looking for. Comparative ellipsis allows for an adverbial stranded by ellipsis to be either present or absent in the antecedent VP, as shown by (108).^{47,48}

- (108) a. John read as many books as Bill did on Thursday.
b. John read as many books on Tuesday as Bill did on Thursday.

The Merge Right approach to structure building predicts that the presence or absence of the stranded adverbial in both conjuncts should affect the availability of right-branching VP structures, leading to just the same contrast as was observed between VPE and VPF. Only when the adverbial is absent from one of the conjuncts should it be possible to find evidence for a right-branching VP structure in the other conjunct. If, on the other hand, the VPF/VPE contrast is

⁴⁶ This is not strictly true, as the following example pointed out to me by David Pesetsky (pc) shows. (i) is marginally possible.

- (i) ? John never read the books, but Mary did on Thursday.

However, the use of a preverbal adverbial in examples like (90) is not a relevant test, because adverbials in this position do not show the scopal ambiguity that they exhibit in VP-final position (cf. *John quickly read all the books*: collective reading only).

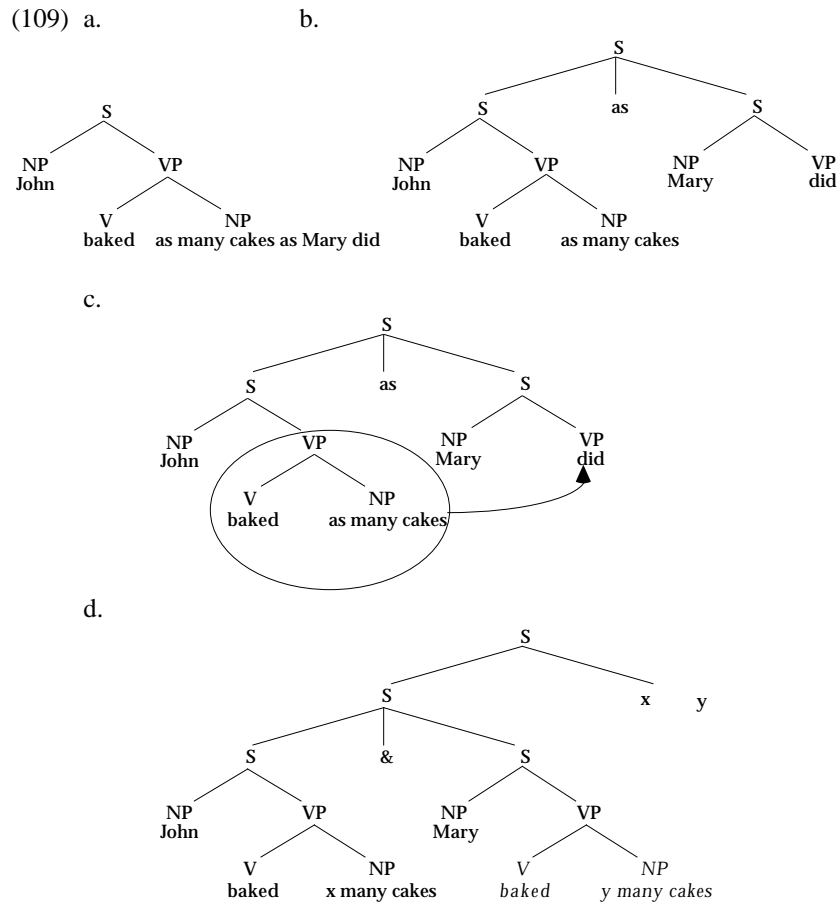
⁴⁷ See Wold 1995 for a proposal for how ACD in comparatives is interpreted, which builds on earlier proposals by Guéron & May (1984), Heim (1985) and Diesing (1992).

⁴⁸ The relevant reading that I focus on for sentences like (108a) is one in which the adverbial that is stranded in the second conjunct is construed with *both* conjuncts, i.e. the ‘Right Node Raising’ interpretation.

Constituency

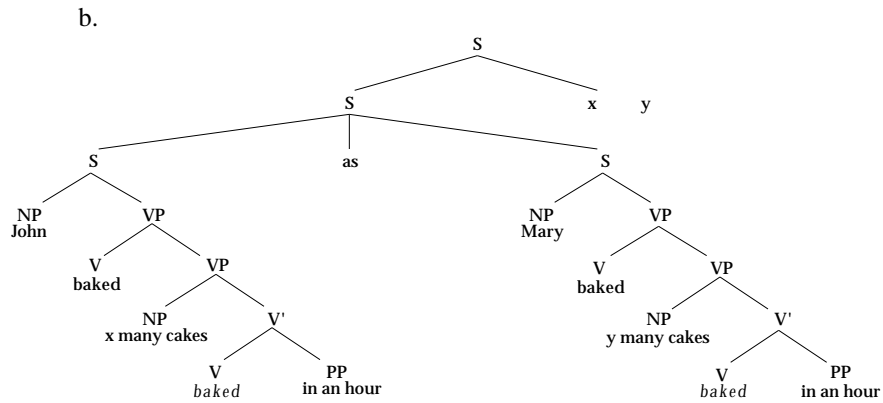
simply due to a quirk of VPE which blocks reconstruction into a right-branching structure, then the presence or absence of an adverbial in the first conjunct should have no effect on the availability of “right-branching” effects in the second conjunct.

(109) shows a sample derivation for a comparative ellipsis construction, basically following Wold 1995, but adapted to the demands of a left-right syntax. (109a–b) shows that initially a normal VP-ellipsis construction is built; accordingly, the VP from the first conjunct is copied into the second conjunct in (109c). The step that is particular to comparatives is the logical form in (109d).



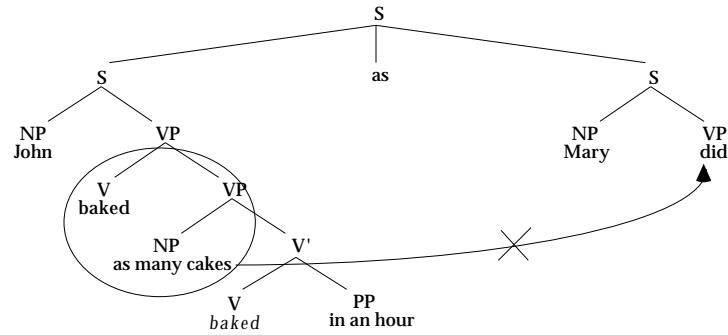
The following examples indicate that the presence or absence of an adverbial in the first conjunct does make a difference to the readings allowed for the second conjunct.

Constituency



Given this, the fact that adding an adverbial to the second conjunct blocks the distributive reading in (110c) and renders pronoun binding in (111b) impossible lends further support to the Merge Right account of the contrast between VPE and VPF.⁵⁰ (113) shows the impossible step that blocks the derivation of the distributive reading for (110c) and (111b).

(113)



(114) shows how the collective reading can be derived for (109c).

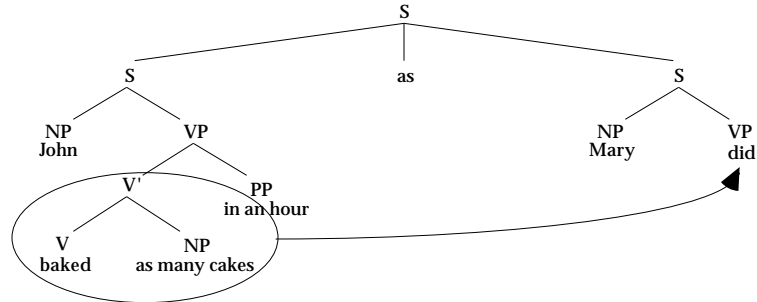
⁵⁰ The examples in (i–ii) contrast with (111a–b) in that the presence or absence of an adverbial in the first conjunct does not have a clear effect on the availability of pronoun binding in (i–ii), unlike (111a–b).

- (i) John gave every girl as many gifts as Bill did at her graduation.
- (ii) John gave every girl as many gifts on her birthday as Bill did at her graduation.

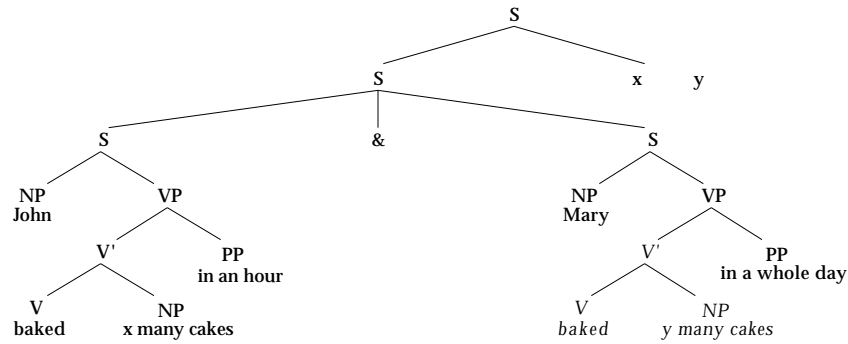
The reason for the difference between (ii) and (111b) may be that the pronoun can be c-commanded and bound by the occurrence of *every girl* in the first conjunct even if it is not c-commanded by the (null) occurrence of *every girl* in the elided VP in the second conjunct.

Constituency

(114) a.



b.



Thus we find internal to comparative ellipsis the same contrast in the distribution of constituency conflicts that we observed between VP-fronting and garden variety VP-ellipsis. Given this, we can rule out the possibility that the VPE/VPF contrast presented in §2.6.1 is due to some as yet unspecified difference between movement and deletion.

2.6.5 Verb-preposition units

In V-PP sequences the verb and the head of the PP appear to behave as a constituent in some constructions but not others. This section demonstrates a contrast in the distribution of verb-preposition units which is similar to the contrasts between VPE and VPF.

The verb and the head of the PP can be coordinated, as (115) shows.

- (115) a. John talked to and gossiped about the kid who sprayed paint on his car.
 b. The cat looked at and then slept on the rug in the middle of the living room.

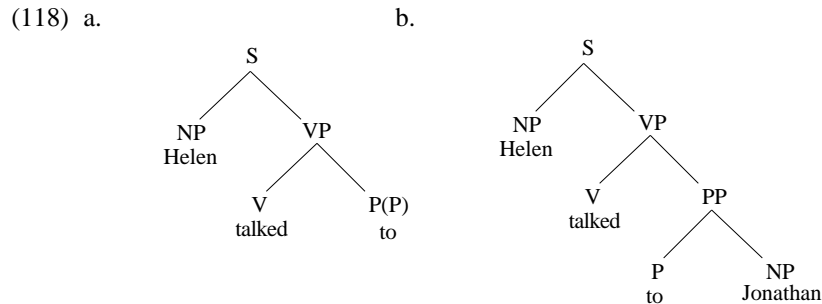
Identical verb-preposition sequences cannot, however, undergo deletion in VP-ellipsis and comparative ellipsis constructions (Postal 1986, Baltin & Postal 1996).

Constituency

- (116) a. Helen talked to Jonathan, and Alice did ___ *(to) Matthew.
 b. The cat slept on the mat, and the dog did ___ *(on) the chair.
- (117) a. Helen talked to Jonathan more often than Alice did ___ *(to) Matthew.
 b. The cat slept on the mat more often than the dog did ___ *(on) the chair.

The contrast between (115) and (116–117) can be explained in much the same way as the contrast between VP-fronting and VP-ellipsis was explained in §2.6.1.

The structures in (118a–b) represent the two stages in the left-to-right derivation of a sentence containing a V-PP sequence.



In (118a), before the complement of the preposition is added, the verb and the preposition form a constituent, but in (118b) the addition of the complement of the preposition has the effect that the verb and the preposition no longer form a constituent. Therefore, using the same reasoning that we have used in a number of other places in this chapter, we expect that V-P sequences will only be treated as a constituent by syntactic relations which are established *before* the object of the preposition is added to the structure. Coordination clearly satisfies this requirement, accounting for the possibility of verb-preposition coordination in (115); in ellipsis the presence of the object of the preposition in the first conjunct accounts for the impossibility of verb-preposition ellipsis in the second conjunct of (116–117). This is because the verb-preposition constituent in the first conjunct is destroyed by the addition of the object, making V-P an impossible target for ellipsis in the second conjunct.

Therefore, the Merge Right approach to structure building provides a common account of two otherwise unrelated contrasts. First, the contrast between binding and scope possibilities in VP-fronting and VP-ellipsis; second, the contrast between the possibility of verb-preposition coordination but the impossibility of verb-preposition ellipsis. These contrasts are all consequences of when different strings are available as constituents at different points in a left-to-right derivation.

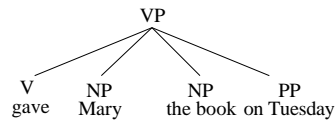
2.7 On the Necessity of Right-Branching Structures

So far in this chapter I have mostly taken for granted the assumption that complex VPs may have a radically right-branching structure. The role of this section is to draw together and extend the motivations for assuming these right-branching structures. This is important, because this approach to VP-structure is far from universally accepted, and because the existence of these structures form a key premise of my arguments here.

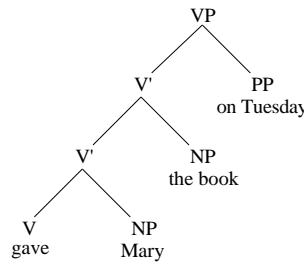
It will be useful to separate three sets of arguments. First, there are the arguments for VP-structures which are right-branching to the extent that the left-to-right ordering of arguments and adverbials corresponds to asymmetric c-command relations. This is the kind of structure proposed by Kayne (1984), Larson (1988), Aoun & Li (1990) among others. Second, there are arguments for the additional assumption that what I have called PP-splitting occurs. This is the possibility first suggested by Pesetsky (1995).

The more 'traditional' assumptions about the structure of VP to which the right-branching structures are to be compared have either a flat n-ary branching structure (119a) or a binary branching left-branching structure (119b), or a mixture of both.

(119) a.



b.



Third, if the arguments for radically right-branching structures go through, we need to ask what the justification is for treating Branch Right as an economy principle rather than an inviolable imperative.

2.7.1 Arguments for Right-Branching VPs

The arguments for right-branching VP-structure in the literature derive primarily from correspondences between constituency motivated by binding and coordination tests. The observation that binding relations in a complex VP are typically possible from left-to-right but not vice versa is consistent with a number of explanations, only one of which involves left-to-right asymmetric c-command relations. However, assuming a right-branching VP-structure makes it possible to give a common account of left-right binding asymmetries and the possibility of coordinating right-edge constituents in complex VPs (cf. 4–5 above).

A further argument for right-branching VP-structures emerges from the discussion of the contrasts between VP-fronting and VP-ellipsis in §2.6. Recall that a central part of the account of why certain scope readings and binding

Constituency

possibilities disappear in ellipsis contexts was that adding an adverbial to a verb-object sequence stops the verb and the object from forming a constituent, and therefore stops the verb and the object from serving as an antecedent to ellipsis. Meanwhile, the same discussion showed that no scope readings or binding possibilities were lost when the entire VP including the adverbial was elided. This implies that the *verb object adverbial* sequence is a constituent. These facts follow under a right-branching analysis of VP, but they are not easily explained under a left-branching analysis like (119b). These facts could be explained under a flat-VP analysis, provided that left-to-right structure building was also assumed.

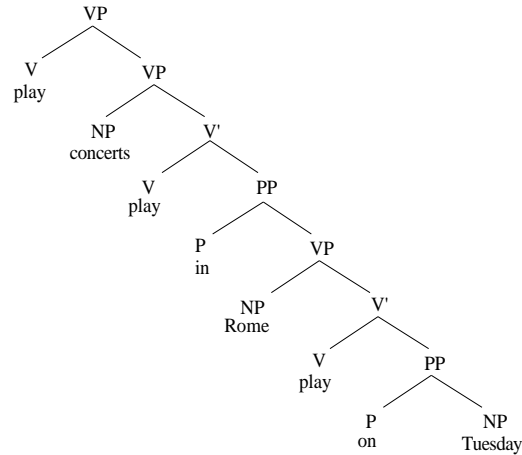
2.7.2 Arguments for PP-splitting

Pesetsky (1995) points out that if we take seriously the logic that leads from binding and coordination evidence to right-branching VP-structures, we also need to assume that non-final PPs are non-constituents in an extremely right-branching VP-structure like the one I have been assuming here. As we have seen already in §2.1 above, an NP can form a conjunct together with a following PP and excluding the preceding preposition that selects it. Also, NPs inside PPs are able to bind elements outside the PP. On the assumption that binding requires c-command, this also implies that prepositions and their complements do not form a constituent, at least when they are followed by additional VP material. Examples of the relevant examples are given in (120–121), and an example of the kind of radically right-branching structure they might motivate is shown in (122).

- (120) *Binding*
- a. Mrs. McGarrick sent a card to **every child_i** on **his_i** birthday.
 - b. The urban-hygiene inspectors departed from **every city_i** during **its_i** rush hour.
 - c. The chef told the guests about **every dish_i** as **it_i** was served.
 - d. Mrs. McGarrick gave a card to **none** of the children on **any** of their birthdays.
- (121) *Coordination*
- a. Andrew wrote to [his mother when he was in LA] and [his aunt Sophie when he arrived in Bogota].
 - b. Sue went to [school in the morning] and [the pottery studio in the afternoon].
 - c. Kremer played concerts in [Rome on Tuesday] and [Somerville on Thursday].

Constituency

(122)



However, one might object that the facts in (120–121) do not justify a move to structures like (122). We might instead suppose that the binding facts arise because prepositions are ‘transparent’ to binding relations, perhaps because they do not count for c-command relations (e.g. Brody 1994), or because the preposition is a case marker which does not dominate the NP in such situations (e.g. Reinhart 1983). Similarly, we might assume that the coordination facts are due to phonological deletion of prepositions.

There are additional facts which favor the PP-splitting analysis over approaches in which prepositions do not count for binding relations or are deletable in coordination. They involve situations where we can show that prepositions PPs sometimes behave as if they are a unit, and sometimes behave as if they have split.

We see this in two examples of the effect of Heavy PP Shift. We have already seen in §2.5.2 above that although NPs can bind out of PPs when the PP is *in situ*, the NP can no longer bind out of the PP when the PP undergoes heavy shift. This is to be contrasted with Heavy NP Shift, which leaves an NP able to bind the positions it was able to bind in its unshifted position. The account I gave for this in §2.5.2 was that Heavy PP Shift requires the PP to be a constituent, and therefore blocks PP-splitting, hence the loss of the binding properties of PP-splitting structures.

A related argument can be found in §4.3.1 in the discussion of restrictions on causer readings for the subject of dative constructions. There I also show that PPs behave as if they undergo splitting when the PP is *in situ* but not when the PP undergoes heavy shift. The diagnostic for PP-splitting in this case involves the availability or not of causer subject readings. See Chapter 4 for further details.

The interaction of Heavy PP Shift with whether or not PPs behave as a constituent is predicted by the account of structure-building that I have presented here, according to which PP-splitting is possible, except where the PP must be a unit for subsequent movement. Meanwhile, the effect of heavy shift is unexpected under the alternative accounts of the facts in (120–121).

2.8 Alternative Approaches to Contradictory Constituency

2.8.1 Points of Agreement and Disagreement

Up to this point I have done little to compare the Merge Right/Branch Right approach to existing approaches to the problem of contradictory constituency. This section attempts to remedy this situation by spelling out where my approach agrees with and where it differs from other accounts of similar phenomena. I focus on two approaches to constituency conflicts: the flexible constituency approach adopted in various versions of enriched categorial grammar (e.g., Steedman 1985, 1988, in press; Dowty 1988; Pickering & Barry 1993), and the parallel structures approach proposed in Pesetsky 1995 (and the related proposal in Brody 1994).

It will be useful to focus the comparison of different theories of constituency around the answers that the various theories give to a set of leading questions about phrase structure and constituency. I take the following to be the main questions to be answered by accounts of constituency conflicts.

- Is a single phrase structure representation sufficient to account for the results of a range of diagnostics?

I think the range of conflicting constituency results is sufficient to rule out the possibility of a single *static* constituent structure that can account for the whole range of results discussed here. On this point all of the theories discussed here agree, including the Merge Right/Branch Right theory.

This conclusion seems at first to be a disappointing one, since the hypothesis that a sentence has a single constituent structure, which is referred to by a wide range of otherwise unrelated syntactic processes, is probably *the* leading idea behind the tradition of phrase structure grammar. We can refer to this as the *Single Structure Hypothesis*. It is a strong and hence extremely interesting hypothesis, but it might not be true, and this is the conclusion reached by Pesetsky, Brody and the Categorial Grammarians.

Although I agree that a single static constituent structure is inadequate, I do not agree that this forces us to depart from the Single Structure Hypothesis in any serious fashion. As I have endeavored to show here, by taking into account how a single derivation for a sentence proceeds from left-to-right, we can account for the kinds of phenomena that pose problems for the Single Structure Hypothesis, but without greatly overgenerating structural possibilities in the process.

- What generalizations, if any, explain why different diagnostics give different results?

Given that not all syntactic processes seem to refer to the same kinds of constituents, the question arises of whether it is possible to predict which kinds of processes will be able to refer to which kinds of constituents. On this question the range of answers is rather diverse. The theory proposed here takes linear order to be the key predictor of what kind of results a constituency test

Constituency

combines with the object by forward application, and then $S \backslash NP$ (i.e. VP) combines with the subject by backward application.

- (125) a. *Forward Application*
 $X/Y \quad Y \quad X$
- b. *Backward Application*
 $Y \quad X \backslash Y \quad X$
- (126) $\text{Leo} \quad \text{saw} \quad \text{Elliot}$
 $\text{---} \quad \text{---} \quad \text{---}$
 $\text{NP} \quad (\text{S} \backslash \text{NP}) / \text{NP} \quad \text{NP}$
 $\text{-----} \rightarrow$
 $\text{S} \backslash \text{NP}$
 $\text{-----} <$
 S

The enrichment of Categorical Grammar which allows the problem of contradictory constituency to be addressed is the existence of type shifting and function composition rules, which make it possible to combine a given set of categories in more than one order, by allowing categories to combine which cannot combine by forward or backward application alone.

For example, by taking advantage of the rules of *Type Raising* and *Forward Function Composition* in (127–128) the sentence in (126) can be derived in a different order, combining the subject and the verb before the object, as shown in (129).

- (127) *Type Raising*
 $X \quad Y / (Y \backslash X)$
- (128) a. *Forward Function Composition*
 $X/Y \quad Y/Z \quad X/Z$
- b. $S / (S \backslash \text{NP}) \quad (\text{S} \backslash \text{NP}) / \text{NP} \quad \text{S} / \text{NP}$
- (129) $\text{Leo} \quad \text{saw} \quad \text{Elliot}$
 $\text{---} \quad \text{---} \quad \text{---}$
 $\text{NP} \quad (\text{S} \backslash \text{NP}) / \text{NP} \quad \text{NP}$
 ---TR
 $\text{S} / (\text{S} \backslash \text{NP})$
 -----FC
 S / NP
 $\text{-----} \rightarrow$
 S

The availability of different derivations for a single sentence makes the description of overlapping constituents a relatively trivial problem. If most sentences have multiple possible derivations, then it is not surprising that conflicting constituency results are predicted. As an illustration of this,

Constituency

constituents are destroyed by subsequent steps of left-to-right derivations, CG derivations add constituents monotonically, so that the non-standard constituents remain at the end of the derivation. I will highlight two consequences of this.

First, if CG derivations are to be equated with PSG structures, then this means that a good many c-command relations are predicted to exist which do not exist in standard phrase structure grammar trees or in the derivations of the Merge Right theory. For example, the direct object *Elliot* in (132) is predicted to c-command the subject of each of the conjuncts.

As we have seen in §2.4 above, though, different c-command tests do not show the kind of variability in their results that we find with constituency tests. C-command tests point fairly uniformly to right-branching structures. Therefore CG structures run the risk of serious overgeneration of c-command relations, and the binding relations captured by c-command relations in phrase structure theories need to be handled in some other way in CG. This fact has not gone unnoticed by CG practitioners, however, and so explicit proposals have been made to account for why c-command tests give much less variable results than constituency tests.

For example, Steedman 1996 argues that binding relations, which form the basis of the majority of c-command tests, are accounted for by a level of predicate-argument structure which is not affected by the different orders in which a given string of words can be combined. In predicate-argument structure binding relations are forced to respect an obliqueness hierarchy of arguments and adjuncts, meaning that more oblique arguments can only be bound by less oblique arguments. This approach to binding is designed to capture the cross-linguistic generalization that the range of possible binding relations is relatively insensitive to word order variation and movement.

The separation of binding relations from how words combine succeeds in preventing massive overgeneration of binding relations, but this move sacrifices an account of the close parallels between coordination and binding relations, as documented in §2.7, and it misses the fact that movement does affect binding possibilities in certain situations.

The literature contains a number of correlations between binding and movement which suggest that movement and binding refer to similar structural representations. The best example of this is certain kinds of local scrambling, often referred to as A-Scrambling, which have been shown to affect binding possibilities (e.g. Yoshimura 1989, Mahajan 1990, Saito 1992). If movement and binding relations depend on the same structural relations, then these correlations are not unexpected, but if movement and binding are unrelated they are somewhat surprising.

The second major shortcoming of the enriched CG approach to constituency conflicts concerns the interactions between different kinds of constituency which were the focus of §2.5–6. If the generalizations about linear order and constituency discussed in §2.5 and §2.6 above are correct, then it is not clear to me how these can be captured by an enriched CG account. Given that (i) if a non-standard constituent is present in a CG derivation it is present *throughout* that derivation, and (ii) CG derivations are not ordered, I see no way of encoding the generalization from §2.5 that certain kinds of constituents are only available *until* material is added on the right that destroys them.

Constituency

The left-to-right order of derivations played a crucial role in explaining the contrast between VP-ellipsis and VP-fronting in §2.6.1, the contrast between two kinds of VP-ellipsis in comparative ellipsis in §2.6.4, and the contrast between verb-preposition coordination and the impossibility of verb-preposition ellipsis in §2.6.5. If my characterization of the facts is correct, then it lends strong support to the claim that the creation of right-branching structures destroys the left-edge constituents that only *appear* to be parts of left-branching structures.

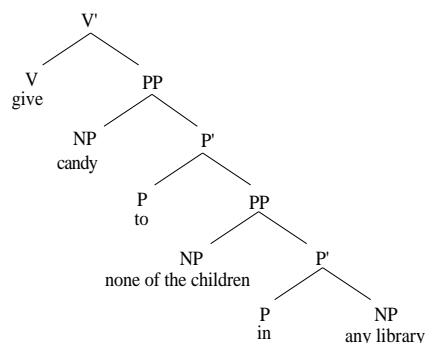
In sum, the Merge Right theory is clearly in agreement with enriched Categorical Grammar theories insofar as it refers to constituents such as *John likes*, which are not allowed under any phrase structure grammar theories that I am aware of. However, while these non-standard constituents are stages in the derivation of (relatively) standard structures in the Merge Right theory, they are retained throughout the derivation in CG theories. I think that this property of CG leads to generalizations being missed about restrictions on what kinds of processes can refer to what kinds of constituents.

2.8.3 Pesetsky 1995: Cascade and Layered Syntax

Pesetsky 1995 gives the most comprehensive account of contradictory constituency effects that I am aware of in a phrase structure grammar approach. Pesetsky's theory represents a much less radical departure from the Single Structure Hypothesis than the enriched Categorical Grammars. Whereas the CG approaches discussed in §2.8.2 allow for an arbitrarily large number of different derivations of a single sentence, Pesetsky proposes that sentences have *exactly two* structures, which are generated by the separate systems *Cascade Syntax* and *Layered Syntax*. The two modes of representation of this *Dual System* differ only in the structures they assign to complex VPs.

The first mode of representation is extremely right-branching Cascade VP structures like (133). These structures are strictly binary branching, and they are just like the structures that I have been assuming here, except that they do not contain multiple copies of V.

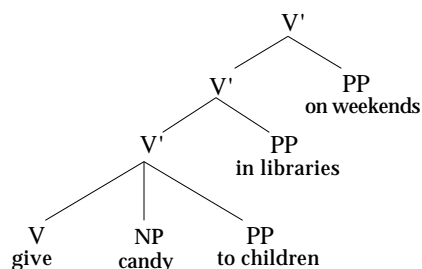
(133)



The second mode of representation is left-branching Layered VP structures like (134). These structures are also binary branching, except that the arguments in multiple complement constructions are assumed to be sisters.

Constituency

(134)



Pesetsky assumes that both structures are represented for all sentences, and that there is a fixed division of labor between the two representations, such that certain syntactic phenomena refer to the constituents of Cascade structures and certain other phenomena refer to the constituents of Layered structures, as shown in (135).

- (135) a. *Layered Syntax*: XP-movement, island conditions on XP-movement, XP-ellipsis, interpretation of modification relations.
 b. *Cascade Syntax*: everything else (binding, polarity item licensing, coordination, etc.)

By including binding relations as something that constituent structure is supposed to account for, Pesetsky's system is able to capture generalizations about parallels between possible binding relations and possible coordinations, which are unavailable in Steedman's Combinatory Categorical Grammar approach, in which binding and constituency are handled by separate representations.

Not surprisingly, the Dual System also makes it straightforward to account for situations in which two properties of a different sentence appear to simultaneously motivate two different constituent structures for that sentence, such as we have seen in sentences where VP-fronting motivates a left-branching structure but the binding relations in the same sentence motivate a right-branching structure (cf. (37) above).

The following are the respects in which I think the Merge Right system differs from or fares better than the Dual System.

First, the same comments about missing generalizations about the relation between linear order and constituency conflicts apply to the Dual System as to Categorical Grammar. If constituency conflicts are the result of independent modes of representation, as Pesetsky assumes, then there appears to be no way of capturing the apparent generalization that the creation of right-edge constituents restricts the possibility of referring to the left-edge constituents that are not constituents in the final right-branching structure.

Second, the Merge Right theory predicts that constituency conflicts should occur in a broader domain than just VP. For example, the conflict between subject-verb and verb-object constituents in Right Node Raising. Given the generality of constituency conflicts, the move of allowing multiple structures for VP will not be sufficient to resolve the contradictory constituency problem.

Constituency

Third, whereas Pesetsky assumes that each sentence has exactly two derivations, in the Merge Right system each sentence has exactly one derivation. This raises the possibility that the Single Structure Hypothesis may be correct after all.

Finally, a comment rather than a criticism. There is an interesting difference between what Pesetsky suggests might underlie the Dual System and the approach to constituency conflicts proposed here. Pesetsky suggests that left-branching Layered Syntax acts as “the only structure that represents semantically contentful relations among items in structure” (p. 289), whereas right-branching Cascade Syntax “might simply prove to be a provisional representation—a ‘sketch’ of syntactic structure, created by the human sentence-processing system as an intermediate step on the way to Layered Syntax.” In the Merge Right system, on the other hand, I have taken almost the opposite view to Pesetsky’s. Structures similar to Pesetsky’s Cascade structures are the only real structures in most derivations, and the appearance of a separate left-branching mode of representation is just an artifact of how right-branching structures are built-up.⁵²

2.9 Conclusion

I have presented a number of reasons in support of assuming that sentences are assembled in a strictly left-to-right fashion, as dictated by the condition Merge Right. The arguments have all been based on the claim that left-to-right derivations of phrase markers provide a better account of constituency facts in English than is otherwise available. First, because of the way in which constituency changes over the course of building a right-branching structure, it is possible to describe the conflicting results of different constituency tests, without needing to posit multiple parallel representations for all sentences. Second, and more importantly, the strictly ordered derivations proposed here make clear predictions about when in a derivation different kinds of constituents should be available, and therefore also which kinds of diagnostics should be able to refer to which kinds of constituents. This makes it possible to begin to explain *why* different kinds of constituency diagnostics yield apparently conflicting results.

It should be clear that there are a good many consequences of Merge Right which I have either ignored or barely touched upon here. In particular, Merge Right entails the near reversal of the way in which syntactic movement operations have standardly been assumed to apply, and further work is required in order to establish the full implications of the change from leftward raising to rightward lowering operations, and whether this is a desirable consequence. Chapter 4 below investigates some of the consequences of left-right structure building for head movement. However, I hope that the consequences of Merge Right that I have discussed here for constituency are sufficiently interesting to

⁵² Note, however, that whereas Pesetsky assumes that his right-branching representations are tentative but are actually constructed, I have been arguing that the appearance of left-branching structures coexisting with right-branching structures is really no more than an artifact of the unjustified inference that if AB is a unit and ABC is also a unit, then they must have the structure [[AB]C].

Constituency

make the possibility of reversing the normally assumed order of syntactic operations worth entertaining.

There is another property of the Merge Right approach which I have said almost nothing about here, but which adds greatly to the interest of exploring its consequences. Merge Right is a necessary property of an incremental parser, and Branch Right is very similar to the local attachment principles which most parsing theories incorporate in some form or another. Chapters 3 and 5 explore this connection in more detail.

Appendix 1: Constraints on Right Node Raising

The aim of this appendix is mostly descriptive, since Right Node Raising has not, in general, been a well-documented phenomenon. §A1.1 presents the (rather small) range of constraints on the *shared* material in RNR, and argues that many of these reduce to Stowell's (1981) *Adjacency Condition* on structural case assignment. §A1.2 focuses on properties of the conjuncts in RNR, and adds to some of the evidence already presented in §2.5.3 showing that RNR does not involve raising.

It might be convenient if we could write off Right Node Raising as a quirk of English which is restricted to a somewhat awkward stylistic register, but RNR is far from a rare curiosity found only in English. It is surprisingly common in normal discourse and writing, and it appears to be an extremely common construction cross-linguistically—I have seen it reported in German, Dutch, Georgian, Polish, Russian, Japanese, Spanish, French and Irish, and I am unaware of languages in which it is claimed that RNR is *not* possible. And yet it is still a rather poorly understood phenomenon. This is not to say that little has been written about it, or that nothing has been learned about it—over the last 25 years a surprisingly large number of squibs and brief discussions (generally towards the ends of papers) have gradually added details to the picture.⁵³ But a complete account of RNR has been hard to come by.

A1.1 Case and Right Node Raising

Beyond its fairly stringent requirements on focus and contrast, RNR imposes remarkably few syntactic requirements. As we shall see below, a number of the syntactic constraints on RNR that have been suggested in the literature disappear once proper attention is given to satisfying the focus/contrast condition on RNR. This section discusses those constraints on the 'shared' material which survive closer scrutiny.

⁵³ For discussions of Right Node Raising see Ross 1967; Hankamer 1971; Maling 1972; Bresnan 1974; Postal 1974, 1994; Abbott 1976; Grosu 1976; Hudson 1976; Wexler & Culicover 1980; Gazdar 1981; McCawley 1982, 1988; Levine 1985; McCloskey 1986; Dowty 1988; Wilder 1994.

Double Focus

RNR constructions generally require a double focus. The coordinated phrases bear contrastive focus, and the shared phrase carries non-contrastive focus. The shared phrase tends to require the same hard-to-pin-down property of ‘heaviness’ which licenses rightward shift of phrases; for example, adding relative clauses to shared NPs very often improves RNR sentences. Both of the two focus requirements become more important as sentence complexity increases, but they apply even in the simplest cases of RNR, as the examples in (136) illustrate. (136a) is fully acceptable; but acceptability decreases markedly in (136b) in which the shared constituent is too light, and in (136c) in which the conjuncts do not contrast.

- (136) a. Heidi ran over and Martha narrowly avoided hitting—the kids who kept dashing out from behind a van to fetch their ball.
 b. * Heidi ran over and Martha narrowly avoided hitting—Rob.
 c. * Heidi ran over and Martha ran over—the kids who kept dashing out from behind a van to fetch their ball.

Controlling for these semantic factors is not always easy, but it is quite crucial. A number of alleged syntactic constraints on RNR disappear once these factors are properly attended to. Not surprisingly, intonation also makes a difference, and therefore reading example sentences aloud often improves otherwise unacceptable sentences, and is recommended in marginal cases.

*ECM Complements

Postal 1974 uses evidence from RNR as one of his arguments for a raising-to-object analysis of exceptional case marking (ECM) constructions. However, extensions of Postal’s data lead to the opposite conclusion from Postal’s, namely that ECM blocks RNR because of its exceptional case properties.

Postal 1974 points out the unacceptability of sentences like (137), in which an ECM complement is shared.⁵⁴

- (137) a. * I find it easy to believe, but Joan finds it hard to believe—Tom to be a dishonest person.
 b. * I wanted to prevent, but I couldn’t prevent—Bob from meeting Sally.
 c. * Tony hopes to stop, but only I can stop—the bomb from going off.
 d. * I didn’t expect to want, but I ended up wanting—Nixon to win.

⁵⁴ Some examples that Postal marks as not so bad, which I’m not sure what to make of at this point. They differ from the other ECM contexts, in that they’re probably NP complements, but why that should make a difference I’m not sure. One possibility would be to just assume that the subjects are getting case from inside the NP.

- i. I didn’t use to favor, but now I do favor—prisoners’ being released.
 ii. I didn’t want to resent, but I did resent—it happening to me.

Constituency

As Postal shows, it is possible to construct various minimal pairs for (137) in which RNR becomes far more acceptable, as witnessed by examples like (138–139). In (138) the non-finite complement of *believe* is replaced by a finite clause complement.

- (138) I find it easy to believe, but Joan finds it hard to believe—that Tom is a dishonest person.

In (139) the complement is still non-finite, but the subject is preceded by a preposition.

- (139) It is possible, but it would be unwise—for Bob to meet Sally.

Both (139) and (140) below, show that infinitival complements do not themselves pose a problem for RNR. It appears to be the ECM subject that is causing the problem, since adding a prepositional case marker (139) or omitting the ECM subject (140) makes RNR possible.

- (140) I want, and Tom really needs—to be out of here by 6:30 to catch the last bus back to Boddington.

The importance of examples like (137) for Postal was that he took the possibility of sharing a phrase in a RNR construction to be a diagnostic of constituenthood. Therefore, the impossibility of RNR in (137) implied that the shared phrases there are not constituents, as predicted under his raising-to-object analysis of ECM complements.

Lasnik (cited in Postal 1974, p.128) suggests a different account of why (137) is bad. He shows that a *that*-less finite complement clause *can* be as bad as an ECM infinitival (140). But in (141a–b) the subject of the shared clause has been modified to the point where speakers are willing to accept the *that*-less finite complement (141a), but they are still unwilling to accept the ECM complement in (141b). Given that I have been unable to construct acceptable RNR sentences with shared ECM infinitivals, though I have been able to rehabilitate otherwise awkward finite complements, I agree with Postal's claim that ECM is responsible for the unacceptability of (137).

- (140) a. * I find it easy to believe, but Joan finds it hard to believe—Tom is dishonest.
b. * I find it easy to believe, but Joan finds it hard to believe—Tom to be dishonest.
- (141) a. I believe, and Harry knows—crossing the Atlantic in a beat-up bathtub would be a stupid thing to do.
b. * I believe, and Harry knows—crossing the Atlantic in a beat-up bathtub to be a stupid thing to do.

However, I believe that Postal's observation is a special case of a more general restriction on RNR.

***Double Object Constructions, Datives OK**

Hankamer (1971: 76) presents (142) as an impossible instance of RNR involving the two objects of a double object construction. He argues that the impossibility of (142) is due to a constraint which restricts the shared material in RNR to a single constituent.

(142) * John offered, and Harry gave—Sally a Cadillac.

Abbott 1976 disputes the claim that only single constituents may be shared in RNR, based on perfectly acceptable examples such as (143–144). (143) shows dative constructions, and (144) shows sequences of direct objects and adverbials. (145) shows further acceptable object-adverbial sequences in RNR, from Grosu 1976.⁵⁵

- (143) a. Joan offered, and Mary actually gave—a gold Cadillac to Billy Schwartz.
 b. Smith loaned, and his widow later donated—a valuable collection of manuscripts to the library.
 c. I borrowed, and my sisters stole—large sums of money from the Chase Manhattan Bank.
- (144) a. Leslie played, and Mary sang—some C&W songs at George’s party.
 b. Mary baked, and George frosted—20 cakes in less than an hour.
- (145) a. John has sliced, and Mary also seems to hav sliced—a large piece of cake with a shining new knife.
 b. Bill may present, and Mary certainly will present—a series of papers at tomorrow’s linguistic meeting.
 c. Mary may have conducted, and Bob certainly has conducted—a number of tests in the large oval lab.

Although I agree with Abbott and Grosu that (143–145) are far better than (142), I do not agree with the implication they presumably wish to be drawn, namely that Hankamer’s example in (142) could also be rehabilitated given appropriate focus, meter etc. I have been unable to construct acceptable right-node-raised double object constructions, even where minimally contrasting dative constructions are acceptable (146).

⁵⁵ Grosu reports that the sentences in (145) are “acceptable to some, but not all, speakers”. I treat these as fully acceptable here, so as to distinguish them from cases like (137) and (142) which I have found nobody who accepts. Those speakers who have difficulty with examples like (145) may place even more stringent conditions on the weight and focus necessary for RNR.

Constituency

- (146) a. The boys sent, and the girls actually gave—a big bunch of flowers to their favorite biology teacher.
b. * The boys sent, and the girls actually gave—their favorite biology teacher a big bunch of flowers.

Abbott 1976 does provide one example of an acceptable RNR involving a double object construction (147), but in this example the first object is contained in each coordinate, and only the second object is ‘shared’. This parallels the improvement of RNR with infinitival complements when the ECM subject was not shared, and clearly implies that it is the first object of the double object construction that is blocking RNR.

- (147) John gave Mary, and Joan presented to Fred, books which looked remarkably similar.

Based on the contrast between double object and dative constructions, I assume that double object constructions, but not complex VP constructions in general, disallow RNR.

If the structure of complex VP constructions is as right-branching as I have argued in this chapter, the acceptability of examples like (143–145) does not challenge the generalization that the shared material in RNR must be a constituent. The focus requirements on the shared constituent may impose this requirement independent of the syntax of coordination. However, this requirement is too weak to exclude all the bad examples presented here, because under my assumptions the two objects of a double object construction are also a single constituent, and likewise ECM infinitivals. Therefore there must be something else that blocks RNR in ECM and double object constructions. What the first object of a double object construction and ECM subjects have in common is that they require structural case. So an initial hypothesis would be that structural case licensing is blocked in RNR environments. This suggestion draws support from a couple of further impossible RNR environments.

*Small Clause

So-called “small clause” complements also appear to resist RNR (148). I have not found any acceptable cases of RNR with shared small clauses.⁵⁶

- (148) a. * I used to find, and Sue still does find—Pete annoying.
b. * I used to consider, and Sue still does consider—the guy who just left the room completely and utterly insane.

⁵⁶ These tests are, of course, only relevant if their non-RNR counterparts are acceptable. Some readers may have difficulty with (ii) in particular, but I find it acceptable, and far better than (148b).

- (i) Sue still finds Pete annoying.
(ii) Sue still considers the guy who just left the room completely and utterly insane.

??NP-CP Complements

Another of Hankamer's starred examples is given in (149). In this case the shared material is an NP-CP double complement construction. (150) shows Abbott's (1976) attempt to rehabilitate the structure (the judgements are Abbott's).

(149) * John told, and Harry showed—Seymour that Sally was a virgin.

(150) ? John tried to persuade, but failed to convince—his skeptical examiners that he knew the right answers.

While I agree that (150) represents an improvement over (149), I find—as Abbott does—that it is still far from entirely acceptable, in contrast to the other complex VP constructions in (143–145) which Abbott manages to completely rehabilitate. I have been unable to improve on Abbott's (150).

Thus far the following generalization emerges. When the NP that begins the shared constituent is dependent on the final head of the conjuncts for *both* case and its theta role (e.g. direct object NP), or for *neither* case nor theta role (e.g. subject of embedded finite clause), RNR is possible. But when the NP that begins the shared constituent depends on the final head of the conjuncts for case alone, RNR is impossible. In (140–141) and (148), in which the shared constituent is an entire non-finite clause, the subject of the clause is an argument of the embedded clause, but is presumably dependent on the matrix verb for case. In the double object constructions in (142) and (146b) the initial NP in the shared constituent is a goal argument, rather than the theme argument which occurs immediately to the right of the verb in dative constructions. Let us suppose that the goal argument is structurally case marked in double object constructions, receiving case from the verb but being the argument of a lower null head (Kayne 1984, Pesetsky 1995 i.a.). And as for the marginality of shared constituents that begin with the NP of an NP-CP double complement construction (149–150), we may assume that this is also a structural case configuration, in which the CP is the underlying complement of the verb but is shifted rightwards for reasons of case assignment (e.g. the *Case Resistance Principle* of Stowell 1981).⁵⁷

Note that the range of complements which cannot be shared in RNR closely matches a class of complements that are impossible in nominalizations (Alec Marantz, p.c.). The contrast in (151) is due to Kayne (1984); the rest of the paradigm in (152–154) is of unknown origin.

⁵⁷ This claim requires further justification. See Stowell 1981, Postal 1986, Pesetsky 1995 for arguments that verbs like *suggest*, *say*, *protest*, *confess*, whose surface complements appear in the order PP-CP, are derived from an underlying structure in which CP c-commands PP. These arguments are based on the impossibility of pseudopassivization (preposition-stranding passivization) with these verbs. I am assuming here an extension of the Stowell/Postal/Pesetsky analysis to NP-CP complement verbs, but this claim cannot be verified using the pseudopassivization test, because pseudopassivization requires the presence of a PP object.

Constituency

- (151) a. * Fred's gift of Helen (of) a dog
b. Fred's gift of a dog to Helen
c. * Sue's handing of James (of) a medal
d. Sue's handing of a medal to James
- (152) a. * Bill's belief of John to be a genius
b. Bill's desire for John to be a genius
- (153) a. * Mark's telling of James that he should stop crying
b. * Steve's persuading of Liam that he should take a nap
- (154) a. * Robin's considering of practice unnecessary
b. * Ian's finding of Robin annoying

If this characterization of the facts is correct, and the main syntactic constraint on RNR involves structural case configurations, then we clearly want to know why structural case should interfere with RNR. One fairly simple possibility is that structural case is subject to Stowell's (1981) *Adjacency Condition*: this requirement can be satisfied between the second conjunct and the shared constituent, but cannot be satisfied between the first conjunct and the shared constituent, for obvious reasons.

A1.2 Constraints and Non-Constraints on the Coordinates

This section focusses mainly on properties of the phrases that are coordinated in RNR. Since there are very few constraints on the coordinates, some of the work of the section is given over to illustrating the range of phrases which RNR could easily disallow coordination of, but which are perfectly possible, given appropriate satisfaction of focus requirements. I also show further evidence that Right Node Raising does not in fact involve raising of the shared material.

CP Material

CP material can be stranded, even in relatives clauses. Bresnan 1974 shows that CP material can be included in the coordinates, leaving just the IP/S as the shared constituent (155).

- (155) a. I can tell you when, but I can't tell you why—he left me.
b. I've been wondering whether, but wouldn't positively want to state that—your theory is correct.

It has been claimed on occasion that Bresnan's observations do not extend to relative clauses. For example, Dowty 1988 points out the impossibility of stranding the CP material from the relative clauses in the examples in (156).

- (156) a. * a man who, and a robot which—can solve this problem
b. * an idea that, and a man who—are much admired

Constituency

For Dowty this is a positive result, because he is attempting to show that RNR can be easily formulated using the flexible constituency allowed by functional composition in categorial grammar, but that RNR is still constrained by island constraints on extraction. Presumably the Complex NP Constraint (or whatever underlies it) is responsible for the impossibility of examples like (167) under this analysis.

However, it is not difficult to find acceptable cases of RNR in which CP material in relative clauses is stranded, once appropriate contrastive focus is supplied. Example (157a) is from Bresnan 1974, and example (157b) is based on a suggestion of Alec Marantz (p.c.).

- (157) a. Tell him almost as much as, but certainly not all that—he'd like to know.
b. I found a box IN which, and Andrea found a blanket UNDER which—a cat could sleep peacefully for hours without being noticed.

What makes examples like (156) impossible is probably the fact that the CP material in the relative clauses lack the content required for them to be used contrastively. (157b) corrects for this problem by relativizing contentful PPs.

Auxiliary Stranding

As Ross showed in one of the earliest discussion of RNR (Ross 1967), the coordinates may end in auxiliaries, leaving an adjective or a predicate nominal as the shared material (158).

- (158) a. Sally might be, and everyone believes that Sheila definitely is—pregnant.
b. Sally might be, and everyone believes that Sheila definitely is—a complete and utter liar.

'Non-Coordinate' RNR

Hudson 1976 points out that RNR is perfectly acceptable in situations where the two "coordinate" phrases are connected by phrases that are not at all coordinate-like, such as verbs (159). See Postal 1994 and section §2.5.3 above for further discussion of this variety of RNR.

- (159) a. Of the people questioned, those who liked outnumbered by two to one those who disliked the way in which the devaluation of the pound had been handled.
b. I'd have said he was sitting on the edge of rather than in the middle of the puddle.
c. It's interesting to compare the people who like with the people who dislike the power of the big unions.

Verb-Final Languages

In languages in which the verb follows all of its arguments, RNR is still often possible, but in these cases the verb, rather than the complements of the verb, is the shared material. The German examples in (160) are taken from Maling 1972.

- (160) a. Weil [Peter den Brief] und [Heidi das Buch] las, wurde keine
Mathematik getan.
because P. the letter and H. the book read was no
math done
'Because Peter was reading the book and Heidi was reading the
letter, no math was done.'
- b. Weil [Peter den Brief geschrieben] und [Heidi das Buch gelesen] hat,
wurde keine Mathematik getan.
'Because Peter wrote the letter and Heidi read the book, no math
was done.'

In matrix clauses in German or Dutch however, in which the verb appears in second position, RNR looks much more like English, as the following Dutch example shows (due to Jan Wouter Zwart, cited in Bobaljik 1995).⁵⁸

- (161) [Jan schrijft], en [Piet leest], artikelen over taalkunde.
J. writes and P. reads articles about linguistics

Although the conjuncts and the shared material is different from English, this kind of RNR is just like English in that contiguous strings from the start of a sentence are coordinated, and the final constituent of the sentence is shared between the two coordinates, as pointed out by Maling (1972).

Medial Coordinates

The coordinates in RNR do not need to include the whole of the beginning of the sentence. RNR constructions can share material at both peripheries of a

⁵⁸ Kayne 1994, pp. 67–68 claims that Dutch does not allow RNR, based on the impossibility of examples like (i), attributed to Teun Hoekstra.

- (i) * Jan heeft gekocht en Marie heeft verkocht de spullen waarmee zij rijk
werden.
J. has bought and M. has sold the things wherewith they rich
became
'Jan bought and Marie sold the things with which they became rich.'

(i) is certainly bad, but this is because it is an attempt to apply RNR in a verb-final clause, presumably requiring rightward movement, in contrast to (161), where no movement is required. See §2.5 and below for evidence that RNR does not involve movement.

Constituency

sentence, coordinating only material from the middle of the sentence, e.g. part of VP.

- (162) a. Mary [bought books] and [sold knitted sweaters] at the flea market.
b. Mary [bought books from an antiquarian] and [sold sweaters to bargain hunters] at the flea market.
c. Mary [bought books from an antiquarian in the morning] and [sold sweaters to bargain hunters in the afternoon] at the flea market.

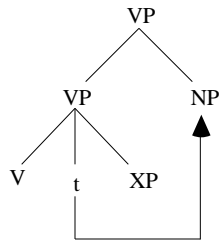
Appendix 2: Heavy NP Shift & Parasitic Gaps

The point of this appendix is to address one of Pesetsky's arguments for his system of dual syntactic representation, based on properties of heavy NP shift. Pesetsky argues that heavy NP shift shows properties of both upward movement in a left-branching VP-structure and left-to-right binding in a right-branching VP-structure. I question the claim that HNPS involves upward movement.

Standard views of Heavy NP Shift (163) treat it as a process which moves a constituent from a position somewhere in the 'middle' of VP to a right-adjoined position at the periphery of VP (164).

- (163) John gave to Mary at her graduation all the old books that he no longer needed.

(164)



This approach assumes a left-branching structure for English VPs. An immediate advantage of this assumption is that it preserves the generalization that movement operations are always 'upward' in character, in the sense that the target of movement c-commands the position in which the movement originated.

Support for the view that HNPS is a raising operation has been drawn from sentences like (165), which have been used to argue that heavy shift licenses parasitic gaps (Engdahl 1983: observation attributed to Tom Wasow). It is standardly assumed that parasitic gaps must at least be c-commanded by the head of a well-formed *wh*-chain.

- (165) Sue offended ____t by not recognizing ____{pg} immediately her favorite uncle from Cleveland.

Constituency

More recently, in the light of facts like (166), some investigators have argued that HNPS shows properties of both raising and lowering operations (Pesetsky 1995; Brody 1994).

(166) Rosa talked ___ about each defendant_i to his_i lawyer.

Therefore, heavy shift poses a challenge for one of the main generalizations that I have been arguing for here. By showing properties of both raising and lowering operations, heavy shift appears to necessitate parallel syntactic representations. There are, however, good reasons to doubt that sentences like (165) involve upwards heavy shift which licenses a parasitic gap: these constructions differ from both heavy shift and standard parasitic gap constructions in a number of respects, and more closely resemble Right Node Raising (RNR) (as argued previously by Williams 1990; Kayne 1994; Postal 1993, 1994). Under the account of RNR argued for in §2.5 and Appendix 1, RNR does not involve raising.

First, the 'true' gap can occur in positions which do not normally license heavy NP shift. Standard heavy NP shift cannot strand prepositions (167), but this is quite possible in alleged pg-licensing cases of heavy NP shift (168), as pointed out by Williams 1990 and Kayne 1994.

- (167) a. * John listened to ___ in front of the fire his favorite Beethoven sonata.
b. * John listened to ___ without recognizing the conductor at once his favorite Beethoven sonata.

(168) John listened to ___ without recognizing ___ at once his favorite Beethoven sonata.

In fact, either gap allows preposition stranding, as (169–170) show.

(169) * John looked at ___ for a few moments his favorite Brazilian bassoonist.

(170) John listened to ___ without looking at ___ for a moment his favorite Brazilian bassoonist.

The second problem for the parasitic gap analysis of (165) involves the fact that the second gap in examples like (165), (168) and (170), which is supposedly licensed by heavy shift from the first gap position, is itself sensitive to constraints on heavy shift. Consider the paradigms in (171–172). (171a) and (172a) are provided for the purposes of orientation: the important thing to notice is that the temporal *until* phrase can only be construed with the higher of the two verbs. It is anomalous if construed with the lower verb. (171b) and (172b) illustrate Ross's (1967) *Right Roof Constraint*: the object of the embedded verb cannot be shifted across material from the higher VP.

(171c) and (172c) show that the object position of the embedded verb is a possible position for a standard adjunct parasitic gap, and (171d) and (172d) show that the two gap positions also license an analog of (165), (168) and (170)

Constituency

above. Given these preliminary tests, we predict that if HNPS licenses parasitic gaps, then the (e) examples should be acceptable. The (e) examples are modeled on the parasitic gap constructions in the (c) examples, except that they involve heavy NP shift to the right. The (e) examples are at least as bad as the Right Roof Constraint violations in (171b) and (172b).⁵⁹

- (171) a. John refused [to notice his nephew] until the end of the movie.
b. * John refused to notice ___ until the end of the movie his nephew who was hiding under the bed.
c. Who did John upset ___ by refusing to notice ___ until the end of the movie?
d. John upset ___ by refusing to notice ___ his nephew who was hiding under the bed.
e. * John upset ___ by refusing to notice ___ until the end of the movie his nephew who was hiding under the bed.
- (172) a. The ambassador refused to start recognizing the breakaway state until the summit.
b. * The ambassador refused to start recognizing ___ until the summit the breakaway state that had just been formed in northwest Somerville.
c. Which state did the ambassador offend ___ by refusing to start recognizing ___ until the summit?
d. The ambassador offended ___ by refusing to start recognizing ___ the breakaway state that had just been formed in northwest Somerville.
e. * The ambassador offended ___ by refusing to start recognizing ___ until the summit the breakaway state that had just been formed in northwest Somerville

The illformedness of (171e) and (172e) is unexpected under the parasitic gap analysis of sentences like (165), which assumes that the only instance of movement is out of the *first* gap position. The parallel with the Right Roof Constraint violations in (171b) and (172b) implies that the heavy NP must have been moved out of *both* of the gap positions in the (e) examples.⁶⁰ In the next

⁵⁹ One difficulty that arises in constructing paradigms like (171–172) is that the alleged parasitic gap is inside a non-finite clause, but Right Roof Constraint effects appear to be less strikingly bad in non-finite clauses. For some speakers, non-finite clauses do not induce Right Roof Constraint violations at all (e.g. (b) examples in (171–172), and therefore the paradigms in (171–172) are not relevant tests for these speakers.

⁶⁰ The fact that both gaps must be possible starting points for HNPS also makes sense of the ill-formedness of (i) below (Pesetsky 1995, ex. 652b).

- (i) * Sue offended ___ by informing ___ about the kids her favorite uncle from Cleveland.

Constituency

section I show how this fact is expected if examples like (165) and (171d, 172d) are analyzed as instances of Right Node Raising.

In addition to examples like (171) and (172), which show that the alleged rightward movement variety of parasitic gap licensing fails to license gaps that are licensed by leftward movement, there are a number of contexts which allow parasitic gaps with rightward movement but do not allow parasitic gaps with leftward movement, as documented extensively in Postal 1994.

The examples in (173–175) illustrate just one of Postal's arguments. Leftward movement tolerates parasitic gaps that are separated from the fronted phrase by one island (173) but not two (174). However, rightward movement does seem to tolerate gaps inside two islands.

- (173) Who did Sue offend ___ by not recognizing ___?
- (174) * Who did Sue offend ___ by not recognizing the people who were supporting ___?
- (175) Sue offended ___ by not recognizing the people who were supporting ___ at that time her favorite uncle from Cleveland.

The reader is referred to Postal 1994 for further discussion of differences between the multiple gaps resulting from leftward and rightward movement.

To summarize, in multiple gap constructions like (165), which have been claimed to be instances of parasitic gaps licensed by heavy NP shift, we observe the following properties:

- both gaps can occur in positions which do not allow heavy NP shift (P-stranding)
- the alleged parasitic gap is impossible in positions which normally allow parasitic gaps (Right Roof Constraint violations)
- the alleged parasitic gap is possible in positions in which gaps are impossible in parasitic gap constructions involving leftward extraction (island violations etc., cf. Postal 1994)

I therefore conclude, in agreement with Postal 1994, that these constructions are not parasitic gap constructions. Therefore, it is not so clear that Heavy NP Shift shows the dual character of raising and lowering operations, contrary to the

Pesetsky argues that (i) is ruled out because of a categorial difference between the complements of *offend* and *inform*, which violates a requirement that the various gaps in a parasitic gap construction must match in category.

However, given that the gaps in (i) do tolerate a standard parasitic gap construction involving leftward extraction, as (ii) shows, it would seem simpler to assimilate the impossibility of (i) to the impossibility of heavy NP shift from the complement of *inform* in (iii) (Pesetsky's ex. 628b).

- (ii) Who did Sue offend ___ by informing ___ about the kids' medical history?
(iii) * Sue informed ___ about the kids her favorite uncle from Cleveland.

I do not have an account of why the complement of *inform* disallows HNPS.

Constituency

suggestion in Pesetsky 1995 and Brody 1994. Meanwhile, the properties of allowing preposition-stranding and failing diagnostics of movement are reminiscent of properties of Right Node Raising, as documented above.

Constituency

Chapter 3

Parsing

3.1 Issues in Parsing

3.1.1 Overview

In this chapter we turn our attention to issues in parsing and sentence comprehension. I argue that the economy condition *Branch Right*, which was introduced in Chapter 2, plays an important role in determining default structural choices in parsing. The evidence for this is that a wide range of structural ambiguities are resolved in favor of the most right-branching of the available alternatives.⁶¹ Ultimately, this parallel between ambiguity resolution strategies and grammatical economy conditions forms one part of the argument for the unification of the parser and the grammar, but this issue will be backgrounded here so that we may focus on more parsing-specific questions.

The first section of this chapter shows where the topic of ambiguity resolution fits in with a general account of sentence processing. This section is primarily intended for readers who are not familiar with the literature on structural ambiguity resolution. Then, in §3.2 I present the results of an experiment conducted in collaboration with Ted Gibson which demonstrates that the structural preference for right-branching structures is stronger than is normally assumed. In most previous studies of ambiguity resolution the more right-branching alternative coincides with the simpler alternative according to a host of other complexity metrics (structural, semantic, probabilistic and discourse-based). Our experiment tests a situation in which most complexity metrics predict a preference for the less right-branching alternative, but in fact the more right-branching alternative is preferred.

⁶¹ This chapter represents an updated and extended version of the approach to structural ambiguity resolution in Phillips 1995, which in turn is a revival of an idea explored by Janet Fodor and Lyn Frazier in unpublished work (Fodor & Frazier 1983). Apart from the different name for the main structural complexity metric (*Branch Right* instead of *Right Association*), which I think better reflects what the principle does, the substantive differences between this chapter and my earlier paper are that the underlying phrase structure theory is made more explicit, there is extra experimental evidence for the strength of local attachment, and there is more discussion of structural parsing preferences in languages other than English.

Parsing

In §3.3 I show how Branch Right accounts for a wide range of structural preferences in ambiguity resolution in English, subsuming and going beyond the empirical coverage of principles like *Minimal Attachment* (Frazier & Fodor 1978, Frazier 1987, Gorrell 1995) or the many versions of the *attach as argument* preference (Ford, Bresnan & Kaplan 1982, Abney 1987, 1989, Pritchett 1988, 1992, Gibson 1991, Crocker 1996). Since almost all of these approaches also incorporate a local attachment strategy, one way of reading this section is as a demonstration that it is possible to get a lot more out of something that most people assume already.

In §3.4 I address the concern that a preference for right-branching structures might encounter difficulties in languages that are claimed to be more left-branching than English. I show that Branch Right accounts for some parsing preferences in head-final constructions in German and Japanese, and also show that languages like Japanese are only left-branching in a limited sense, and are otherwise rather similar to English.

In §3.5 I discuss some challenges to the generality of Branch Right in parsing, involving cross-linguistic differences in parsing preferences for superficially identical constructions, and concerns that Branch Right is too inflexible. §3.6 addresses some residual issues involving the modularity (or not) of different kinds of information in sentence comprehension.

3.1.2 Representations and transitions

In this section and §3.1.3 I give a brief overview of the main issues in sentence parsing and how they relate to one another.

The question of how sentences are understood can be most usefully divided into separate questions about what representations are built at various stages in parsing a sentence and how the parser progresses from one representation to the next in incremental parsing. In other words, the questions can be partitioned into the following:

- (1) a. What representation(s) are the end-point of understanding a sentence?
- b. What kinds of representations are the intermediate stages of understanding a sentence?
- c. How does a speaker proceed from one intermediate representation to the next in incremental parsing?

For each of these questions the range of possible answers covers a variety of language-specific and general cognitive representations and processes. And in fact, a large part of the work in the field of sentence processing has been devoted to assessing the relative importance of language-specific and general cognitive processes.

In general, there has been a shift of emphasis over the last 30 years in research on sentence processing. In the 1960s questions (1a) and (1b) appear to have been the focus of attention, but since then work on these particular questions has been pursued in most detail in the field of computational linguistics. Much of the action here is in evaluating different kinds of

declarative linguistic knowledge representations and how they may be exploited under the time, memory and linear order constraints of real-time parsing.

Meanwhile, work on sentence processing in experimental psychology over the past 20 years has focused mostly on question (1c). This is not so much because a consensus was reached on the answers to (1a) and (1b) as because question (1c) has proven to be best suited to the techniques and concerns of experimental psychology. Questions about the nature of representations (i.e. 1a–b) have been experimentally investigated in the small number of cases where different theories make clear enough predictions for reaction time studies, as in the literature on whether the parser's representations include movement traces or not,⁶² but such cases are rare.

3.1.2.1 Endpoints

Experimental work in the 1960s seems to have reached a consensus that part of understanding a sentence involves building a grammatically defined structural description for the sentence. Rather than reviewing the evidence for this, I refer the reader to the useful surveys of work on this topic in Fodor, Bever & Garrett 1974 and Levelt 1974 (vol. III).

Most work on sentence parsing since then has assumed that a surface structure phrase marker is constructed for the incoming sentence. If a non-transformational theory of syntax is assumed, then this is the only syntactic structure for the sentence; if a transformational theory with movement traces is assumed, then the surface structure contains all of the properties of the underlying structure, in the form of traces. To the extent that there has been any controversy on this point, it mostly reduces to reasons for one choice or another of syntactic theory, and does not question the assumption that a surface structure representation is built.

In what follows I will take it for granted that understanding a sentence involves recovering its structural description as defined by the grammar of the language. By 'structural description' I mean the entire derivation from the surface string of the sentence to an underlying representation which encodes properties of both the D-structure and Logical Form levels of Extended Standard Theory approaches to syntax, as described in more detail in Chapter 2.

3.1.2.2 Intermediate representations

The assumption that sentence parsing involves building a structural description for the sentence leaves open a number of possibilities for how this structural description is arrived at.

A possibility that was entertained for a while in the 1960s and then largely ignored was that the intermediate stages in understanding a sentence correspond to intermediate stages in a 'backwards' transformational derivation of a sentence from its surface structure to its deep structure. I will not discuss this alternative in much detail here, because it is taken up in more detail in Chapter 5 below.

⁶² See Bever & McElree 1988, MacDonald 1989, Nicol & Swinney 1989, Pickering & Barry 1991, Hickok 1993, Gibson & Hickok 1993, Gorrell 1993, Fodor 1993, Sag & Fodor 1994.

Parsing

Although this is often not made explicit, there seems to have been a broad consensus over the last 20 years that the intermediate stages in parsing a sentence correspond to incomplete surface structure representations of a sentence, which are incrementally assembled from left-to-right. Since phrase structure trees with material missing on the right-hand side are not well-defined objects in most theories of grammar,⁶³ these partial structures are generally assumed to be assembled by a special-purpose parsing device which builds partial phrase markers for partial sentences. A good deal of the work in this area involves the determination of how the kinds of rules and constraints that the grammar specifies are to be exploited in these incomplete phrase markers, and for this reason it is a topic that has been more profitably explored in computational linguistics.

One issue concerning intermediate representations which has received more attention in the psychological literature on parsing is the question of whether the parser pursues just one parse of a sentence at any time (*serial* models) or whether it sometimes pursues multiple analyses at once (*parallel* models). The dominant view has been that the parser is serial, but a number of interesting arguments for parallelism have been given (cf. Kurtzman 1985, Gorrell 1987, Gibson 1991, Hickok 1993).

In what follows I assume that the intermediate stages of processing a sentence correspond to the intermediate stages in left-to-right grammatical derivations as defined in Chapters 2 and 4. I assume that the parser is serial in the sense that it only carries one representation from one step of parsing to the next, although it may evaluate multiple possible representations at various decision points in parsing.⁶⁴

3.1.3 Transitions: ambiguity resolution

Most contemporary experimental work in sentence processing has focused less on the nature of representations and more on the question of how incoming information is used by speakers to guide the transitions from one intermediate representation to the next, for example, how does the parser move from its parse of *John saw* to a parse of *John saw Mary*.

In some cases the grammar determines a unique answer to how the incoming word and the current representation can be combined. The main issue in such cases is how the parsing device accesses the information that there is only one grammatical option. This question is therefore most naturally combined with investigations of the properties of partial phrase markers and how they relate to grammatically defined complete phrase markers, and has not been addressed in much detail in the experimental psychological literature on parsing.

⁶³ Some versions of categorial grammar which allow a given sequence of words to be combined in a number of different orders are exceptions to this (e.g. Ades & Steedman 1982). See Chapter 2, section 8 for discussion of this approach.

⁶⁴ In fact, as far as the arguments of this chapter are concerned the choice between a serial model and a model with limited parallelism does not make much of a difference. If every claim that the more right-branching structure is chosen is replaced with the claim that the more right-branching structure is the highest ranked of the alternatives that are pursued, the same predictions hold.

Parsing

Experimental studies of sentence processing have tended to focus on situations in which the incoming word and the existing representation can be combined in more than one way consistent with the grammar, i.e. structural ambiguities. This is a particularly rich area for experimental investigation, because it is possible to experimentally manipulate the various factors that might influence people in choosing between the different possible continuations of an incomplete sentence (structural, pragmatic, probabilistic, lexical etc.). In fact, the study of ambiguity resolution dominates research in sentence processing to such an extent that it can sometimes seem as if ambiguity resolution is the only topic in sentence processing. But a theory of ambiguity resolution is just one component of an account of sentence parsing, let alone a theory of sentence comprehension.

The term ‘ambiguity resolution’ covers a number of different topics. First, there is the question of what choices the comprehender makes in structurally ambiguous situations (*first pass* parsing); second, there is the question of what happens when the comprehender realizes that an earlier choice was incorrect (recovery and reanalysis).

Within the area of first pass parsing, we can ask both what speakers do when the ambiguity is completely unbiased, i.e. where there are no clues from context or lexical frequencies about which of the alternative continuations is correct. We can also ask what speakers do in biased situations, i.e. situations where there are a number of potential clues which could be used to help choose an analysis from among the various alternatives.

For the situations in which there are no clues to help the comprehender, the question reduces to the question of what structural biases speakers come equipped with, either by virtue of their grammar or by virtue of parsing specific principles. In other words, these are questions about *structural complexity metrics*. It can be difficult to find situations in which there really are no additional contextual or probabilistic ‘clues’ to help the comprehender choose an analysis, as a good deal of recent work in sentence processing has shown, but this does not detract from the interest of investigating syntactic complexity metrics, it only makes them more difficult to study.

For the situations in which there are potential clues to how the structural ambiguity will be resolved, the investigator’s task is to find out which of these potential sources of information are available to the comprehension system, what their relative importance is, and how and when the different sources of information are integrated. In other words, here the focus is on questions of *modularity*; how much cross-talk is there between different information sources in the brain.

It is important to separate the question of what the relevant structural complexity metric is from the question of how modular the flow of information is in language comprehension. The issues are sometimes conflated. If it is true that people are able to integrate multiple sources of information very quickly in language comprehension, then this is an important discovery, but it does not make it any less relevant to investigate questions of structural complexity.⁶⁵

⁶⁵ Different views are found, as for example in the following passage from MacDonald, Pearlmuter & Seidenberg 1994:

Neither pursuit is any more or less valuable than the other *a priori*; nor is any particular set of data (e.g. commonly occurring sentences) privileged.

Similarly, the investigation of different possible structural complexity metrics does not imply any commitment to one or another position on the modularity question. I stress these points here merely because some recent discussions of these issues might present the impression that these lines of inquiry are mutually exclusive.

The main reason for this conflation of issues seems to be the unfortunate accident that interests in modularity or structural complexity do seem to cluster with views on the modularity issue. For example, a good deal of the most influential work on sentence processing in the 1970s and 1980s focused on complexity metrics and also argued that the flow of information is highly encapsulated in sentence processing. This may have led to the impression that these interests were in some way dependent on one another.

Throughout most of this chapter I focus on questions of structural complexity. Sections §3.5 and §3.6 contain some discussion of the status of the modularity question, but I will have little to say on this topic here, not because I do not consider the issues important, but because it is not of central importance to the main argument of the thesis involving the relation between the parser and the grammar.

3.1.4 Structural complexity metrics

A number of different structural complexity metrics have been explored in the literature on sentence processing. Here I give a brief summary of some of the most well-traveled lines of inquiry in this area, treating some closely related proposals together, and ignoring certain differences between related proposals. For more detailed discussions of the various different theories I refer the reader to the review chapters found in a number of recent books on sentence processing (Crocker 1996, Frazier & Clifton 1996, Gibson 1991, Gorrell 1995, Pritchett 1992).

'As the dominant theory of sentence processing, the garden path theory suggested what kinds of empirical questions are worth pursuing. The central question, of course, was whether there is an autonomous parser that obeys the minimal attachment principle. In our theory, this question is no longer the focus because whether or not the interpretation described as "minimal attachment" is preferred depends on other factors. These factors become the new focus of attention.' (p. 697).

Claims like this present the misleading impression that the study of structural complexity metrics and other biases are mutually exclusive lines of research. If the nature of the enterprise is to predict how ambiguities are resolved in naturally occurring corpora of sentences, with some clinical or engineering goal in mind, then this may be true. However, if the aim is to understand the components of the organism that parses sentences, then these are entirely independent questions. Note that this point does not question MacDonald et al.'s empirical findings about how multiple sources of information can interact in ambiguity resolution. What I am questioning is whether their findings bear at all on questions of the existence of structural complexity metrics, except to the extent that their results make it harder to study structural complexity metrics, because of the many additional factors which must be controlled for.

Parsing

It will be convenient to divide the different proposals into three groups. First, complexity metrics based on the overall formal simplicity of a phrase marker; second, metrics based on formal simplicity of a very local nature; third, metrics based on maximal satisfaction of grammatical requirements such as case and thematic role assignment.

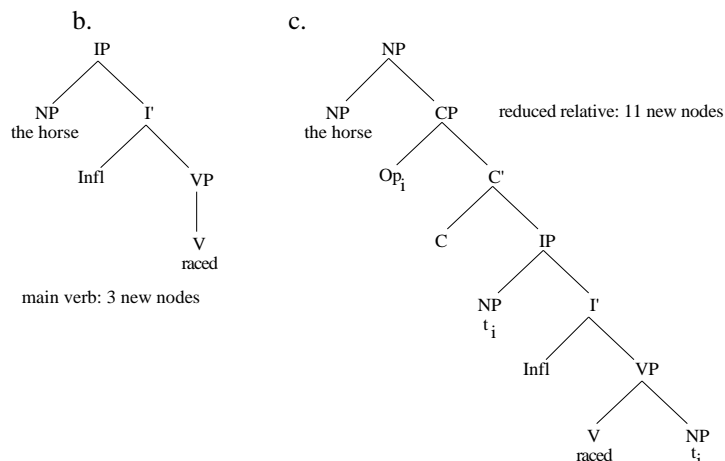
Global Formal Simplicity. The most well-known instance of a global formal simplicity metric is the principle of *Minimal Attachment* of Frazier & Fodor (1978). This principle states that the transition to the next parse state which adds the fewest number of new nodes to the current phrase marker is the preferred one. When this principle was originally proposed in the 1970s it was assumed that it probably did not need to be explicitly stated, but was instead an emergent property of a parsing device which chose the first available analysis of incoming material. This argument relied on the assumption that phrase markers are constructed by accessing rewrite rules (of the form A → B C), and that different rewrite rules take a uniform amount of time to access. These assumptions had the consequence that the first available analysis would be the one that required reference to the fewest phrase structure rules and hence the fewest new nodes, thereby deriving the principle of Minimal Attachment.

The question of whether Minimal Attachment is an explicitly stated component of the parser or an emergent property has become less clear as grammatical assumptions have shifted. For example, Gorrell 1995 adopts a version of Minimal Attachment, which he dubs *Simplicity*, in the context of a parser which embeds a Government-Binding style syntactic theory. Gorrell makes no claims to a speed-of-access grounding of this principle, and instead assumes that the principle is the parser's implementation of the economy of representation principles that have been explored by a number of linguists in recent years (e.g. Chomsky 1991, 1995, Grimshaw 1996, Speas 1990).

The application of the Minimal Attachment principle to a classic ambiguity from Bever 1970 is shown in (2). In this ambiguity, the reduced relative clause parse of the verb *raced* requires many more new phrase structure nodes than the main verb treatment.

Parsing

- (2) a. # The horse raced past the barn fell.



Locality. There are a number of different versions of formal complexity metrics that select the new parse state which attaches the incoming word as close as possible to immediately preceding words. The best known of these are Kimball's (1973) principle of *Right Association* and Frazier's (1978) *Late Closure* strategy. These strategies differ in their details, but they both require that incoming words attach to the lowest possible existing node in the phrase marker. A number of other versions of this principle have been proposed which by and large have the same effects: *Attach Low* (Abney 1989), *Recency* (Gibson 1991, Gibson et al. 1996), *Attach Bottom-up* (Stabler 1994), among others. Attempts to provide grounding for the local attachment preference have tended to draw connections with well-known recency effects in memory.

Some well-known examples of ambiguous sentences which have been taken to motivate a local attachment principle are given in (3) (cf. Kimball 1973). In each case there are two preceding heads which the final word of the sentence could associate to, and association to the most recent word is consistently preferred.

- (3) a. John said Bill left yesterday.
 b. Joe figured that Sue wanted to take the train to New York out.
 c. I met the boy who Sam took to the park's friend.

Grammatical-constraint satisfaction. A further set of structural ambiguity resolution strategies involves principles which care less about the purely formal properties of the phrase marker than about the grammatical relations which the partial phrase marker satisfies or leaves unsatisfied. The principles typically place some premium on the satisfaction of relations such as thematic role assignment, case assignment or chain completion. In this area most attention has been given to the role of argumenthood. Some proposals require that when an incoming item can be treated either as an argument (of an existing head) or as an adjunct, the argument option is to be chosen (Ford,

Parsing

Bresnan & Kaplan 1982, Abney 1987, 1989, Gibson 1987, Crocker 1996, Schütze & Gibson 1996), whereas other proposals have required that ambiguities be resolved in such a way that as few as possible predicate-argument relations are left unsatisfied (Pritchett 1988, 1992, Gibson 1991).

Argument-based principles have been generally fairly successful, for the reason that many of the structural ambiguities that have been studied involve a choice between an argument and a non-argument treatment of the incoming word, as we shall see in §3.3 below. Some of the parsing preferences which are straightforwardly handled by such approaches are shown in (4). In each case the speaker is faced with a choice between treating an incoming phrase as an argument of the verb or as part of a modifier of the first object of the verb. Speakers generally pursue the argument attachment in these cases (see sources cited in §3.3 below).

- (4)
- a. John put the book on the table on the shelf.
 - b. Mary gave the man the dog bit a package.
 - c. The patient persuaded the doctor that he was having trouble with to leave.

Almost all structural accounts of ambiguity resolution have assumed that more than one of the strategies listed above is necessary to account for the full range of structural ambiguities in English, let alone other languages. Most of the theories listed above choose either a global structural complexity metric like Minimal Attachment or an argumenthood-driven strategy, in either case together with some version of the local attachment principle. It is generally taken to be impossible to do without a local attachment preference, because of simple examples like (3a) above, in which neither of the two possible parses involves an argument attachment and neither parse involves the creation of more nodes than the other—the only difference between the two alternatives is the locality of the attachment site to the position of the immediately preceding word. My goal here is to show that the local attachment preference in fact subsumes all or most of the empirical coverage of other structural complexity metrics.

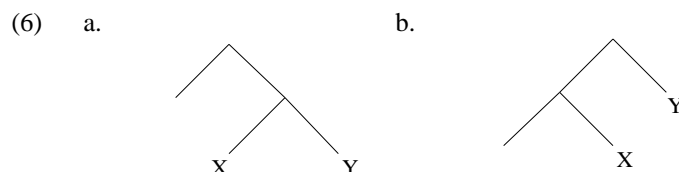
3.1.5 Branch Right

My aim in this chapter is to argue that not only is the local attachment preference, here in the form of the Branch Right principle, an indispensable component of theories of ambiguity resolution, it can also subsume most or all of the work of other structural parsing principles. Branch Right is stated in (5). It locally forces the construction of right-branching structures by requiring the paths through the phrase marker from one terminal node to the next terminal node to be as short as possible.

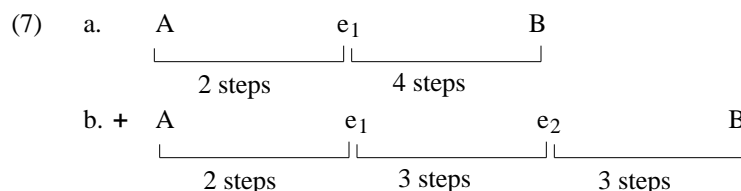
- (5) **BRANCH RIGHT**
Metric: select the attachment that uses the shortest path(s) from the last item in the input to the current input item.
Reference set: all attachments of a new item that are compatible with a given interpretation.

Parsing

For example, if the last item in the input was X and the current input item is Y, and Y may be grammatically added to the phrase marker in the two ways shown in (6), then the structure with the shorter path from X to Y must be chosen (i.e. 6a).



The previous input and the current input are *phonetically adjacent* terminals. In many cases they are also syntactically adjacent terminals, but this is not always the case, because phonetically null terminals sometimes need to be postulated between the previous and the current input items. In these cases, Branch Right evaluates the paths between the each pair of syntactically adjacent terminals between the two phonetically adjacent inputs. To illustrate, consider the hypothetical situation. A is the last (overt) input item, and B is the current (overt) input item and each of two possible attachments of B requires postulating intermediate null heads, which are labelled e_1 , e_2 . The lengths of the paths required in each case are shown in (7).



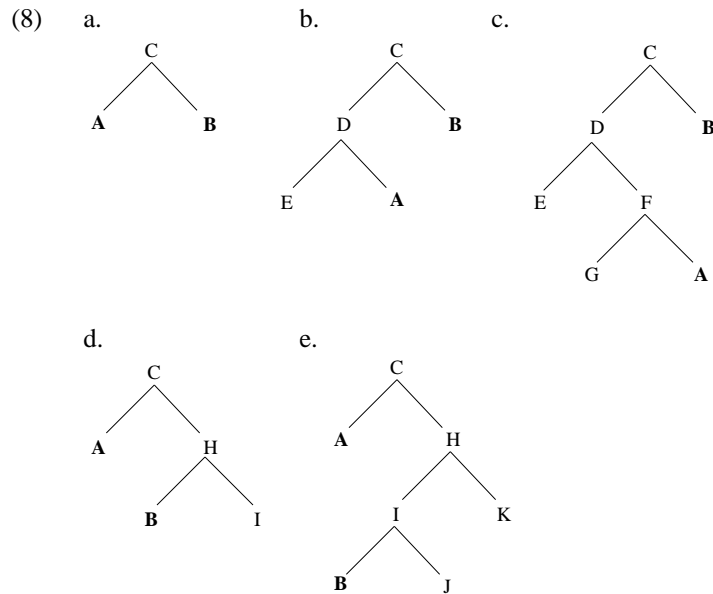
The attachment of B in (7b) is selected by Branch Right because it only requires no paths longer than 3 steps, whereas (7a) requires a path of length 4 steps. This point is important. Branch Right does *not* sum the total length of all paths between A and B (this would favor 7a), nor does it worry about the total number of terminals between A and B (if it did, this would presumably also favor 7a). All that matters is which analysis requires the longest single path between any pair of syntactically adjacent heads.

Note that this way of handling situations involving pairs of inputs that are phonetically but not syntactically adjacent is not arbitrary. The total number of heads in a structure has no bearing on how right-branching that structure is, and the distance between phonetically adjacent elements also has no bearing on how right-branching a syntactic structure is. It is the length of paths between syntactically adjacent heads that locally ensures precedence/c-command correspondences, and hence right-branching structures.

To see the effect of path length on precedence/c-command correspondences, consider the diagrams in (8). (8a) shows the shortest possible path between adjacent terminals A and B, consisting of two steps. Almost any way in which the path between A and B is lengthened leads to precedence/c-command

Parsing

mismatches. Adding an extra step in the upward path from A to C has the effect that nodes E and A precede but fail to c-command B (8b). Adding further steps to the upward path adds to the number of nodes which precede but do not c-command B (8c). (8d–e) shows the effects of lengthening the downward path from C to B. (8d) shows that adding one step to the downward path does not in fact create any precedence/c-command mismatches. This is the one instance where lengthening the path from A to B does not add precedence/c-command mismatch. (8e) shows that when the downward path is lengthened further, new precedence/c-command mismatches are created.



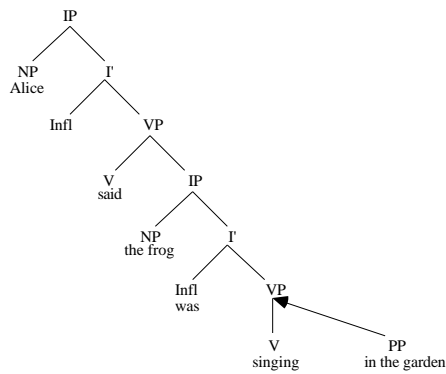
In the sections that follow I investigate the empirical coverage of Branch Right. But rather than embarking on a survey of well-known structural ambiguities at this point, a task which I delay until §3.3, the next section presents experimental evidence for the importance of local attachment effects in structural ambiguity resolution, based on a previously unstudied syntactic ambiguity.

3.2 On the Strength of the Local Attachment Preference⁶⁶

3.2.1 The Locality Puzzle

This section reports joint work with Ted Gibson which investigates the strength and the generality of the *Local Attachment Preference*. We use this term to refer in a theory-neutral way to whatever underlies the interpretive preference that speakers of English share when presented with a sentence like (9a). The first reading that becomes available is one in which the locative adverbial PP *in the garden* associates with the lower clause, such that it is the frog's singing that is happening in the garden, and not the announcement that Alice made about the frog. In other words, the adverbial is preferentially attached to local material in the tree rather than to non-local material. The Local Attachment Preference is a pervasive phenomenon in parsing, and some more well known cases of Local Attachment from Kimball 1973 are shown in (9c-e).

- (9) a. Alice said the frog was singing in the garden.
 b.



- c. # Joe looked the friend who had smashed his new car up.
 d. # I met the boy who Sam took to the park's friend.
 e. # The girl applied for the jobs that was attractive.

Our aim here is not to try to motivate the *existence* of the Local Attachment preference. We will take that for granted. Nor, for the most part, do we intend to choose among the many different versions of the Local Attachment preference that have been proposed.⁶⁷ What we are more concerned with here is the

⁶⁶ This section represents work conducted in collaboration with Ted Gibson, whom I thank for permission to include this material here. It is a minimally modified version of a draft of Phillips & Gibson 1996. Although the text of this section has intentionally not been seamlessly integrated with the rest of the chapter, the point of the section is quite relevant to the rest of the chapter.

⁶⁷ Among the family of proposals which we group under the heading Local Attachment Preference we include *Right Association* (Kimball 1973, Phillips 1995), *Late Closure* (Frazier 1978), *Recency* (Gibson 1991, Gibson et al. in press), *Attach*

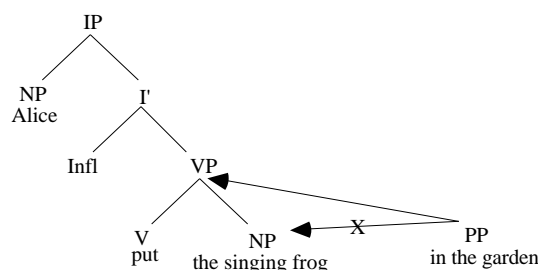
question of how the Local Attachment preference interacts with other parsing biases, and in particular the question of how *strong* the Local Attachment preference is relative to other proposed structural factors in parsing.

As we will show, in asking about the strength of the Local Attachment preference we are faced with what seems at first to be a straightforward contradiction. First, there is some long-standing and well-known evidence that the Local Attachment preference is a weak bias in ambiguity resolution, and that there are a number of other biases that can overrule its choices. To contrast with that we present the results of an experiment on a novel structural ambiguity which points to just the opposite conclusion, namely that the Local Attachment Preference is rather stronger than has usually been assumed. In the last part of the section we show how it is possible to reconcile what seem to be opposing findings.

The evidence that the Local Attachment Preference is a relatively weak bias comes from examples like (10). In (10a–b) we are again interested in the prepositional phrase *in the garden* and where it is preferentially attached. As in (9) there are two possible attachments, but the choices are slightly different in this case. The alternatives are to make the PP an argument of the verb or a modifier right-adjoined to the object NP. (10c–e) show further cases of V NP XP sequences in which XP could be attached to a projection of either the verb or the noun phrase.

Most versions of the Local Attachment Preference predict a preference for the NP-modifier reading of the PP *in the garden* in (10a), because this involves attachment to more recently built material. However, this prediction is incorrect, as both intuitions and a sizable body of experimental literature on the topic demonstrate (cf. Clifton, Frazier & Rayner 1983, Clifton Speer & Abney 1991). Speakers reliably prefer the VP-attached interpretation of (10a).

- (10) a. # Alice put the singing frog in the garden.
 b.



- c. # Joe bought the book for Susan to the party.
 d. # Alice saw the singing frog in the garden in the bathroom.
 e. # Henry told the intruder that he met to leave.

Low (Abney 1989), *Minimal Connections* (Fodor and Frazier 1983). These proposals differ with regard to whether locality is defined in terms of terminal strings, tree geometry or grammatical attachment sites. The differences among these various versions of the LAP are not important in this paper, except where noted.

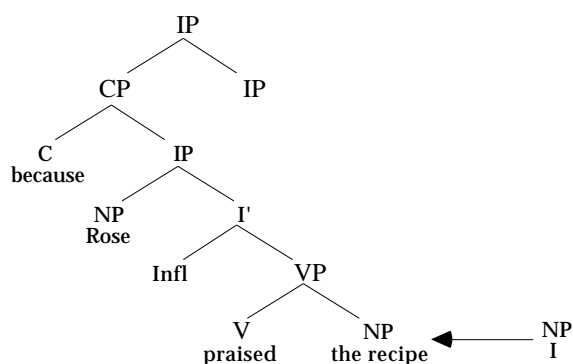
Parsing

The relevance of cases like those in (10) for the status of local attachment has long been recognized, and the conclusion that has typically been drawn from them is that there must be additional factors influencing the preferences speakers show in examples like (10a–e). Moreover, these other factors must be *stronger* than the Local Attachment Preference. The literature contains a number of proposals about what these other factors might be, including *Minimal Attachment* (Frazier 1978, 1987), argument attachment preferences (Ford et al. 1982; Pritchett 1988, 1992; Abney 1989; Gibson 1991; Schütze & Gibson 1996), discourse accommodation (Crain & Steedman 1985, Altmann & Steedman 1988, Percus 1995), frequency (Spivey-Knowlton & Sedivy 1995).

The other piece of the locality puzzle, which is the main focus of this section, is that there is a further ambiguity which leads us to the opposite conclusion from (10), namely that Local Attachment is *stronger* than the factors listed in the preceding paragraph. This is the ‘matrix-relative’ ambiguity shown in (11) (cf. Gibson & Broihier 1996). Given the incomplete sentence *because Rose praised the recipe I ...* there are a couple of possible continuations. The NP *I* could be made either the subject of a *that*-less relative clause modifying *the recipe*—this would mean that the clause that *I* is the subject of should contain an object gap, as in *because Rose praised the recipe I made for her birthday I also made it for her graduation*. Alternatively, *I* could be attached as the subject of the matrix clause, in which case the clause that *I* is the subject of should not contain an object gap, as in *because Rose praised the recipe I made it for her birthday*. Further examples of this ambiguity are given in (11c–d) and in the appendix.

Although almost all parsing principles that we are aware of predict that subjects should always opt for the matrix clause attachment, §3.2.3 shows evidence that subjects in fact opt for the relative clause attachment, which as far as we can tell *only* Local Attachment favors.

- (11) a. Because Rose praised the recipe I made ...
 i. ... for her birthday I also made it for her graduation. (rel. cl.)
 ii. ... it for her birthday. (matrix cl.)
 b.



Parsing

- c. Although the dissident eventually destroyed the letters he had kept
 - i. ... beneath the floorboards it pained him to do so. (rel. cl.)
 - ii. ... them beneath the floorboards for months. (matrix cl.)

- d. Since the students had enjoyed the classes they completed ...
 - i. ... with no trouble they encouraged their friends to take the same classes. (rel. cl.)
 - ii. ... them with no trouble and got good grades. (matrix cl.)

This preference leads us to believe that Local Attachment is a strong factor in ambiguity resolution. This conclusion clearly conflicts with what is usually concluded from the anti-locality effects shown in (10). In §3.2.4 we show how this conflict can be resolved.⁶⁸

3.2.2 Competing Biases

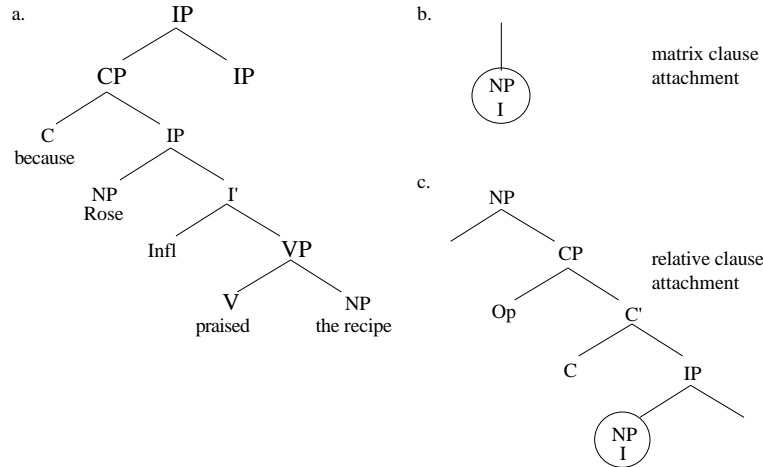
Consider now what different well-known parsing principles predict about how the matrix-relative ambiguity will be resolved. The choice that the parser has to make is as follows. The subject NP can be attached into the matrix clause, adding the structure shown in (12b). Alternatively, it can be made the subject of a relative clause modifying the object of the subordinate *because* clause (12c). In this case rather more structure needs to be built in order to attach the overt NP.

First, it should be clear from looking at the alternative chunks of structure required for each of the two possible continuations in (12) that Frazier's *Minimal Attachment* principle (Frazier 1978, 1987) should prefer the matrix clause attachment. This is because the relative clause attachment requires more new structure to be built than the matrix clause attachment, under any way of counting new structural material.

⁶⁸ See Cuetos & Mitchell 1988, Gibson et al. 1996 for evidence that Local Attachment is outranked by another factor in parsing ambiguous Spanish NPs of the form *the daughter of the colonel who I met last week*. Spanish speakers show a (weak) preference for attaching the relative clause to the first NP in examples like this. Gibson et al. 1996 attribute this preference to Predicate Proximity (see §3.2.4 below). We will have nothing to say about the English/Spanish contrast here.

Parsing

(12)



Principles that favor attaching arguments or obligatory constituents over optional constituents (cf. Ford et al. 1982, Abney 1989) should also choose the matrix clause attachment, since the matrix subject is an obligatory constituent in English, whereas a relative clause modifying the direct object of a subordinate clause is not at all obligatory.

Principles based on the idea that the parser prefers to leave as few predicate-argument relations unsatisfied as possible (e.g. Pritchett 1988, 1992; Gibson 1991) should also favor the matrix clause attachment. If the ambiguous NP is attached as the matrix subject, then just subject NP is lacking a theta-role assigner. If, on the other hand, the ambiguous NP is attached as the subject of an object relative clause, then two arguments will be lacking a theta-role assigner—both the subject of the relative clause and the null relativized object in SpecCP of the relative clause. Therefore, the matrix clause attachment is predicted to be preferred.

In addition, a selection relation holds between the *because* or *while* clause and the main clause. The simple fact that the subordinate clause requires the main clause in order to form a complete sentence could be putting additional pressure on the parser to pursue the matrix clause continuation.

Furthermore, any theory in which the parser chooses the continuation which entails the simplest accommodation of the current discourse model (Crain & Steedman 1985, Altmann & Steedman 1988, Percus 1995) the prediction is also that the matrix clause continuation should be chosen. This is because the relative clause implies the existence of some contrast set of recipes which has not yet been inserted into the discourse model. The matrix clause attachment, on the other hand, entails no such unsupported implicatures.

These predictions are summarized in the table in (13). As is immediately clear, the only principle that favors the relative clause continuation for the matrix-relative ambiguity is the Local Attachment Preference.

Parsing

(5) factor	preference	reason
Minimal Attachment	matrix	fewer nodes required to build matrix SpecIP than CP complex plus SpecIP in relative clause
Attach arguments/ obligatory constituents	matrix	Matrix subject is obligatory, relative clause is not
Complete predicate– argument relations	matrix	matrix attachment has 1 argument missing a predicate; relative clause attachment has 2 arguments missing a predicate
Semantic completeness	matrix	<i>because</i> or <i>while</i> clause needs an antecedent matrix clause; relative clause does nothing for this
Discourse Accommodation	matrix	In null context there's no contrast set for <i>the recipe</i> , so no pressure to modify
Tense Matching Constraint (see below)	matrix	Increased matrix clause activation due to tense-parallel imposed by <i>while</i> , <i>after</i> , <i>until</i> clauses
Local Attachment, Right Association, Recency	relative	closer association to most recently attached lexical material

Thus, all parsing principles that make any choice about how the matrix-relative ambiguity in (11) should be resolved choose the matrix clause attachment, except for local attachment. The fact that *only* Local Attachment favors the relative clause continuation is useful, because it means that if we find any evidence for a relative clause preference in resolving this ambiguity, then we may conclude that local attachment and none of the other factors listed is responsible for this preference. It would also show fairly unequivocally that Local Attachment is stronger than all of the other parsing principles listed in (13), because that would be the only way of explaining how it could override the conflicting preferences that the principles in (13) predict.

If, on the other hand, we find that subjects show a matrix clause attachment preference, this is not very informative. Given that there are so many different reasons to prefer the matrix clause attachment over the relative clause attachment, it would be hard to know which factor(s) are responsible for the parser's choice of the matrix clause attachment.

3.2.3 Experimental Evidence

We conducted an experiment to test the intuitive preference for relative clause attachment in the matrix-relative ambiguity. The predictions of local attachment and other proposed parsing strategies were already discussed in §3.2.2.

Subjects. Forty-seven native English speakers from MIT (primarily undergraduate students) participated, for \$8.00 each.

Materials. There were 4 conditions: matrix clause and relative clause resolutions, with ambiguous and unambiguous versions of each. The difference between the matrix and relative clause conditions was that the relative clause conditions contained an object gap where the matrix clause conditions contained an overt pronoun. In the unambiguous relative clause condition the relative clause was introduced by the complementizer *that*. In the unambiguous matrix clause condition the direct object NP in the subordinate clause contained a possessor, e.g., *the recipe* was replaced by *my recipe*. The relevance of this is that possessors block relativization, so that when the following (ambiguous) NP is encountered the parser knows that it must be the matrix subject, because the relative clause continuation is not a possible option.⁶⁹

- (14) a. *Relative clause, ambiguous (RA)*
Because Rose praised the recipe I made for her birthday it was worth all the effort.
- b. *Relative clause, unambiguous (RU)*
Because Rose praised the recipe that I made for her birthday it was worth all the effort.
- c. *Matrix clause, ambiguous (MA)*
Because Rose praised the recipe I made it for her birthday as a surprise.
- d. *Matrix clause, unambiguous (MU)*
Because Rose praised my recipe I made it for her birthday as a surprise.

Two additional properties varied in our items. Half of the items described a *non-temporal* relation, using complementizers like *because* (cf. 14), *since* or *although*, and the other half described a *temporal* relation, using complementizers like *while*, *when* or *after*. In addition, the ambiguous NP was a pronoun in 8 of the items and a full NP in the 16 remaining items (e.g., *While I talked with the lawyer John was watching (him) at the party ...*). A complete list of materials is given in the appendix.

We should point out that although the factor \pm temporal was balanced in our materials, it was not varied independently of the factor \pm pronominal. All 8 of the items in which the ambiguous subject NP was a pronoun occurred in items describing a non-temporal relation. This fact becomes relevant below.

Twenty-four items each with four forms like those shown in (14) were constructed. The 24 experimental items were combined with 90 fillers to form four lists. The fillers were of approximately the same length and complexity as the experimental items. The experimental items were counterbalanced across the lists so that each list contained six items from each condition and exactly one

⁶⁹ It is not strictly true that NPs containing a possessor do not allow relativization (thanks to Gregory Ward for pointing this out). However, the combination of the fact that relative clauses following NPs containing possessors are generally interpreted as non-restrictive and the fact that *that*-less relative clauses must receive a restrictive reading means that relative clause readings of examples like our (14d) should be either impossible or extremely marked. Furthermore, the results of the experiment imply that the (d) condition was indeed an effective disambiguator.

Parsing

version of every item. Ten practice items were also constructed to be similar to the fillers.

Procedure. The stimuli were presented on a Macintosh Centris computer using a word-by-word self-paced moving-window paradigm (Just, Carpenter & Wooley 1982). A subject saw two screens of instructions, followed by 10 practice items and then the 114 experimental and filler items. The experimental and filler items were presented in a different random order for each subject. Each sentence was followed by a question, to ensure that subjects had understood the sentences. All trials on which the question was incorrectly answered were excluded prior to further analysis. One subject was excluded because of an extremely high error rate in the questions.

Results. A regression equation predicting reading time from word length was constructed for each subject, using all items (filler and experimental). At each word, the reading time predicted by the subject's regression equation was subtracted from the actual measured reading time, and all analyses were performed on these differences (residual reading times). This transformation removes extraneous variance by subtracting out a baseline for each subject, and by controlling for noise due to length effects (Ferreira & Clifton 1986; Trueswell & Tanenhaus 1991).

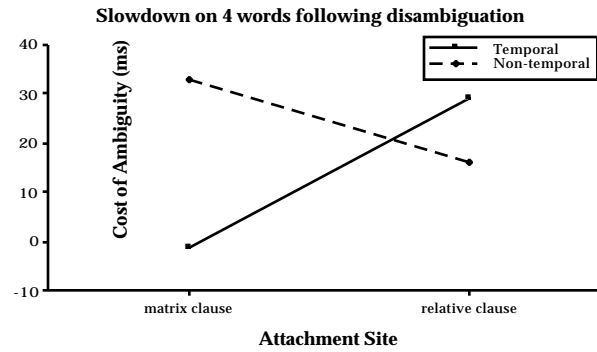
Trials on which the subject answered the comprehension question incorrectly were excluded from all of the following analyses. All trials with residual reading times greater than 1000 msec were also excluded prior to analyses. This affected less than 1% of the trials.

No differences were found preceding the point of disambiguation (object pronoun in matrix clause conditions, PP in relative clause conditions).

On the 4 word region beginning at the point of disambiguation there were main effects of ambiguity, attachment site and connective (i.e. \pm temporal). These results are shown in the table in (15b). There was also an ambiguity x attachment site x connective interaction, shown in (15a), which we focus on in what follows. The non-temporal examples show evidence for a relative clause attachment preference, whereas the temporal examples show evidence for a matrix clause attachment preference.

Parsing

(15a)



ambiguity x site x connective interaction
 $F(1,45) = 4.99, p < 0.05. F(2,1,11) = 3.01, p < 0.1$

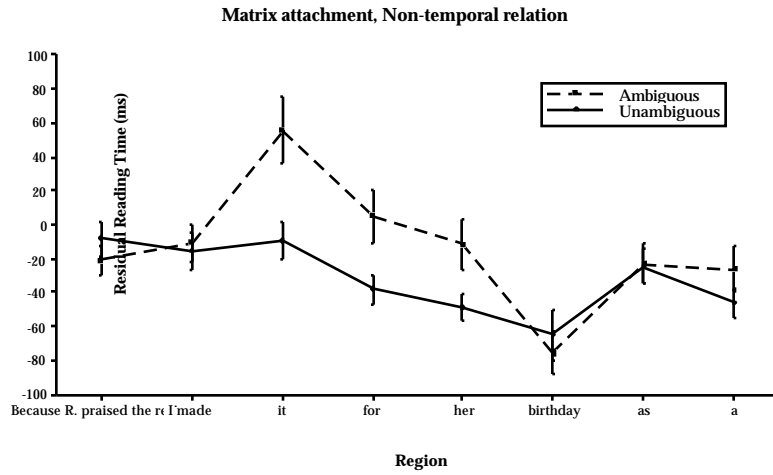
b.

ambig = 1.30ms	unamb = -18.44ms	$F(1,45) = 8.15$	$F(2,1,11) = 10.2$
		$p < 0.01$	$p < 0.005$
matrix = -19.55ms	relative = 2.42ms	$F(1,45) = 12.3$	$F(2,1,11) = 7.06$
		$p = 0.001$	$p < 0.05$
temp = -1.58ms	nontemp = -15.55ms	$F(1,45) = 6.06$	$F(2,1,11) = 2.08$
		$p < 0.05$	$p = 0.16$

(16) compares residual reading times for ambiguous and unambiguous versions of the matrix clause conditions with the non-temporal items. There is a highly significant slowdown in the ambiguous condition which begins as soon as the disambiguating pronoun is read, and extends onto the first two words of the following prepositional phrase. This is a classic ‘filled gap’ effect (cf. Stowe 1986). No other regions showed significant effects of ambiguity. In the relative clause conditions (17) the non-temporal items show a small slowdown at the PP following the gap, but it is not significant on any individual word or the PP as a whole, in clear contrast to the matrix clause conditions.

Parsing

(16)



Because Rose praised the recipe / I made / it / for / her / birthday / as / a ...

it ambiguous = 55.45ms, unambiguous = -9.57ms
 F1(1,45) = 8.71, p = 0.005. F2(1,11) = 8.30, p = 0.015

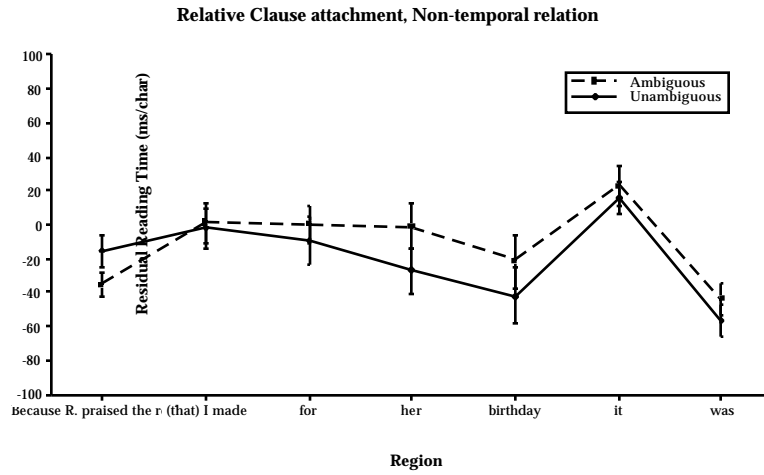
for ambiguous = 4.78ms, unambiguous = -37.7ms
 F1(1,45) = 5.85, p = 0.02. F2(1,11) = 7.98, p = 0.017

her ambiguous = -11.78ms, unambiguous = -48.38ms
 F1(1,45) = 5.24, p < 0.05. F2(1,11) = 7.81, p = 0.017

it for her ambiguous = 16.24ms, unambiguous = -31.48ms
 F1(1,45) = 12.1, p = 0.001. F2(1,11) = 11.55, p = 0.006

Parsing

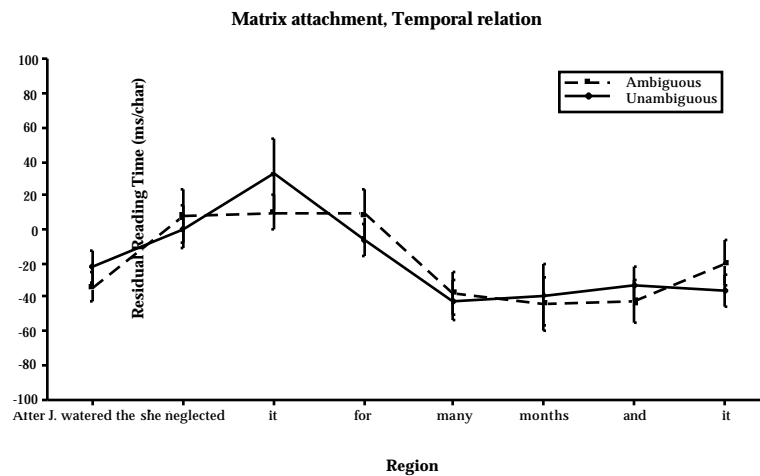
(17)



Because Rose praised the recipe / I made / for / her / birthday / it / was ...

The items describing a temporal relation show the opposite pattern of results (18–19). The matrix clause conditions show no effect of ambiguity.

(18)

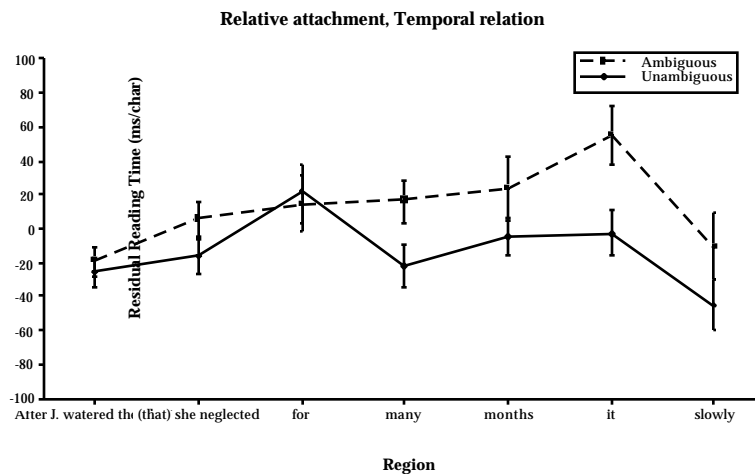


After Julie watered the plant / she neglected / it / for / many / months / it / ...

Parsing

In the relative clause conditions, however, we find a strong garden path effect, starting at the second word of the disambiguating PP and extending for a region of 3 or 4 words. This garden path effect is not as immediate as in the non-temporal matrix clause conditions, where it occurred immediately at the disambiguating pronoun. This slight delay is not surprising, though, given that disambiguation in these conditions occurs when the subject notices the *absence* of an overt object NP in the relative clause conditions, whereas it only requires noticing the *presence* of an overt object NP in the matrix clause conditions.

(19)



After Julie watered the plant / she neglected / for / many / months / it / slowly

many	ambiguous = 16.72ms, unambiguous = -21.73ms F1(1,45) = 4.57, p < 0.05. F2(1,11) = 6.41, p < 0.05
months	ambiguous = 23.84ms, unambiguous = -4.82ms F2 (1,45) = 1.4, p < 0.3. F2(1,11) = 0.49, p = 0.5
it	ambiguous = 55.71ms, unambiguous = -2.64ms F1(1,45) = 7.82, p < 0.01. F2 (1,11) = 9.84, p < 0.01

At the first word following the PP there was a marginally significant effect of ambiguity: ambig = 3.33ms, unambig = -11.0ms; F1(1,45) = 3.44, p < 0.1. There was a highly significant effect of attachment site: matrix = -30.51, relative = 22.85; F1(1,45) = 25.66, p < 0.001. There was also an ambiguity x attachment site interaction: F1(1,45) = 6.28, p < 0.05. At the following word (i.e. second word after PP) there was a main effect of ambiguity only: ambig = -24.97, unambig = -45.47; F1(1,45) = 6.56, p < 0.05.

Discussion. Most importantly, the fact that in the non-temporal conditions we find a preference for the relative clause attachment confirms that Local Attachment is an extremely strong structural factor in parsing. It also shows

Parsing

that Local Attachment is strong enough to override the opposing forces to pursue the matrix clause attachment that the other principles listed in (13) predict. As we pointed out above, we are unaware of biases other than Local Attachment which would lead speakers to pursue the relative clause attachment.

A comment is in order at this point on the issue of punctuation. Our stimuli were presented with no punctuation, and it has been suggested to us that this may be a confounding factor in our results. The objection usually goes as follows: in written text the matrix clause attachments would often be disambiguated by means of a comma. Therefore the absence of a comma in these stimuli could be responsible for leading people to initially pursue the relative clause attachment. There are a couple of reasons why we think this is an unlikely explanation for the results.

First, the split in our results between temporal and non-temporal items shows that speakers are able to pursue the matrix clause attachment in the absence of a comma. In written text the subordinate clause and the matrix clause are often not separated by a comma. Second, although we agree that the role of a comma when it occurs in examples like these is to aid comprehension, it is important to ask why the comma should ever be necessary in these cases. The answer is presumably that the comma is used to counteract the local attachment preference for the relative clause parse. Therefore, the fact that commas are sometimes used in examples like ours may confirm rather than confound our claim that there is a local attachment preference in these sentences (cf. Frazier & Rayner 1982 for a similar argument).

We owe an explanation for why the contrast between the temporal and the non-temporal items should lead to such a clear difference in parsing preferences. As pointed out above, the factors \pm temporal and \pm pronominal were not independent in our materials, so further experiments will be required in order to determine whether one or both of these factors is responsible for the split in the results. We can, however, offer the following suggestions.

Tense Matching. If the factor \pm temporal is responsible for the split in the results, then this may be due to the fact that the temporal items observe a tense matching requirement between the subordinate and the matrix clauses, but the non-temporal items do not.

As (20a) and (20b) show, when the two clauses describe a temporal relation, the choice of tense in the first clause tightly constrains the choice of tense in the second clause. In (20a) for example, the subordinate clause *while John was eating his lunch*, which contains a past tense verb, can be followed by a past tense matrix clause like *he was watching TV*. But if we replace this with a present tense or future tense main clause the sentence becomes ungrammatical. Sentences describing temporal relations are therefore subject to a Tense Matching Constraint.

- (20) a. While John was eating his lunch ...
- | | |
|------------------------------|-----------|
| ... he was watching TV | past |
| * ... he watches TV | * present |
| * ... he will be watching TV | * future |

Parsing

- b. After Mary gets off the bus on Mondays ...
- | | | | |
|---|----------------------------------|---|---------|
| * | ... she bought a cup of coffee | * | past |
| | ... she buys a cup of coffee | | present |
| * | ... she will buy a cup of coffee | * | future |

In (20c) and (20d), though, which give examples of non-temporal sentences, no Tense Matching Constraint is operative. As long as general constraints on tense sequencing in English are respected, more or less any tense can combine with more or any other tense. In (20d) for example, the subordinate clause *although Helen drives a microbus* can be followed by past, present, or future main clauses with no difficulty.

- (20) c. Because John was eating a greasy sandwich at his desk ...
- | | | | |
|--|---|--|---------|
| | ... he was in trouble with the boss | | past |
| | ... he is in trouble with the boss | | present |
| | ... he will be in trouble with the boss | | future |
- d. Although Helen drives a VW microbus ...
- | | | | |
|--|----------------------------------|--|---------|
| | ... she used to own a Cadillac | | past |
| | ... she also owns a Cadillac | | present |
| | ... she will soon own a Cadillac | | future |

To see why the Tense Matching Constraint should make a difference to the resolution of the matrix-relative ambiguity, imagine that the parser is working its way through a subordinate clause headed by a temporal complementizer like *while*. We assume that the Tense Matching Constraint is immediately active during parsing. So as soon as the complementizer *while* is encountered the parser knows to expect both a subordinate clause and a matrix clause and it also knows that the two clauses will match in tense. Therefore, as soon as the tensed verb is reached, the parser is immediately in a position to build tense features into the Infl node of the matrix clause, and so these features are built right away.

The relevance of this is that by working on building the matrix inflection the matrix IP projection is activated to a certain degree. The effect of this when the parser encounters the ambiguous NP is to add to the other factors that are lobbying for a matrix clause attachment, and this turns out to be just enough to override the strong pull for relative clause attachment that comes from the Local Attachment preference. In the non-temporal cases, on the other hand, the absence of the Tense Matching Constraint means that Local Attachment is able to override any biases towards matrix clause attachment.

Pronominals vs. Full NPs. Another possible reason for the split in our results involves the contrast between sentences in which the ambiguous NP is a pronoun and sentences in which it is a full NP (e.g. *the janitor*). Recall that 8 of the 12 sentences describing a non-temporal relation (relative clause preference) had a pronoun as the ambiguous NP, and all 12 of the sentences describing a temporal relation had a full NP as the ambiguous NP (matrix clause preference). The pronouns always had an antecedent in the subordinate clause, and the full NPs always introduced novel discourse referents.

The full NPs may be preferentially attached in the matrix clause because of discourse factors. All of the ambiguous relative clauses lack a relative pronoun,

and are therefore restrictive relative clauses. In a felicitous discourse the content of a restrictive relative clause is given information rather than new information. But the fact that the full NPs pick out novel discourse referents entails that the relative clauses cannot contain given information. This may provide sufficient pressure to avoid the local attachment and pursue the matrix clause attachment. In the examples with ambiguous pronouns, on the other hand, the first word of the relative clause picks out an existing discourse reference, so the same ‘discourse penalty’ is not incurred if the relative clause attachment is pursued.

This gives just a sketch of why the factors \pm temporal or \pm pronominal could lead to the bifurcation that we observe in our results. Further experiments are underway which attempt to separate these two factors, by independently manipulating the factors \pm pronominal and \pm temporal.

3.2.4 Resolving the Locality Puzzle

We are now in a position to return to the Locality Puzzle that we began with. Recall what the problem is: the results of our experiment lead to the conclusion that the Local Attachment preference is strong, and can override a variety of other potential factors in parsing. But this has to be reconciled with the anti-locality effects that we pointed out at the beginning, and that are motivated by a sizable body of work. The key evidence is summarized in (21).

- (21) a. # Because Rose liked the recipe I made it for her birthday.
 (Local Attachment > other factors in (13))
 b. # Alice put the singing frog in the garden in the circus.
 (Local Attachment < other factors in (13))

We can see a couple of ways of resolving the Locality Puzzle, each of which we have pursued elsewhere. The first approach focuses on the fact that the Locality Puzzle only arises if it is assumed that parsing biases are ranked, so that one factor will always win out over another. If the strict hierarchy view is dropped and a series of weaker factors is allowed to ‘gang up on’ a stronger factor, then it is fairly easy to solve the Locality Puzzle. The second approach questions the assumption that the VP-attachment preferences that are taken to show anti-locality effects are really anti-locality effects. Reexamination of the syntax of these examples shows that they are actually consistent with a Local Attachment preference. If this is the case, then the Locality Puzzle does not arise either.

(22–24) show how the puzzle is resolved under an approach in which different structural constraints are weighted and can conspire (cf. Gibson 1991; Gibson, Pearlmuter, Hickok & Canseco Gonzalez 1996; Gibson & Broihier 1996). The relevant constraints are shown in (22). The strongest constraint, Recency, favors attachments to recently built structure; the next strongest constraint places a cost on parses in which there are arguments whose theta-role assigner has not yet been determined; finally Predicate Proximity favors attachments to projections of verbs and other predicates.

Parsing

- (22) a. RECENCY favors attachments to projections of recent items.
b. LOCAL THETA VIOLATIONS incurred by arguments lacking a theta-role.
c. PREDICATE PROXIMITY favors attachment as structurally close to a predicate as possible.

In the matrix-relative ambiguity shown in (23) the relative clause continuation is chosen, because one recency violation and one local theta violation is worse than two local theta violations. Neither representation is associated with a predicate proximity violation, because the attachment of the subject NP *I* is equally close to a predicate in each. In both cases, this NP is the subject of a predicate VP to come.

- (23) # Because Rose liked the recipe I made it for her birthday.
- | | |
|------------------------------|--|
| + Relative clause attachment | 0 recency violations
2 local theta violations
(subject & relativized object)
0 predicate proximity violations |
| Matrix clause attachment | 1 recency violation
1 local theta violation
0 predicate proximity violations |

It follows from other results that a recency violation is associated with greater cost than is a theta violation (Gibson, 1991).⁷⁰ Thus the preferred attachment according to this theory is the more local relative clause attachment, in spite of its two theta violations.

Turning now to the anti-locality effect repeated in (24), notice that if we consider only recency and local theta violations this theory incorrectly predicts that the PP will be attached to the NP rather than to the verb. This is because the matrix-relative ambiguity showed that recency violations must incur a greater

⁷⁰ Specifically, a difference of a single recency violation is enough to cause a strong preference for the more local attachment, as in (i):

- (i) The teaching assistant told the professor that the students were confused during the class.

The PP *during the class* can ambiguously attach to either the matrix IP or the embedded IP. There is a single recency violation difference between the two attachments, but there are no other cost differences. The preference for low attachment is strong enough that semantic factors cannot override the attachment preference initially, as evidenced by the difficulty associated with the processing of (ii):

- (ii) #The teaching assistant told the professor that the students will be confused yesterday during the class.

In contrast, a difference of a single theta violation does not cause as such a strong preference that semantic and pragmatic factors cannot override the thematic preference. See Gibson (1991) and Gibson, Hickok & Schütze (1994) for more details.

Parsing

cost than local theta violations. However, this is one of the situations where Predicate Proximity plays a role: by conspiring with the theta-attachment constraint it is able to override the strength of the Recency preference.

(24) # Alice put the singing frog in the garden in the circus.

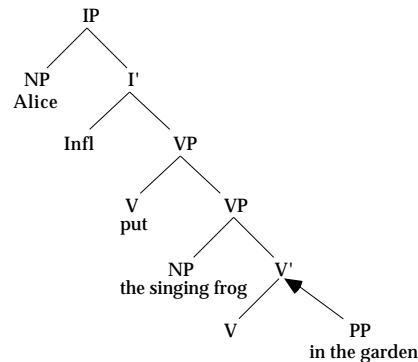
Modifier attachment	0 recency violations 1 local theta violation 1 predicate proximity violation
+ Argument attachment	1 recency violation 0 local theta violations 0 predicate proximity violations

Under this view, then, the Locality Puzzle involves a genuine conflict, but such conflicts are the hallmark of this approach.

The second approach to the Locality Puzzle, argued for in the remainder of this chapter (building in part on Fodor & Frazier 1983), claims that there is really no conflict between the locality and anti-locality effects. What has been thought of as an anti-locality effect is in fact entirely consistent with Local Attachment.

A body of work on constituent structure from the last 10 years argues that VP-modifier phrases and the second argument in double complement structures are actually sisters of the verb in a right-branching VP structure like (25) (cf. Larson 1988, 1990; Stroik 1990; Pesetsky 1995).⁷¹

(25)



These syntactic results become relevant if the Local Attachment is instantiated by the principle of *Branch Right* (cf. §3.1 above), a principle which

⁷¹ The evidence for this involves the results of tests of coordination, binding, idiom-formation, polarity item licensing etc., in conjunction with the assumption that these tests transparently diagnose constituency or c-command relations. See Chapter 2 for justification of this assumption, which has often been challenged in the literature.

favors construction of right-branching structures. As (26) and (b) show, both the matrix-relative ambiguity and the V-NP-PP ambiguities are resolved in favor of the more right-branching alternative, and therefore conform to the predictions of Branch Right.

- (26) a. # Because Rose liked the recipe I made it for her birthday.
- | | | |
|---|----------------------------|-----------------------------------|
| + | Relative clause attachment | more right-branching continuation |
| | Matrix clause attachment | less right-branching continuation |
- b. # Alice put the singing frog in the garden in the circus.
- | | | |
|---|---------------------|---|
| | Modifier attachment | less right-branching continuation
(adjunction to NP) |
| + | Argument attachment | more right-branching continuation
(sister of V) |

We do not intend to choose between these two alternatives here. Nor is the choice among them important from the perspective of this particular paper, because our main points, which are consistent with both alternatives, are that the Local Attachment Preference is a strong structural factor in parsing and that biases must be weighted rather than ranked.

3.2.5 Interim Conclusions

Based on our experiment on the matrix–relative ambiguity, we hope to have shown that the Local Attachment preference is stronger than has often been assumed. In particular, its choices are able to override all of the other potential factors listed in (13) above.

Our results showed a split in attachment preferences between sentences describing temporal relations and sentences describing non-temporal relations. We suggested an explanation for this split, based on the effects of a Tense Matching Constraint, and an alternative explanation based on whether the ambiguous word is a pronoun or a full NP.

Although our findings regarding the strength of the Local Attachment Preference may appear to conflict with a body of well-known evidence that Local Attachment is not so strong, we have shown a couple of ways in which the ‘anti-locality’ results can be reconciled with our results.

The sections that follow show how the local attachment preference, in the form of Branch Right, accounts for a broader range of structural ambiguity resolution preferences than has normally been assumed.

3.3 Ambiguity resolution in English

The previous section gave evidence that a local attachment preference is not only a necessary component of ambiguity resolution, it is also much stronger than has generally been assumed. In the sections that follow I aim to show in addition that the local attachment preference also has a much broader empirical coverage than is usually assumed. The effect of this is to make a number of other structural parsing principles redundant.

3.3.1 Two types of structural ambiguity

In this section I survey a number of different structural ambiguities in English and show how Branch Right accounts for the way in which they are preferentially parsed, highlighting the structural properties that different kinds of ambiguities share.

For the purposes of this section I will mostly be taking the existence of structural preferences in parsing for granted, and not addressing the various recent claims that there is no structural component to ambiguity resolution. I delay discussion of this issue until §3.6. This is not because I consider this issue to be minor or irrelevant, but because I am defending a different structural generalization (Branch Right) from the one that has most commonly been criticized (Minimal Attachment), and should therefore first spell-out the generalization and what it predicts.

It will be useful to separate the different structural ambiguities that have been studied in English into different groups, according to the kind of choice that the parser is faced with in each case. The first kind of ambiguity is a situation in which the parser may form a grammatical dependency (argumenthood, modification, agreement etc.) between the input word and a preceding word, or it may choose to not establish any direct dependency.⁷² I refer to these as *(In)dependence ambiguities*. Some examples are given in (27). As is clear from the table, in these situations, the parser always chooses to establish a dependency in preference to the absence of one. Such a generalization is fairly easy to capture in a number of different approaches, and therefore these ambiguities will not require much discussion.

(27) (In)dependence ambiguities

Choice	Example	Dependency formed?
direct obj. vs. embedded subj.	John knew the answer to the problem was wrong.	yes
direct obj. vs. matrix subj.	While Mary was mending the socks fell off her lap.	yes
rel. cl. on obj. vs. matrix subj.	Because the conference organizers liked the caterers they hired them for the banquet.	yes
embedded subj. vs. matrix subj.	Whenever Bill started to complain the class was boring as a result.	yes

On the other hand, a second class of ambiguities in English requires of the parser that it choose between two possible dependencies with existing elements in the phrase marker. I will refer to these as *Association ambiguities*. Some examples are given in (28), together with which element the incoming element is reported to preferentially associate to. ‘1’ refers to the attachment site that hierarchically dominates the other attachment site (‘2’).

⁷² Alternatively, it may choose between a direct or an indirect dependency with an existing element (e.g. complement vs. subject of complement).

Parsing

(28) Association ambiguities

1	2	Example	Preferred associate
VP	NP	John gave the man the dog bit a package.	1
VP	NP	The spy saw the cop with a telescope.	1
VP	NP	The patient persuaded the doctor that he was having trouble with to leave.	1
IP	NP	The horse raced past the barn fell.	1
NP1	NP2	I know the daughter of the colonel who Mary met	2
NP	N	The cotton clothing is made of grows in Mississippi.	2
IP	N	The desert trains young people to be very tough.	2
VP1	VP2	John said Bill left yesterday	2
VP	Adj	I gave her hamsters to my cousin.	2
VP	N	John expressed his interest in a hurry.	2

As table (28) shows, there are a number of different association ambiguities in English, and the preferred associate cannot be reliably predicted from which of the sites is higher or lower in the tree at the point at which the ambiguity occurs.

In what follows, I aim to show that the manner in which all of these ambiguities is predominantly resolved is predicted by Branch Right. I begin in §3.3.2 with the cases of association ambiguities in which the preferred associate is the lower position, according to table (28), and then in §3.3.3 move on to the remaining association ambiguities, which mostly have being the second argument of a verb as one of the alternative structures. I discuss these examples in the light of the analysis of complex VPs given in Chapter 2. Finally in §3.3.4 I discuss the (in)dependence ambiguities.

3.3.2 Association ambiguities I

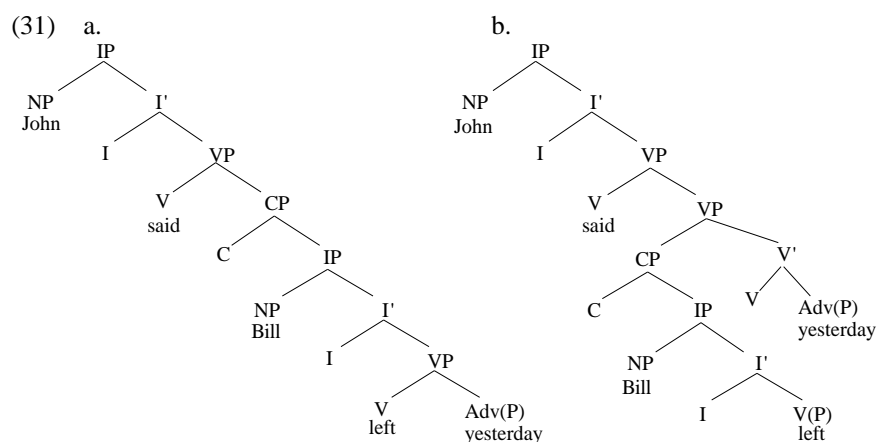
‘Traditional’ local attachment effects. The simplest examples of association ambiguities are the cases which have traditionally been attributed to locality principles like Right Association (Kimball 1973) and Late Closure (Frazier 1978). Some examples are shown in (29–39). (29) shows cases (based on examples in Kimball 1973) in which an element can associate to either a matrix or an embedded VP. There is a strong preference to associate to the embedded verb in these examples. (30) shows cases in which a relative clause can attach to one of two preceding NPs. Here there is also a preference to attach to the more recent NP, but this preference is weaker than the preference for examples like (29) (cf. Cuetos & Mitchell 1988, Gilboy et al. 1995, Ganger 1996).⁷³

⁷³ In a number of other languages (e.g. Spanish, Dutch) ambiguities like (30) are predominantly resolved in favor of the higher attachment site (cf. Cuetos & Mitchell 1988, Gilboy et al. 1995, Gibson et al. 1996, Brysbaert & Mitchell 1995). See §3.5 for further discussion.

Parsing

- (29) a. John said Bill left yesterday.
 b. Joe figured that Sue wanted to take the train to New York out.
 c. Sarah saw the book that you were reading in the library.
- (30) a. I met the daughter of the colonel who had had the accident.
 b. I spoke to the butler of the countess who had placed the emergency call.

(31) shows how the low attachment preference follows from Branch Right, using (29a) as an illustration. The path between the last word (*left*) and the incoming word (*yesterday*) is much shorter when the adverb is attached in the embedded clause.⁷⁴



low attachment:
 2: V-VP-Adv(P)

high attachment:
 5: V(P)-I'-IP-CP-VP-V
 2: V-V'-Adv

Examples (32a–b) show temporary ambiguities for which preferences have standardly been attributed to the local attachment preference. The sentences in (32–33) involve ambiguous scope of the conjunction *and*, for which the intuitive preference is always to favor the low attaching, narrow scope for *and*. In (32) the NP *her sister* is initially construed part of a conjoined object NP rather than the subject of a conjoined sentence (cf. Frazier 1978).

- (32) John kissed Mary and her sister laughed.

⁷⁴ Note that although I assume the right-branching VP structures of Chapter 2 (and the many other references cited there) in this chapter, most of the results would follow equally well under the more traditional assumption that complex VPs are n-ary branching structures in which all arguments and adjuncts of the verb are daughters of the VP node.

Parsing

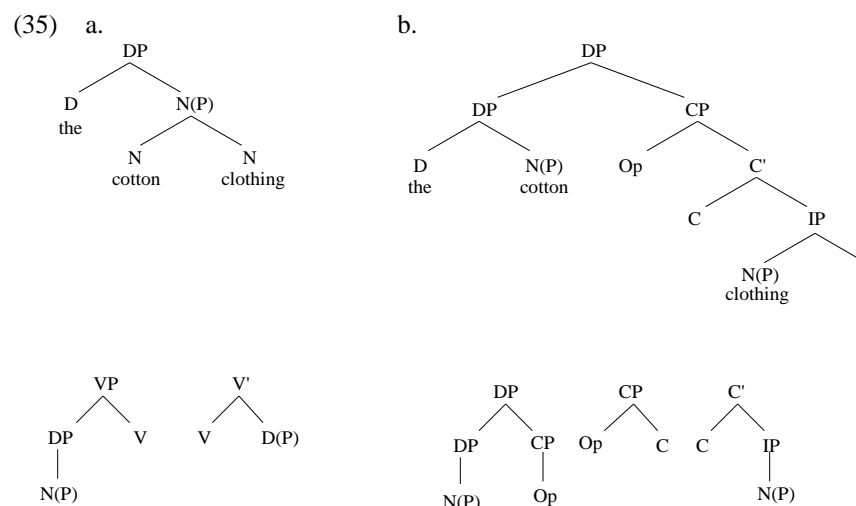
In (33) the parser must choose between an N-coordination and an NP-coordination analysis. If the NP-coordination is chosen, then the second noun *doctors* will be interpreted as a bare plural, and the relative clause will only restrict this noun. If, on the other hand, the N-coordination is chosen, then *doctors* will not be interpreted as a bare plural, but the relative clause will restrict both nouns.

- (33) The janitors and doctors who supported the strike picketed in large numbers.

There is an intuitive preference to interpret the relative clause as restricting both nouns, indicating a preference for the low-attaching N-coordination.

N-N compound–relative clause ambiguity. The examples in (34) show an ambiguity in the analysis of N–N sequences, which may be analyzed as N-N compounds or as a head noun followed by the subject of a relative clause (Marcus 1980). Disambiguation in favor of the relative clause analysis leads to a strong garden path effect. (35) shows the relative cost of each attachment according to Branch Right. The partial tree resulting from each attachment is shown first, followed by the individual subpaths required in order to make each attachment. These have been added for ease of exposition, because the path lengths cannot always be read directly from the output tree. This is because of the assumption that there are no vacuously projecting nodes, only branching nodes, which means that some nodes which project in the output structure did not need to project in order to be initially attached.

- (34) a. # The cotton clothing is made of grows in Mississippi.
 b. # The construction linguists work in is scheduled to be pulled down.
 c. # The cable service users ask for most is CNN.



Parsing

compound attachment:
2: N–N–N

relative clause attachment:
4: N(P)-DP-DP-CP-Op
2: Op-CP-C
3: C–C'–IP–N(P)

Possessor–goal argument ambiguity. The examples in (36) show an advantage of not explicitly building an argument attachment preference into a parsing theory. In these examples, the words following the ditransitive verb may be interpreted either as a single argument containing a possessor, or as both arguments of a double object construction. To the extent that speakers' intuitions show a preference for one of these two interpretations of *her hamsters*, they slightly favour the single argument interpretation in (37b) (cf. Gibson 1991, Kennison & Clifton 1994).^{75,76}

- (36) a. I gave her hamsters on her birthday.
b. I gave her hamsters to Mary.

⁷⁵ I am currently uncertain of the relative frequency of the two readings of *her*. This is hard to determine from corpora, because of the homophony. However, it is possible to make estimates based on non-homophonous pairs like *him/his* and *them/their*. *His* occurs more than twice as often as *him* in the Francis & Kucera 1982 corpus. *Their* is also more frequent than *them*, but the contrast is much smaller than with the masculine singular pronouns.

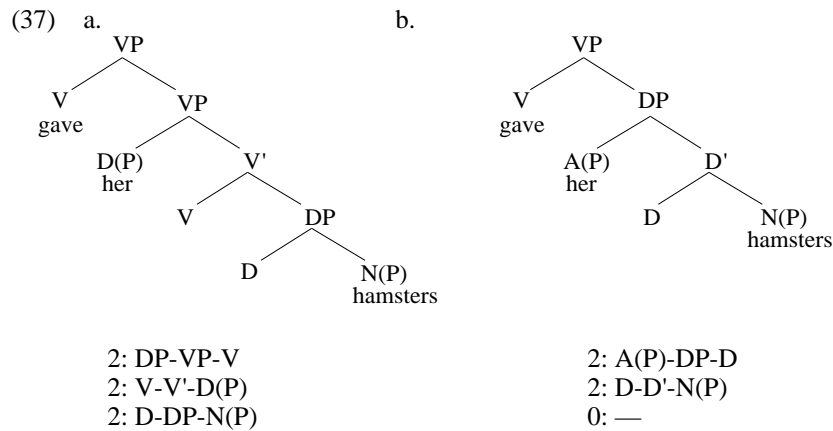
Therefore, it is likely that there is a frequency advantage for the genitive reading of *her* over the accusative reading. However, this does not automatically translate into a frequency bias for the genitive reading of *her* when it occurs in object position. Genitive *her* can occur in NPs in subject, object and adjunct positions. Therefore it is likely that when the form *her* appears in a potential object position it is more frequently accusative than genitive. It is unclear what kind of frequency predictions are most appropriate in this case.

⁷⁶ There is a clearer preference for the single argument interpretation when *give* in (36) is replaced by an optionally ditransitive verb like *find* or *carry*, as Frazier (1994) points out.

- (i) I found her hamsters.
(ii) I carried her shopping.

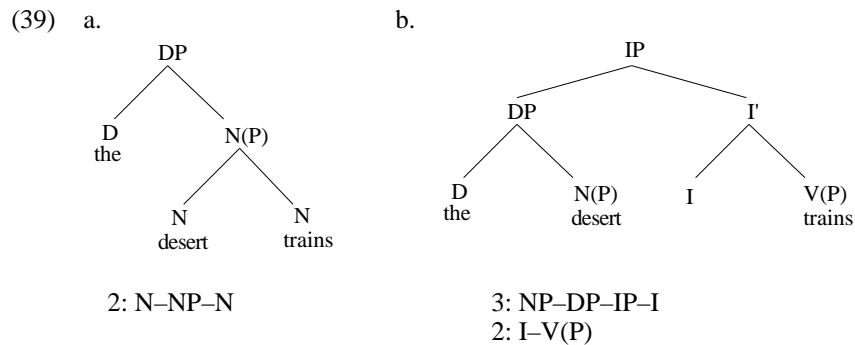
This increased preference for the single argument reading may be due either to the parser's avoidance of benefactive roles, or to the fact that in the double object construction reading of (i–ii) the obligatory argument is preceded by an optional argument.

Parsing



Noun-verb ambiguity. The final ambiguity in this section involves a choice between a N-N compound and a N-V sequence. There is a preference for the N-N reading (Frazier & Rayner 1987), although the difficulty of the N-V reading may be greatly reduced by lexical factors (MacDonald 1993). The N-N analysis is preferred by Branch Right, because it requires only a minimal 2 step path between the two relevant words.

- (38) a. The desert trains young people to be very tough.
 b. The warehouse fires most of its employees within a month of hiring them.



3.3.3 Association ambiguities II

3.3.3.1 Multiple complement verbs

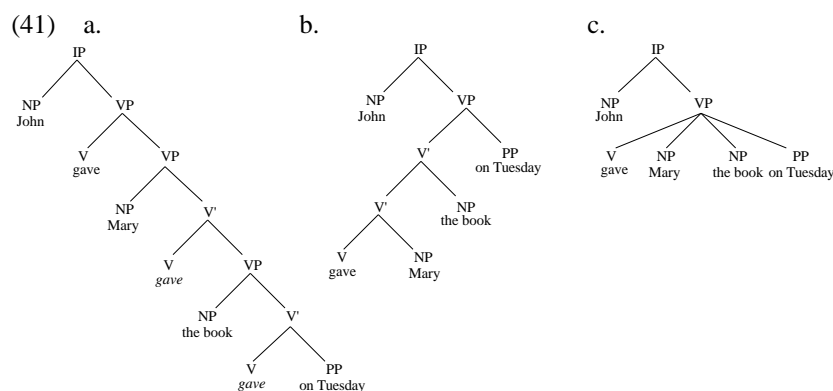
Consider the sentences in (40). These are all ambiguous sentences in which speakers typically show a preference to treat the ambiguous phrase as an argument of the verb rather than as a modifier.

Parsing

- (40) a. # I sent the child the bookcase fell on a message.
 b. # Bob convinced the woman that the president had met that the budget would never pass.
 c. # Sue put the book on the table under the table.

In traditional formulations of local attachment preferences these were the kinds of example that provided the best evidence that more than just a local attachment strategy was necessary to account for structural parsing preferences—arguments to this effect can be found in both Kimball 1973 and Frazier & Fodor 1978. The assumption was that the local attachment preference should prefer attachment to the most recently built phrase, i.e. NP, and therefore the modifier should be preferred in each of (40a–c). This prediction is incorrect, given the observed preference to treat the ambiguous phrase as an argument of the verb.

However, the preferences in (40) follow equally well from Branch Right, when we take into account the syntactic arguments for a modified analysis of complex VPs that have accumulated over the last 10 years, and which are reviewed and extended in Chapter 2 above. I refer the reader to Chapter 2 for a more detailed presentation, but in brief, a number of syntactic arguments have been advanced for an analysis of multiple complement constructions that is rather more right-branching than traditional analyses in which complex VPs were assumed to have either a left-branching (41b) or a ‘flat’ n-ary branching structure (41c). These different alternatives are shown in (41) below, for the sentence *John gave Mary the book on Tuesday*.



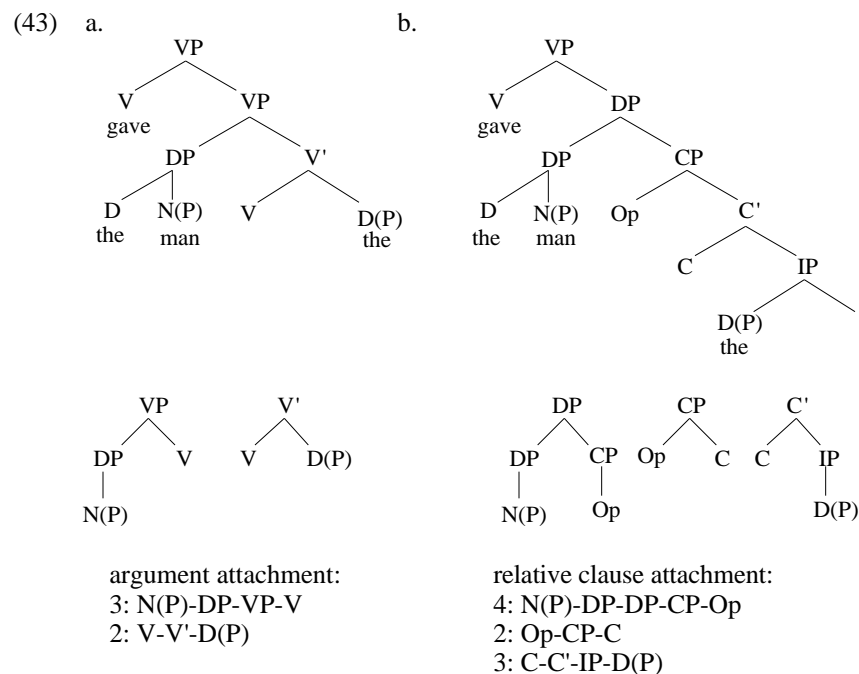
More details of the motivations for this analysis, together with some sample derivations can be found in Chapters 2 and 4.

Object–relative clause ambiguity. (42) gives examples of the theme object–relative clause ambiguity.

- (42) a. # The man gave the woman the dog bit a package.
 b. # I sent the child the bookcase fell on a message.
 c. # Dan bought the dog the ball hit a steak.

Parsing

The direct object attachment of the underlined NP is preferred to the reading in which that NP is the subject of a relative clause adjoined to the first NP. This is because the argument attachment can be made using paths of no longer than 3 steps, whereas the relative clause attachment requires a 4 step path between the head of the relative clause ('man') and the null operator in SpecCP of the relative clause.



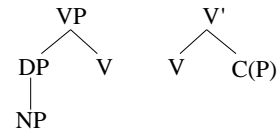
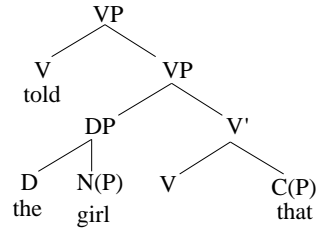
Complement clause–relative clause ambiguity. (44) is a different kind of argument/relative clause ambiguity. Here the choice is between treating a clause headed by *that* as the clausal complement of the matrix verb or as a relative clause modifying the first object of the verb.

- (44) a. # John told the man that he knew to brace himself for a storm.
 b. # The patient persuaded the doctor that he was having trouble with to leave.
 c. # Bob convinced the woman that the president met that the Sox had won a game.

Subjects prefer the attachment as an argument of the verb (cf. Frazier 1978, Crain & Steedman 1985, Altmann 1988, Mitchell, Corley & Garnham 1992), as predicted by Branch Right. The reason for the argument advantage is exactly the same as with the object–relative clause ambiguity above.

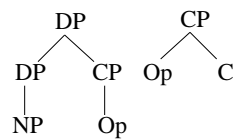
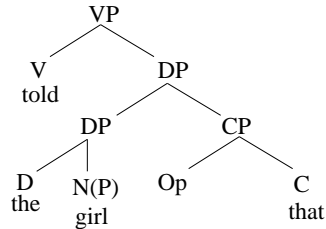
Parsing

(45) a.



argument attachment:
3: N(P)-DP-VP-V
2: V-V'-C(P)

b.



modifier attachment:
4: N(P)-DP-DP-CP-Op
2: Op-CP-C

3.3.3.2 Main verb–reduced relative ambiguity

The examples in (46) illustrate the main clause–reduced relative ambiguity pointed out by Bever (1970) which has been the most intensively studied of all structural ambiguities (see the comprehensive list of references in MacDonald et al. 1994). (47) shows how the alternative structures are evaluated by Branch Right. The relative clause attachment is dispreferred because of the path from the noun *horse* to the subject of the relative clause.^{77,78}

⁷⁷ Note: this is just one possible structural analysis of reduced relatives; they could also be treated as CPs with an operator in SpecCP. Since the argument in the reduced relative that is coreferential with the head of the relative clause undergoes passivization inside the relative clause it is hard to probe for standard properties of A-bar movement which would make it possible to decide between these two analyses of the reduced relative.

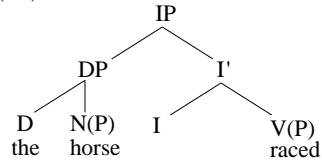
However, since either analysis involves a null specifier position between the noun and the participle verb, the choice of analysis makes no difference to the predicted structural preference.

⁷⁸ The argument structure of the verb seems to affect the strength of the garden path in this ambiguity. The verbs used in (46) *float*, *race* and *walk* all alternate between an intransitive and a transitive (causative) frame. Because the second argument is not obligatory, the sentences are disambiguated only when the sentence final verb is encountered. If, on the other hand, obligatorily transitive verbs like *find* are used, as in *The bird found in the bush was rare*, disambiguation is reached much earlier, and the initial mis-parse is much easier to recover from. See Pritchett 1988 and Gibson 1991 for alternative accounts of the difference between obligatorily and optionally transitive verbs in this ambiguity.

Parsing

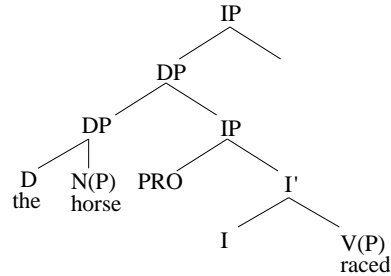
- (46) a. The horse raced past the barn fell.
 b. The dog walked to the park chewed the bone.
 c. The boat floated down the river sank.

(47) a.



main verb attachment:
 3: N(P)-DP-IP-I
 2: I-I'-V(P)

b.



reduced relative clause attachment:
 4: N(P)-DP-DP-IP-PRO
 2: PRO-IP-I
 2: I-I'-V(P)

3.3.3.2 PP-attachment: argument-by-category interaction

So far in 3.3.3 we have seen how to capture argument preferences in a theory which does not build in an explicit argument preference. Next I show an advantage of this approach over a theory with an explicit argument preference, by showing that the same structural property which creates an argument preference in other ambiguities captures related ‘anti-locality’ effects involving non-arguments.

PP-attachment ambiguities like (48) have been studied intensively in the experimental literature on parsing (cf. Rayner, Carlson & Frazier 1983, Clifton, Speer & Abney 1991, Spivey-Knowlton & Sedivy 1995, Schütze & Gibson 1996).

- (48) The spy saw the cop with the revolver.

In V-NP-PP sequences at least two choices must be made in attaching the PP. First, a choice of category—whether to attach the PP to a projection of N or V; second, the choice of whether to attach the PP as an argument or as a modifier of the chosen category. Some typical examples are shown in (49–50) (adapted from Clifton, Speer & Abney 1991). In (49) the PP can be either an argument of N or a modifier of V, and in (50) the PP can be either an argument of V (50a) or a modifier of N (50b).⁷⁹

⁷⁹ See Schütze 1995 for a useful survey of diagnostics for argumenthood and adjuncthood.

Parsing

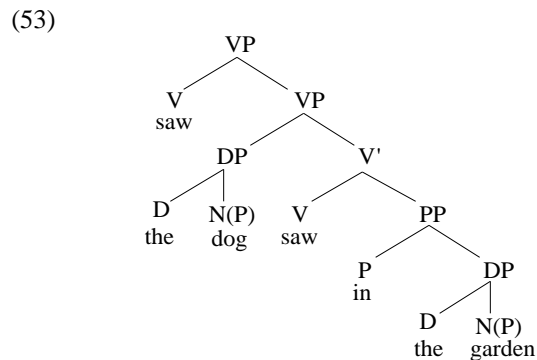
- (49) a. The man expressed his interest in the car at the sale. (N, argument)
 b. The man expressed his interest in the car rather than outdoors. (N, modifier)
- (50) a. I put the candy on the table for the children. (V, argument)
 b. The cook placed the cake in the oven on the table. (V, modifier)

If we apply the kinds of constituency tests familiar from Chapter 2 to examples like these, we find that arguments and adjuncts of V behave alike, whereas arguments and adjuncts of N behave differently, implying that the argument/adjunct distinction does not correspond to a consistent structural contrast.

PPs which are arguments and modifiers of V behave alike on coordination and reciprocal binding tests (51–52; see §2.7 for further examples).

- (51) a. I saw [the dog in the garden] and [the cat on the roof]
 b. The baker put [the cake in the oven] and [the pastry on the shelf]
- (52) a. I saw the dogs in each other's kennels.
 b. He put the children in each others beds.

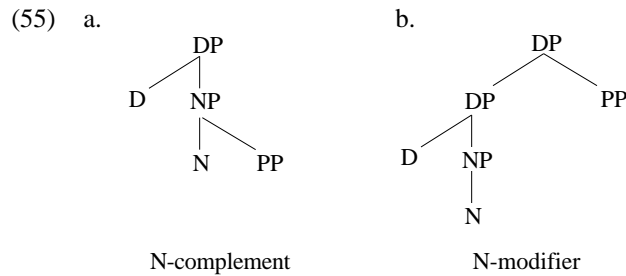
These parallels were noted by Larson (1990) and Stroik (1990), who concluded that PP modifiers of V must be structurally identical to PP arguments of V, and therefore attached as complements of the lower V in a nested VP structure like (53).



In contrast to arguments and modifiers of verbs, which behave alike on constituency tests, arguments and modifiers of nouns show an asymmetry in coordination tests. N-PP sequences can be coordinated to the exclusion of a determiner if PP is an argument of N (54a), but N-PP cannot be coordinated if PP is a modifier of N (54b). I take this as evidence for the two structures in (55), in which an argument of N is the sister of N, and a modifier of N is a sister of NP (or DP). The binding tests are inapplicable as tests for the constituency of NP.

Parsing

- (54) a. Everybody always appreciates the [giving of advice] and [showing of support].
 b. * My uncle gave me two [gifts in a box] and [receipts still stapled to the outside].



Why there should be this structural contrast between modifiers of verbs and modifiers of nouns is not clear to me, but the distinction does appear to be motivated by the facts.

Therefore there is a structural distinction between arguments and modifiers of nouns. Putting this together with the parallelism between arguments and adjuncts of V we obtain the following predictions for attachment preferences in V-NP-PP sequences in English (56). Arguments of N are predicted to be most preferred, followed by any attachment to V, with attachment as a modifier of N predicted to be worst.

- (56) *Predicted preferences in V-NP-PP sequences:*
 a. N-arguments
 b. V-arguments and V-adjuncts
 c. NP-adjuncts

These predictions appear to be supported by a combination of intuitions and experimental results. The preference for arguments of N over adjuncts of V was found by Schütze & Gibson (1996),⁸⁰ and the preference for arguments of V over modifiers of N has been reported by a number of studies (Rayner, Carlson & Frazier 1983, Clifton, Speer & Abney 1991). Both of these preferences also follow from an explicit ‘argument first’ parsing heuristic. What does not follow from the argument first approach is the intuitive preference for modifiers of verbs over modifiers of nouns, as in sentences like (57).

- (57) I saw the man in the park.

⁸⁰ Clifton, Speer, and Abney (1991) independently manipulated argumenthood and the category of the attachment site for PPs and failed to find any immediate preference for argument attachments to nouns, contrary to what Branch Right predicts. However, Schütze (1995) and Schütze & Gibson (1996) point out that many of Clifton et al.’s intuitively chosen argument and adjunct PPs fail to pass a number of syntactic tests of argumenthood or adjuncthood.

Branch Right predicts this ambiguity to be resolved in favor of the V-modifier reading rather than the N-modifier reading, for exactly the same reason that V-arguments are predicted to be preferred over N-modifiers.

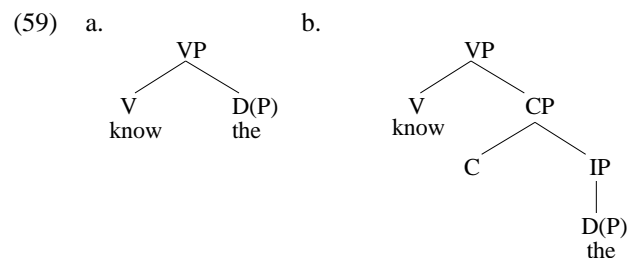
Thus, we can show that phrases which are not arguments, but which occupy a similar phrase structure position to arguments, are parsed in a similar manner to arguments. This parallel is captured by a purely structure-based principle like Branch Right, but not by an approach which explicitly favors argument attachments.

3.3.4 (In)dependence ambiguities

This section deals with situations in which the parser has to choose whether or not to form a dependency with a preceding element. The resolution of these ambiguities is handled by a simple generalization: dependencies are established where possible. This generalization follows from Branch Right, but it could be captured by other theories too.

Object–Subject of complement ambiguity. In (58a–b) the ambiguous NP can be either the direct object of the matrix verb, or it can be the subject of a clausal complement. There is a preference for a direct object interpretation of the ambiguous NP, and (59) shows how Branch Right predicts this.

- (58) a. I know the answer to the physics problem is wrong
 b. Bill heard the gossip wasn't true



direct object attachment: 2: V-VP-D(P)
 embedded subject attachment: 2: V-VP-C(P)
 3: C-CP-IP-D(P)

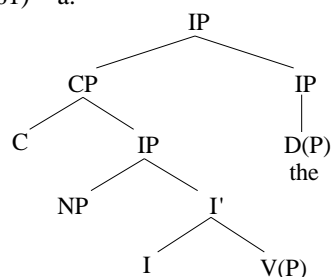
There has been some controversy in the recent literature over whether subjects show any reliable general preference for one attachment over the other (cf. Frazier & Rayner 1982; Ferreira & Henderson 1990; Juliano & Tanenhaus 1993; Trueswell, Tanenhaus & Kello 1993). However, although it has been shown that the structurally more complex analysis can be made easier or harder by varying lexical factors, I am unaware of evidence that the structurally simpler analysis is ever difficult.

Object–Matrix subject ambiguity. In (60) the NP *the sock* attaches to a different category with a different role in the two alternative structures to be

evaluated. In the first (61a), *the sock* is the subject of the matrix clause, and in the other alternative it is the direct object of the verb *mending*. Therefore Branch Right correctly predicts (60) to be a garden path, because the direct object analysis requires a shorter path from the verb to the NP. The alternative structures are shown in (61).

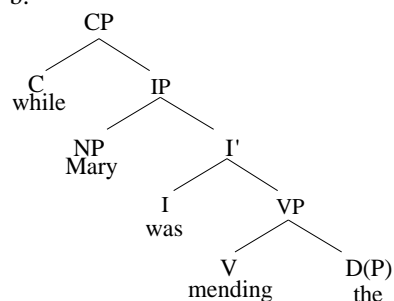
(60) While Mary was mending the sock fell off her lap.

(61) a.



matrix attachment:
6: V(P)-I'-IP-CP-IP-IP-D(P)

b.



embedded attachment:
2: V-VP-D(P)

Embedded subject-matrix subject ambiguity. A number of verbs appear in the following range of frames: they can either be intransitive, or select a complement clause (without *that*), but they do not allow an direct object NP complement. This has the effect that it is possible to construct local ambiguities in which the NP in a V-NP sequence is either the subject of the verb's clausal complement or the subject of an independent higher clause. Some examples are given in (62–64) and the alternative structures at the choice point are shown in (65).⁸¹

(62) a. Whenever Mary started to think the class had finished there was still an hour remaining.
b. Whenever Mary started to think the class had finished and everybody had left.

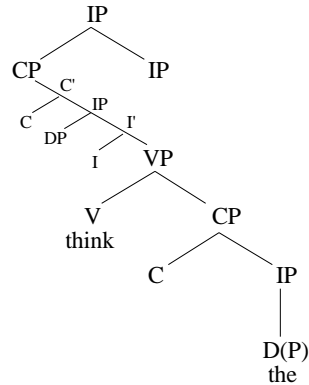
(63) a. When Bill complained the class was boring it usually was.
b. When Bill complained the class was annoying as a result.

(64) a. If Tom confessed the gun was his he would be in trouble.
b. If Tom confessed the gun was his.

⁸¹ This ambiguity has similar structural properties to the ambiguity studied in the experiments in §3.2, except that in this case the alternative structures use different subcategorization frames of the verb, whereas the examples in §3.2 hold the verb's subcategorization frame constant. This makes the results of the experiments in §3.2 easier to interpret than preferences on the subject-subject ambiguity.

Parsing

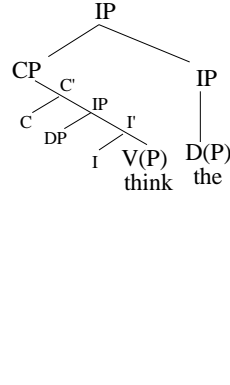
(65) a.



embedded subject attachment:

2: V-VP-C(P)
3: C-CP-IP-D(P)

b.



matrix subject attachment:

7: VP-I'-IP-C'-CP-IP-IP-DP

In the continuation in which the ambiguous NP is the subject of the complement of *think*, many new nodes must be built (viz. C, CP, IP, DP), but the paths between adjacent heads are kept short. In the analysis in (65b), on the other hand, fewer new nodes must be built than in (65a), but the path between *think* and the ambiguous NP is extremely long, because it includes the right edge of the entire clause. Therefore, the structure in (65b) contains the more ‘minimal’ attachment of *the*, but the structure in (65a) is favored by Branch Right, a clear divergence of predictions.

Similar examples can be constructed with a handful of other verbs (e.g. *agree*, *dissent*, *concur*, *joke*, *cry*), and speakers generally show an intuitive preference for the embedded subject continuation (the (a) examples in 62–65). The preferences have not been experimentally confirmed, but if the intuitive preference is reliable, then these subject-subject ambiguities favor a locality based account (e.g. Branch Right) over a global node-counting metric such as Minimal Attachment.⁸²

⁸² A concern that arises about these examples is whether the intuitive preference for low attachment in these examples is due to a strong frequency bias for the transitive frames of the verbs involved. I have checked for this possible confound by extracting frequencies of transitive and intransitive frames in the parsed Brown Corpus (available in the Penn Treebank) for verbs displaying the subject-subject ambiguity. It turns out that the frequency of a given subcategorization varies greatly according to the tense and aspect of the verb. Some tenses/aspects show strong frequency biases, while other tenses/aspects of the same verb are roughly equibiased. For example, present tense *think(s)* is strongly biased in favour of a sentential complement, but past tense *thought* occurs with approximately equal frequency with a complementizerless sentential complement or with no complement at all. Since it is not yet known how frequency biases carry over from one part of a verbal paradigm to another, it is hard to say whether it would be possible to experimentally verify this preference using balanced materials.

Summarizing the survey of structural ambiguities in this section, we have seen that with the principle Branch Right it is possible to achieve rather broad empirical coverage with a local attachment principle, a principle which almost any theory of parsing has to incorporate in any case. Therefore, this section adds to the argument from §3.2 an argument for Branch Right based on the redundancy of other well-known parsing principles.

Nevertheless, there is a possible concern that Branch Right is successful in English just because English is a rather ‘right-branching’ language, and that it will run into difficulties in languages which are generally considered to be less right-branching. Therefore in the next section we examine some structural ambiguities in head-final constructions in German and Japanese, which have generally been considered to be less right-branching languages.

3.4 Some attachment preferences in head-final languages

In this section I consider some parsing preferences in head final languages. Apart from the general interest of testing the scope of Branch Right in languages other than English, the particular interest of head-final languages is that they are often claimed to have left-branching phrase structure. Given that Branch Right favors the building of right-branching structures, it might appear that Branch Right imposes preferences that go against the canonical structural configurations of these languages. This section examines how accurate it really is to refer to head-final languages like Japanese as ‘left-branching’, and how the predictions of Branch Right match up with what is currently known about ambiguity resolution in these languages.

From a historical perspective, one of the earliest formal theories of human parsing, Yngve’s *Depth Hypothesis* (Yngve 1960, 1961), was rather similar to Branch Right in that it placed a cost on any deviations from strictly right-branching structures. A criticism that has often been leveled against Yngve’s model is that it falsely predicts that ‘left-branching languages’ should be extremely difficult to parse (cf. Frazier 1985). Since we know that Japanese is parsable, this is potentially a very serious problem. In fact, it now appears that Japanese sentences become incomprehensible under almost exactly the same conditions as comprehension breakdown occurs in English (Babyonyshev & Gibson 1995, Babyonyshev 1996). An auxiliary aim of this section, therefore, will be to see whether the criticism of Yngve’s proposal (to the extent that it

It is important to use verbs that do not allow a direct object NP subcategorization frame, because this would create an independent reason for favoring the embedded subject attachment over the matrix clause attachment. If the NP was first attached as the direct object of the verb, it would then be easy to reanalyze the NP as the subject of the complement clause.

In fact, it is not even clear that we can rule out the possibility that this kind of initial misanalysis of the NP as a direct object does not occur here in any case, given the claim that some people have made that initial NP attachments can be made that violate subcategorization requirements of the verb. [If this were possible, then the NP could initially be misanalyzed as a direct object and then reanalyzed as an embedded subject.] This assumption has been argued for by Mitchell (1987), but called into question by Boland and Tanenhaus (1991).

was ever well-founded) also applies to Branch Right. My conclusion will be that Branch Right fares well in these languages, and that they are not so left-branching after all.

3.4.1 Head position and PP-attachment in German

This section briefly discusses how elements are combined in head-final constructions and how Branch Right predicts variations in head position to lead to variations in parsing preferences.

In a series of papers Konieczny and colleagues present an interesting set of findings about preferences for PP-attachment in German (Konieczny, Hemforth, Scheepers & Strube 1995, Konieczny, Hemforth & Scheepers 1995). PPs in German show the same range of ambiguities as they do in English between argument and adjunct status and between N-attachment and V-attachment. The additional property of German which makes PP-attachment particularly interesting is that the position of the verb can be varied such that it either precedes or follows the ambiguous NP-PP sequence. Konieczny et al. found that preferences are reversed by changing the position of the verb.

(66–68) show sample items from Konieczny et al.’s stimuli. The correct attachment sites for the PPs are determined pragmatically. (The stimuli used in these experiments all contained a strong plausibility bias towards one or the other attachment.) Verbs like *watch* and *hit* show a strong preference for a PP headed by *with* to have an instrumental (V-attached) reading, whereas verbs like *notice* and *catch sight of* create a preference for an attributive reading of the PP.⁸³ The four possible combinations of verb bias and PP bias are shown for verb second clauses in (66), and then schemata for the 4 conditions in two types of verb-final clauses: a matrix clause perfect tense construction (67) and past tense in an embedded clause (68).

- (66) *Verb-second ambiguities*
- a. Marion beobachtete das Pferd mit dem neuen Fernglas.
M. watched the horse with the new binoculars.
(V2: instrumental bias V, instrumental bias PP)
 - b. Marion beobachtete das Pferd mit dem weißen Fleck.
with the white mark.
(V2: instrumental bias V, attributive bias PP)
 - c. Marion erblickte das Pferd mit dem neuen Fernglas.
caught sight of
(V2: attributive bias V, instrumental bias PP)
 - d. Marion erblickte das Pferd mit dem weißen Fleck.
(V2: attributive bias V, attributive bias PP)

⁸³ It was not clear from the paper that was available to me how representative these examples are of Konieczny et al.’s stimuli, in particular whether all of their PPs were ambiguous between arguments of the verb and modifiers of the noun, or whether there were V-modifier and N-argument cases as well.

Parsing

(67) *Verb-final ambiguities*

Marion hat das Pferd $\left\{ \begin{array}{l} \text{mit dem neuen Fernglas} \\ \text{mit dem weißen Fleck} \end{array} \right\} \left\{ \begin{array}{l} \text{beobachtet} \\ \text{erblickt} \end{array} \right\}.$
M. has the horse PP V

(68) Ich habe gehört, daß Marion das Pferd $\left\{ \begin{array}{l} \text{mit dem neuen Fernglas} \\ \text{mit dem weißen Fleck} \end{array} \right\}$
 $\left\{ \begin{array}{l} \text{beobachtete} \\ \text{erblickte} \end{array} \right\}.$
I have heard that M. the horse PP V

Konieczny et al. used stimuli like these in a number of different reading studies, using a variety of different presentation methods. They used self paced reading with phrase-by-phrase presentation (e.g. *Marion / hat / das Pferd / mit dem neuen Fernglas / beobachtet*) and with word-by-word presentation (e.g. *hat / das / Pferd / mit ...*). They also used an eye-tracking paradigm.⁸⁴

In both the word-by-word self-paced reading study and in the eye-tracking study Konieczny et al. found a contrast between attachment preferences in verb-second and verb-final clauses. In verb-second clauses, in which the verb preceded the ambiguous NP-PP sequence as in English, there was evidence that readers preferred to treat the PP as an argument of the verb rather than a modifier of the noun, thereby replicating the finding of many studies of PP-attachment in English. In verb-final clauses, on the other hand, a preference for treating the PP as a modifier of the NP was found.⁸⁵

This finding is, in effect, the same pattern of results that I discussed under the heading of the 'locality puzzle' in §3.2 above. The XP in an NP-XP sequence is analyzed as a modifier of the NP in some syntactic contexts but not in others. In English the contrast is between a V-NP-NP sequence on the one hand, in which the verb is monotransitive and the NP could be the subject of a relative clause or the subject of an independent clause, and a V-NP-XP sequence on the other hand in which the verb is ditransitive and the XP could be the second object of the verb or part of a relative clause adjoined to the first NP. Konieczny et al.'s German contrast is in some ways a more minimal pair, because just the reordering of the NP-PP sequence with respect to the verb causes the reversal of attachment preferences.

⁸⁴ Eye-tracking has the advantage of allowing faster, more natural reading of the stimulus sentences. This is potentially a big advantage when looking for misanalyses which may be corrected extremely rapidly. The disadvantage of eye-tracking is that the data can be rather difficult to interpret, because fixations typically do not progress smoothly from the start to the end of the sentence, and include many regressive eye movements. This means that there are a number of different possible measures of the time taken to read any given word, depending on which of the various fixations on a particular word are taken into consideration.

⁸⁵ Different results were found in the phrase-by-phrase self-paced reading study. Here no preference for treating the PP as a modifier of the NP was found. Konieczny et al. argue that this difference is likely to be due to bias introduced by presentation of the NP and the PP as two separate phrases.

Parsing

Konieczny et al. propose that their findings are due to a ranking of parsing strategies which they refer to as *ParameterizedHeadAttachment*. The strategies in (69) are assumed to apply in order, until one of them determines a choice between the possible parses.

- (69) PARAMETERIZED HEAD ATTACHMENT (Konieczny et al. 1995)
- a. *Head Attachment*: prefer to attach an item to a phrasal unit whose lexical head has already been read.
 - b. *Preferred Role Attachment*: prefer to attach an item to a phrasal unit whose head *preferentially* subcategorizes for it.
 - c. *Most Recent Head Attachment*: prefer to attach an item to the head that was read most recently.

This theory predicts that in verb-final clauses the PP in an NP-PP sequence will be treated as a modifier of the NP by virtue of (69a), because the head of the NP has been encountered and not the head of the VP. In verb second clauses (69a) does not make any choice between the two parses of the V-NP-PP sequence, because the heads of both possible attachment sites have been encountered. (69b) ensures VP-attachment when the PP is a potential argument of the verb and not a potential argument of the N. (69c) is the elsewhere condition.⁸⁶

However, Konieczny et al's findings follow equally well from the Branch Right principle. Verb second clauses work just like English; verb-final clauses require that I first specify how head-final phrases are assembled from left-to-right.

(70a–b) shows how the different treatments of the PP are evaluated by Branch Right in a verb-second clause—the predictions here are just the same as with PP attachment in English. The VP-attachment is more highly valued by Branch Right; the path from the NP to the PRO subject of the PP modifier is longer than any path required to build the VP-attachment structure, and therefore proves fatal for the NP-attachment.

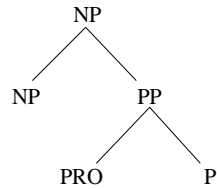
⁸⁶ Notice that this set of conditions predicts that in a V-NP-PP sequence in which neither the V nor the N can take the PP as an argument, the PP will be preferentially interpreted as a modifier of N rather than of V. This predicts a V-attachment preference in examples like (i–ii):

- (i) The spy saw the cop on a boat.
- (ii) Susan played the piece in 4:4 time.

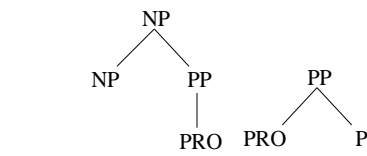
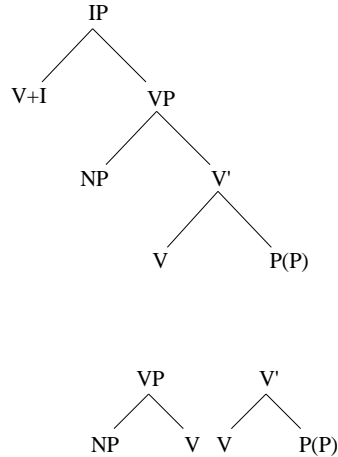
To my knowledge this prediction is not supported, at least for English.

Parsing

(70) a.



b.



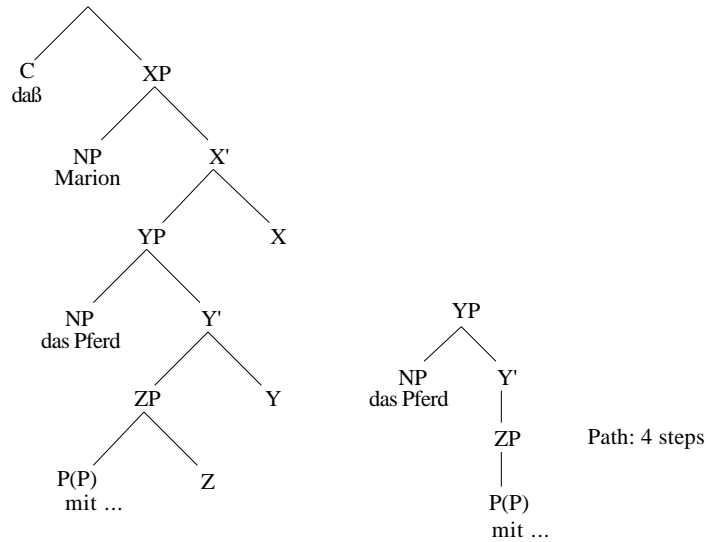
(71) shows how the non-modifier treatment of the PP is treated in a clause in which the main verb appears in final position. The account depends on the following assumptions. I assume that when a word is encountered that cannot be merged with the existing structure in any way by satisfying any requirements of itself or of an existing head, an empty head is projected to its right. This head lacks any content until a head with content is inserted to fill it. Until it is filled, it serves merely as syntactic ‘glue’ in order to allow incoming material to be added to the existing structure.⁸⁷

In an ambiguous string like *daß Marion das Pferd mit dem neuen Fernglas ...* there are no contentful heads with which the NPs can directly merge, and therefore a series of empty heads must be projected. The relevant path to pay attention to in (71) is the one from the NP *das Pferd* to the preposition *mit*.

⁸⁷ An alternative would be to explicitly predict phrasal nodes before their heads are encountered. This would mean, for example, that V heads would be posited instead of X and Y heads in examples like (71). For the cases under consideration the choice between this and the mechanism suggested in the text makes no difference.

This mechanism was implicitly assumed earlier in my treatment of the attachment of subjects in English, where I assumed that a node labeled IP intervened between CP and the subject, even before the head of IP was encountered. Since the head of CP and the subject cannot be directly merged, this additional node was necessary.

(71)



As the figure shows, the path from the NP to the PP when the PP is not treated as a modifier of the NP is longer than any paths required for VP-attachment in a verb-second clause (70b). This is also longer than the path required for the NP-modifier attachment of the PP, which is identical in verb second and verb-final clauses (70a). Therefore, the modifier attachment of the PP is correctly predicted to be preferred in verb-final clauses.⁸⁸

Since the reversal of preferences between verb-second and verb-final clauses follows from Branch Right, there does not seem to be any need to explicitly invoke Konieczny et al.'s Head Attachment principle (69), given that its effects can be captured with just the locality condition Branch Right.⁸⁹

3.4.2 Japanese

In this section I examine parsing ambiguities in Japanese, in order to clarify the status of my right-branching parsing strategy in a language that is often alleged to be left-branching. My conclusion will be that Japanese does not pose obvious problems for Branch Right, and that the key aspects of ambiguity resolution in Japanese can in fact be captured by a wide range of different accounts of ambiguity resolution. This is because Japanese shows a much more

⁸⁸ Clauses in which there is an auxiliary in second position but the main verb is in final position will generate identical preferences for PP attachment. Although an empty head does not need to be projected in order to attach the subject NP in these cases, because the subject can form an agreement relation with the auxiliary, this makes no difference to the status of the object NP and the PP.

⁸⁹ I assume that the empty heads Y and Z in (71) become V heads once the verb is encountered. The verb is first inserted into the lower position (Z), and then raises to the higher position (Y).

restricted range of structural ambiguities than English. In explaining Japanese sentence comprehension the most interesting questions arise not in determining how to proceed from one appropriate parse state to the next, but in explaining how to recover from inappropriate parse states, as the literature on Japanese sentence processing shows (see, for example, the papers in Mazuka & Nagai 1995).

The first step, though, will be to clarify a couple of claims that often appear in discussions of parsing in Japanese: Japanese is a ‘left-branching’ language, and Japanese is massively ambiguous. These claims are both true, but only in a restricted sense.

3.4.2.1 How left-branching is Japanese?

This section focuses on syntax rather than parsing. The aim is to show that the kinds of arguments which have motivated syntacticians to claim that English has extremely right-branching clause structure apply equally well to Japanese. As an illustration of this, I show that subject-oriented secondary predicates, which are generally assumed to have an underlying position outside VP show properties of phrases deep inside a right-branching VP, parallel to the behavior of adverbial phrases in English in Chapter 2.

The structure of this argument is as follows. Probing order and c-command relations in underlying structures in Japanese is difficult, because of the disrupting effects of scrambling. I take a test involving binding relations and quantifier float which has previously been used to diagnose underlying relations in subject–object clauses and double object constructions. Applying this test to subject-oriented secondary predicates shows that secondary predicates are underlyingly lower than arguments in Japanese, just as in English.

Secondary predicates in Japanese are marked by the suffix *-de*.⁹⁰ The sentences in (72), taken from the discussion of Japanese secondary predicates in Koizumi 1994, illustrate the free ordering possibilities for the subject-oriented secondary predicate. The predicate need not be adjacent to either the subject or the verb. This is not surprising, given the liberal scrambling that is widely attested in Japanese.

- (72) a. Taroo-ga **hadaka-de** hon-o yonda
 -nom naked book-acc read
 ‘Taro read a book naked.’
- b. **Hadaka-de** Taroo-ga hon-o yonda
 naked -nom book-acc read
- c. Taroo-ga hon-o **hadaka-de** yonda
 -nom book-acc naked read

More interestingly, Koizumi shows that objects can bind variables contained in subject-oriented secondary predicates in Japanese (73). The plural accusative

⁹⁰ *-de* is also used to mark phrases denoting locations, manners etc.

Parsing

object NP binds a variable inside the secondary predicate in (73a), and (73b) shows the same for a dative object.⁹¹

- (73) a. Hanako-wa [Taroo to Ziroo]_i-o [soitu_i-no mottomo
 -top and -acc SOITU-gen most
 sukina kakko]-de demukaeta
 favorite fashion met
 ‘Hanako met [Taro and Ziro]_i in his_i most favorite dress.’
- b. Hanako-wa [Yamada-to yuu otoko to Tanaka-to yuu otoko]_i-ni
 -top called man and called man -dat
 [soitu_i-no mottomo sukina kakko]-de tanzoo purezento-o watasita
 SOITU-gen most favorite fashion birthday present-acc gave
 ‘Hanako gave a birthday present to [a man called Yamada and a man
 called Tanaka]_i in his_i most favorite dress.’

Although these facts are similar to what we have seen in English in §2.7, they do not tell us much about where subject-oriented secondary predicates are generated in Japanese. Given the ease of scrambling in Japanese, sentences like (73) could straightforwardly be derived by generating the secondary predicate above the object and then scrambling the object to a position higher than the secondary predicate. This is the analysis of (73a–b) which Koizumi assumes.

However, further elaboration of sentences like (73) allows us to show that subject-oriented secondary predicates underlyingly lower than objects, and that the order in (73) does not need to be derived by object scrambling. This argument is based on an extension of tests used by Miyagawa (1995) in the context of an examination of double object structures in Japanese. Miyagawa provides a diagnostic for whether an accusative object is scrambled or not. (74) shows that object scrambling to a pre-subject position produces ungrammaticality when the subject is a reciprocal (74a), but not when the subject *contains* a reciprocal (74b). Given that reciprocal binding is possible in (74b) Miyagawa assumes that the object scrambling in (74a–b) is A-scrambling, and that the problem with (74a) is that it violates the *Chain Condition* (Rizzi 1986).

- (74) a. * [John-to Mary]_i-o otagai_i-ga t_i mita
 [John-and Mary]-acc_i each other_i-nom t_i saw
 ‘John and Mary, each other saw.’

⁹¹ The Japanese personal pronouns *kare* ‘he’ and *kanozyo* ‘she’ do not generally tolerate bound variable interpretations. However, Hoji (1985) shows that a certain class of elements, which have roughly the force of anaphoric epithets in English (e.g., *the guy*) can be construed as bound variables if they are c-commanded by an appropriate quantificational element. These ‘pronouns’ are *soko* ‘that place’, *sore* ‘that thing’, *soitu* ‘that person’.

Parsing

- b. [John-to Mary]_i-o otagai_i-no sensee-ga t_i mita
 [John-and Mary]-acc_i each other_i-gen teachers-nom t_i saw
 ‘John and Mary, each other’s teachers saw.’

In contrast, no similar problems arise when direct objects precede indirect objects, despite the fact that the ‘unmarked’ order in double object constructions in Japanese is IO DO—(75) is far better than (74a). Miyagawa’s conclusion from this, and other contrasts similar to (74–75), is that direct objects may be generated either above or below indirect objects in Japanese.

- (75) John-ga [Hanako-to Mary]_i-o (paati-de) otagai_i-ni syookaisita
 John-nom [H.-and M.]_i-acc (party-at) each-other_i-dat introduced
 ‘John introduced Hanako and Mary to each other at the party.’

However, Miyagawa shows that there are situations in which the direct object appears to have been moved across the indirect object, and in these situations we find the same kind of contrast as in (74a–b) involving Chain Condition violations. (76) illustrates the well-known fact that direct objects can strand numeral classifiers in Japanese.

- (76) a. John-ga Mary-ni **pen-o ni-hon** ageta
 John-nom Mary-dat pen-acc 2-cl gave
 ‘John gave two pens to Mary.’
 b. John-ga **pen_i-o** Mary-ni **t_i ni-hon** ageta
 John-nom pen_i-acc Mary-dat t_i 2-cl gave

The paradigm in (77) shows that in just the cases where the direct object and its associated numeral classifier appear on either side of the indirect object, the Chain Condition effect in (74a–b) is found (cf. 77c–d).

- (77) a. John-ga gakusei-tati_i-o otagai_i-ni syookaisita
 John-nom students_i-acc each other_i-dat introduced
 ‘John introduced the students to each other.’
 b. John-ga **gakusei-tati_i-o futa-ri** otagai_i-ni syookaisita
 John-nom students_i-acc 2-cl each other_i-dat introduced
 ‘John introduced two students to each other.’
 c. * John-ga **gakusei-tati_i-o** otagai_i-ni **t_i futa-ri** syookaisita
 John-nom students_i-acc each other_i-dat t_i 2-cl introduced
 d. (?) John-ga **gakusei-tati_i-o** [otagai_i-no sensei]-ni **t_i futa-ri**
 John-nom students_i-acc [each-other_i-gen teacher]-dat t_i 2-cl
 syookaisita
 introduced
 ‘John introduced two students to each other’s teachers.’

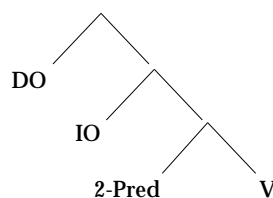
Parsing

In the light of this paradigm, we are now in a position to ask whether the objects in (73a–b) are above the subject-oriented secondary predicates because that is their underlying position, or whether they scrambled over the secondary predicates. Extending Miyagawa’s paradigm in (77) we can construct situations in which the direct object must have been generated higher than an indirect object, and ask whether a subject-oriented secondary predicate can occur below the direct object position in such cases.

The paradigm in (78–80) suggests that the secondary predicate may be underlyingly below the direct object. (78) shows that configurations like (73), in which the object binds a variable inside the secondary predicate, are possible even when the object also binds an indirect object reciprocal. Since there is no Chain Condition violation, we must assume that the direct object did not cross the indirect object, which in turn entails that it did not cross the secondary predicate either. This implies a schematic underlying structure like (79).

- (78) Hanako-ga [Taroo-to Jiroo]_i-o otagai_i-ni
 Hanako-nom [T.-and J.]-acc each other-dat
 [soitu_i-no mottomo sukina kakko]-de syooksaisita
 in his most favorite dress-de introduced
 ‘Hanako introduced Taroo and Jiroo to each other in his most favorite dress.’

(79)



The examples in (80) serve as controls for other possible accounts of why (78) is possible. (80a) shows that—in parallel to (77c)—if a numeral classifier associated with the direct object is added to the right of the secondary predicate, presumably marking the underlying position of the direct object, the sentence becomes sharply ungrammatical. This is presumably due to a Chain Condition violation. (80b) shows that the intervention of the secondary predicate between the direct object and the numeral classifier does not present any problem. This guarantees that the reciprocal is the offending element in (80a).

- (80) a. * Hanako-ga gakusei-tati_i-o otagai_i-ni [soitu_i-no mottomo
 Hanako-nom students-acc_i each other_i-dat in his_i most
 sukina kakko]-de t_i futa-ri syooksaisita
 favorite dress-de t_i 2-cl introduced
 ‘Hanako introduced two students to each other in his most favorite dress.’

Parsing

- b. Hanako-ga gakusei-tati-o [soitu_i-no mottomo sukina kakko]-de t_i
futa-ri demukaeta
Hanako-nom students-acc_i in his_i most favorite dress-de t_i 2-cl met
'Hanako met two students in his most favorite dress.'

So subject-oriented secondary predicates appear to be freely generated deep inside a right-branching VP structure in both English and Japanese.⁹²

This finding is particularly interesting in the case of Japanese, since the secondary predicate is presumably therefore generated as the phrase *closest* to the verb, despite being the *least* semantically related to the verb.

The more general conclusion that I draw from this is that Japanese clause structure is left-branching only in a restricted sense, involving the relation of NP and clausal complements to their heads. However, as far as the hierarchical organization of phrasal constituents is concerned it seems likely that Japanese clause structure is just as 'right-branching' as English.

3.4.2.2 How ambiguous is Japanese?

Discussions of parsing in Japanese often begin with the remark that Japanese is massively ambiguous, as a result of the following three properties:

- (81) a. Phrases are strictly head-final
b. Quite liberal scrambling of phrases is allowed.
c. Dropping of both subjects and objects is allowed.

Another fairly common observation about Japanese sentence processing, however, is that it is remarkably difficult to confuse Japanese speakers.⁹³ To a non-native speaker it can seem that Japanese sentences ought to be extremely difficult to parse, but native speakers experience very little difficulty in most cases. Obviously, the supposed massive ambiguity of Japanese and the remarkable ease with which it is understood do not sit well together. One

⁹² Koizumi 1994 presents a number of arguments in support of his claim that subject-oriented secondary predicates in Japanese are either adjoined to VP or to I', whereas object-oriented secondary predicates are inside VP. However, the focus of most of these arguments is to show a *contrast* between subject- and object-oriented secondary predicates, a claim which I do not dispute. The facts presented here about subject-oriented secondary predicates do not undermine the contrast between the two types of secondary predicate, they merely show that the contrast must involve structures deeper inside VP.

Koizumi's tests include the ability of secondary predicates to be stranded by *soo su* ellipsis (cf. §2.6.2) or by VP fronting: he finds that the subject-oriented type can, but the object-oriented type cannot. In this respect, subject- and object-oriented secondary predicates contrast in exactly the same way that VP arguments and VP modifiers contrast under similar tests in both English and Japanese (see Chapter 2 for relevant English data).

⁹³ Hence the title of a paper 'Can Japanese speakers be led down the garden path' (Mazuka & Itoh 1995), and the following remark from Inoue & Fodor 1995: 'It is easier to argue that parsing Japanese is impossible than to explain how it is done.' (p. 9).

Parsing

response to this has been to argue that the enormous amount of ambiguity in the language is actually the explanation for why it is so hard to garden-path Japanese speakers (Inoue & Fodor 1995). Inoue & Fodor suggest that the more possible points of ambiguity there are in a sentence, the lower the level of confidence that a speaker assigns to any given analysis, and therefore the more prepared the speaker is to deal with errors when they arise.

The claim that Japanese is massively ambiguous deserves closer attention, though. I aim to show here that Japanese shows structural ambiguities of only a restricted kind, which are not very complex and which it is relatively easy to account for.

It is true that a sentence-initial clause may be arbitrarily deeply embedded, and that a listener has no way of knowing just how deeply embedded the clause is until the end of the sentence occurs, as the following examples from Mazuka & Itoh 1995 illustrate.

- (82) a. Hiroshi-ga Masao-o mita ...
H. -nom M. -acc saw
'Hiroshi saw Masao.'
- b. [Hiroshi-ga Masao-o mita] toki ...
H. -nom M. -acc saw when
'When Hiroshi saw Masao, ...'
- c. [[Hiroshi-ga Masao-o mita] toki Takashi-ga futari-o yonda] node ...
H. -nom M. -acc saw when T. -nom the two-acc called because
'Because Takashi called the two when Hiroshi saw Masao ...'

But although examples like these show that a listener may need to treat what he initially took to be a matrix clause as an embedded clause, if we look for the kinds of local structural ambiguities that are typically studied in English, in which there is more than one grammatical transition from the current parse state to the next parse state, we find that these are restricted to a narrow class of cases.

In English, most structural ambiguities fall into one of two categories, as we have seen in §3.3.1. First, there are the *association ambiguities* discussed in §3.3.2–3 in which an incoming word or phrase can associate to more than one existing node. Second, there are the *(in)dependence ambiguities* discussed in §3.3.4, in which the parser must decide whether or not to establish a grammatical relation between the last word/phrase and the incoming material. This was also the choice to be made in the case of PP attachment in German verb-final clauses. Some examples are repeated in (83).

- (83) *Two types of structural ambiguity (mostly English)*
- a. *Association ambiguities*
object–relative clause (VP, NP)
adverbial attachment (VP1, VP2)
PP-attachment (VP, NP, N)
main verb–reduced relative (NP, IP)

Parsing

- b. *(In)dependence ambiguities*
matrix–relative (modification)
object–subject (thematic)
PP-attachment in German verb-final clauses (modification)

In English, much of the difficulty in characterizing structural parsing preferences arises in connection with association ambiguities (83a), as we have already seen. In some situations association with the earlier site is structurally preferred, in other situations association with the more recent site is preferred. Characterization of the (in)dependence ambiguities, on the other hand, is rather simple: if there is a possibility of establishing a relation between the new word and the most recent word/phrase, this possibility is always taken. Numerous theories are able to capture this.

Japanese, on the other hand, only allows one of these ambiguous situations, as far as I can tell, namely (in)dependence ambiguities in which the incoming word or phrase has to choose whether to associate to the most recent phrase or not. This kind of ambiguity arises all the time, but there seems to be little question of how it is resolved. A relation is established where possible.

In order for Japanese to show an association ambiguity, it would need to be the case that there was a string $XP \dots YP Z^0$, in which the head Z must take one of XP or YP as an argument. In order for the possibility of Z associating with XP and not YP to arise, it would need to be the case that XP had scrambled out of the domain of Z . However, due to the impossibility of scrambling nominative-marked phrases and of scrambling out of relative clauses in Japanese, the possibilities for an association ambiguity configuration are severely restricted.

Consider a simple sequence of NPs followed by a verb, as in (84: adapted from Inoue & Fodor 1995).

- (84) Bob-ga ringo-o tabeta
B. -nom apple-acc ate

Let us assume that the parser has reached a point where it has already analyzed the two NPs as independent argument phrases, and it is deciding what to do with the verb. One possibility would be to treat the verb as the predicate that selects both of the argument phrases, thereby creating a possible complete sentence (85a). Alternatively, the parser could decide to treat the verb as part of a relative clause, with the consequence that one or more of its arguments is phonetically null, and therefore one or both of the NPs is not an argument of the verb, as would be consistent with the continuations in (85b–c).

- (85) a. Bob-ga ringo-o tabeta
B. -nom apple-acc ate
'Bob ate the apple.'
- b. Bob-ga ringo_i-o [[pro_i pro_j tabeta] inu_i-ni] ageta.
B. -nom apple-acc ate dog-dat gave
'Bob gave the apple to the dog that ate it.'

Parsing

- c. Bob-ga [[pro_i ringo-o tabeta] inu_i-ni] hone-o ageta.
B. -nom apple-acc ate dog-dat bone-acc gave
'Bob gave the dog that ate the apple a bone.'

Not surprisingly, speakers of Japanese seem to adopt the monoclausal analysis in (85a) over the more complex analyses in (85b–c). Inoue & Fodor (1995) cite as evidence for this the fact that speakers report a 'surprise' effect when a verb is used which cannot select all of the arguments that precede it, as in (86).⁹⁴

- (86) Bob-ga Mary-ni ringo-o tabeta
B. -nom M.-dat apple-acc ate

In general, an (in)dependence ambiguity arises in Japanese whenever a head is encountered that potentially takes the preceding word or phrase as an argument or modifier. As in English, the empirical generalization appears to be that dependencies are established wherever possible.

But Japanese is different from English in the respect that the parser never needs to decide which of two preceding words or phrases to associate to, and this is the area where the action is in the literature on ambiguity resolution in English.

3.4.2.3 Japanese syntactic ambiguity resolution

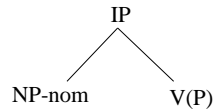
Given the limited scope of structural ambiguities in Japanese, there is not very much that needs to be said about structural ambiguity resolution of the kind studied in English, involving choices between different *grammatical* transitions between parse states. If it is correct that speakers of Japanese do build dependencies between adjacent words where possible, then this is quite consistent with Branch Right, though the generalization may be easily captured in a variety of approaches.

Consider the following two analyses of an NP-V sequence. In the first a selection relation is established between the verb and the NP, in the second they are independent, because the verb is analyzed as part of a relative clause with a null subject.

⁹⁴ Mazuka & Itoh 1995 dispute Inoue & Fodor's interpretation of this surprise effect as evidence for early commitment in parsing decisions. They show that a variety of factors can modulate the strength of the surprise effect, which are not expected under the early commitment analysis. However, while the additional variance that Mazuka & Itoh point to is not explained by Inoue & Fodor's theory, their account of (86) remains the most adequate.

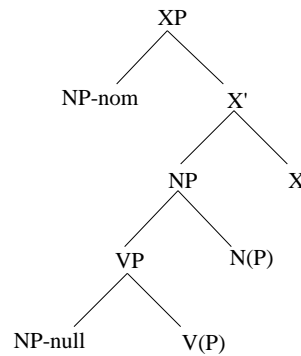
Parsing

(87) a.



2: NP-V(P)

b.



5: NP-X'-NP-VP-NP
2: NP-V(P)

The diagram given for (87b) is likely to be a conservative assessment of the complexity of the relative clause analysis based on the Branch Right metric, given that I have ignored any CP and IP nodes in the structure of the relative clause. But whatever the structure of the relative clause, there is sure to be a complexity difference between establishing a dependency and failing to do so.

3.4.2.4 Recovery from error

As we have seen, Japanese lacks the kind of association ambiguities which have generated most of the interest in structural ambiguity resolution in English. But of course this does not mean that a parser for Japanese is never faced with a range of choices. Such situations are easy to construct, but they all involve reanalysis situations, in other words, situations in which the parser has more than one way of parsing the incoming word or phrase, but none of these involve grammatically defined transitions from the immediately preceding parse state. However, reanalysis is where the action is to be found in the literature on Japanese syntactic ambiguity resolution.

Examples of the kinds of facts which have drawn a lot of attention in the literature involve examples like (85b-c) above in which what at first seemed to be a main clause needs to be reanalyzed as a relative clause, and possibly the selection relation between the verb and one of the NPs needs to be withdrawn.

Inoue (1991) claims that in these situations speakers prefer to include as much of the existing material as part of the relative clause, and calls this the 'Minimal Expulsion Strategy'. Inoue claims that it is not difficult to 'expel' one of the NPs from the relative clause, but that speakers find it rather difficult to expel more than one NP. Examples with zero, one and two expelled NPs are shown in (88a-c).

Parsing

- (88) a. [Mary-ga sinseihin-wo $t_{loc/i}$ kaihatu_{sita}] kaisy_a-ga tubureta.
M. -nom new product-acc developed company-nom went bankrupt
'The company where Mary developed the new product went bankrupt.'
(Inoue 1991)
- b. Yamasita-ga [$t_{nom/i}$ yuuzin-wo houmonsita] siriai_i-ni tegami-wo kaita.
Y.-nom friend-acc visited acquaintance-dat letter-acc wrote
'Yamashita wrote a letter to an acquaintance who visited his friend.'
(Sturt & Crocker 1996)
- c. # Yamasita-ga yuuzin-wo \emptyset_{nom} $t_{acc/i}$ houmonsita] kaisy_a-de mikaketa.
Y.-nom friend-acc visited company-loc saw
'Yamashita saw his friend at the company he visited.'
(Sturt & Crocker 1996)

A variety of different theories have been proposed to account for this basic paradigm, or to show that it is inaccurate. I will not review these here, but instead refer the reader to the papers in Mazuka & Nagai 1995.

The observations relevant to my proposal are the following.

First, since Branch Right only evaluates the path from the most recent to the incoming word, it cannot distinguish between the different scenarios in (88a–c). This is because in each of these cases the relevant path is from the verb to the head of the relative clause.

Second, it is not clear that I should want Branch Right to account for recovery from error. I assume that Branch Right is a property of the system that generates and parses sentences in a single left-to-right pass, and that reanalyses requiring backtracking are handled by other mechanisms. Additionally, the accuracy of the paradigm in (88) has been questioned by some researchers in the field (e.g. Mazuka & Itoh 1995).

The aim of this section was to clarify what predictions Branch Right makes about structural ambiguity resolution in head-final constructions. The special relevance of such constructions to Branch Right is that they induce a limited form of left-branching structure, and therefore might be seen as problematic for a parsing principle which favors construction of maximally right-branching structures. What I hope to have shown in this section is first that head final constructions only entail left-branching structures in a rather limited sense, and second that structural ambiguities involving choices between different grammatical transitions to a new parse state are rather limited in head-final constructions. They are easily handled by Branch Right, but also by a good number of other approaches to ambiguity resolution.

3.5 Some residual questions

3.5.1 Fragility of the predictions of Branch Right

This section discusses the concern that Branch Right is too fragile to properly account for the various parsing preferences in English in which an argument attachment is chosen over a relative clause.

Specifically, for any case in which I argue that Branch Right avoids building a relative clause at some point of ambiguity, it ought to be possible to increase the complexity of the preceding NP such that Branch Right will predict the relative clause to suddenly become preferred. For example, consider the following examples from Gibson 1991, which he brings up in a discussion of exactly this issue.

- (89) a. The horse that John saw beside the dog raced past the barn fell.
 b. The canoe that Susan observed beside the sailboat floated down the river sank.
 c. The dog that was fed next to the cat walked to the park chewed the bone.
- (90) Susan bought a house with two bedrooms, a garage, and a fireplace for Mary.

Gibson points out that the examples in (89) are still garden paths, despite the fact that locality should now predict the ambiguous verb to be preferentially attached as a reduced relative modifying the immediately preceding NP. (90) is also claimed to still show an argument attachment preference (Milsark 1983 makes the same claim; Frazier & Fodor 1978 make the opposite claim).

If there is a structural effect that overrides variations in locality here, then this is a serious problem for my claim that Branch Right subsumes the work of principles like Minimal Attachment or attach-as-argument.

However, I think that these examples need to be more carefully controlled in order to run the relevant test. For example, my intuition is that some counterparts to (90), do show a reversal of preferences when the NP is made more complex, as in the pair in (91), also from Gibson 1991.

- (91) a. I put the book on the floor onto the table.
 b. I put the book, the pencil and the eraser on the floor onto the table.

My intuition is that the PP *on the floor* is initially read as an argument of the verb in (91a), but not in (91b).

With respect to the examples in (89), these examples all contain ambiguous verbs which show a massive bias against the reduced relative reading (see the discussion in MacDonald et al. 1994 for the relevant factors that contribute to verb bias). In order to control for the interfering effect of verb bias, it is important to use verbs which are not massively biased towards one reading or another. For these verbs I then predict that increasing or reducing the

complexity of the subject NP in examples like (89) should lead to a reversal of interpretive preferences. This test has not yet been run.⁹⁵

3.5.2 One more anti-locality effect: complex NPs

Probably the best challenge to the claim that Branch Right is the only structural measure in ambiguity resolution comes from the existence of apparent anti-locality preferences in complex NPs in a number of languages. Here I briefly discuss the status of the issue.

Although English speakers show a preference to attach the relative clause in complex NPs like (92) to the more embedded NP, a number of studies have shown that this is not the case in superficially identical constructions in a number of other languages (Spanish: Cuetos & Mitchell 1988, Gilboy et al. 1995, Gibson et al. 1996ab; German: Hemforth et al 1996; Dutch: Brysbaert & Mitchell 1995).

(92) The daughter of the colonel who I met....

The preference of English speakers to attach the relative clause to the more embedded NP is weak, but replicable. Similarly, the strength of the high attachment preference in speakers of other languages is not overwhelmingly, but it has been confirmed in a number of studies by different groups.

There have been a number of suggestions for why the cross-linguistic difference exists (e.g. Mitchell & Cuetos 1991, Gilboy et al. 1995, Gibson et al. 1996a, Sauerland 1996), but this question is still very much open. Similarly, the question of what underlies the high attachment preference is still open, though progress has been made in this area. For example, Gibson et al. 1996ab and Hemforth et al. 1996 have shown for Spanish and German respectively that in more complex versions of (92) containing three levels of NP embedding Spanish and German show effects of a low attachment preference. Gibson and colleagues have also found evidence for a weak anti-locality effect in English. These findings indicate that the cross-linguistic difference is not so categorical as the initial findings about Spanish suggested.

However, this is the most compelling case that I am aware of of a structural preference that is unlikely to be reducible to Branch Right. If Sauerland (1996) is correct that the reason for the high attachment effects is a case matching preference, then it may nevertheless be possible to account for the structural component of how this ambiguity is resolved in terms of complexity measures which are independently justified in the grammar. But this question still awaits resolution.

⁹⁵ Another factor which might make it difficult to properly test constructions like (89) involves the fact subjects often treat NP-V sequences as subject-verb sequences, even when this is ungrammatical, as for example in **the horse raced past [the barn fell]*. This means that subjects might (incorrectly) entertain an analysis of examples like (89) in which the verb takes the NP embedded inside the subject NP as its subject.

3.6 Modularity

An enormous amount of experimental work in sentence processing has been devoted to issues of modularity, i.e. to the question of how independent sentence processing is of other cognitive processes.

The received view for a long time was that the sentence processor was a special purpose module of the mind which, at least at the initial stages of parsing incoming material, operated independently of other cognitive processes and many sources of potentially relevant information. This view was influenced by Fodor's (1983) notion of modules of mind, and it received a good deal of empirical support. Structural complexity metrics played an important role in this approach. See Frazier & Fodor 1978 and Frazier 1987 for classic formulations of this position.

More recently, there have been a number of challenges to the classical view of a relatively autonomous syntactic processor, often under the rubric of *constraint-based models* (e.g. MacDonald et al. 1994, Trueswell & Tanenhaus 1994). A large number of studies have provided evidence for a range of non-syntactic factors which affect the course of syntactic ambiguity resolution. This line of work has been taken to challenge the classical 'modular' view of parsing, and in doing so has questioned a number of assumptions that were standard in earlier models. One of the assumptions that has been questioned is whether there is any motivation at all for the existence of structural complexity measures in an account of sentence processing. I therefore take this opportunity to briefly comment on what I think the status of the modularity question is, and how I think it affects approaches to parsing of the kind that I have been developing in this chapter.

Most importantly, we must distinguish between two quite different claims which are combined in Fodor's notion of modularity. One claim is that there are many distinct modules for many distinct mental functions; in other words, this is a claim about *individuation*. A quite separate claim is that at least some of the modules are highly *encapsulated* systems, which share relatively little information with other systems. These are quite separate claims, so it is unfortunate that they have often had to share the same term *modularity*.

Most of the recent challenges to modularity in sentence processing have addressed the issue of encapsulation of information processing, and have argued that information processing is much less encapsulated than was assumed in models like Frazier's Garden Path model. Very rapid effects of context, plausibility and lexical frequency have been found, typically in experiments which show that classic garden path effects can be reduced or eliminated by using differently biased lexical items.

However, we cannot draw any conclusions about the individuation of a language processor or structure building system based on findings about reduced encapsulation of information flow. The issues are not related. This also means that evidence for individuation of language or different subsystems of language is irrelevant to the validity of evidence against encapsulation.

I assume here that the syntactic parser/grammar may be individuated as a distinct mental faculty, but that it is not necessarily an encapsulated information processor. In this view it is quite consistent for there to exist structural complexity metrics which play a role in ambiguity resolution and for there also

to be very rapid influences of non-structural information in ambiguity resolution. From this perspective, one of the most important lessons of the work on constraint-based models is that if we want to investigate structural complexity metrics in parsing we need take extreme care to control for the variety of other factors that have rapid effects on parsing.

The question of how successful the arguments for non-encapsulation have been addressed elsewhere (cf. Frazier 1995), so I will not discuss them here. However, the question of encapsulation is not of primary importance to my goals in this thesis. It is the individuation of language and of the subparts of language that I am most interested in. The evidence for the individuability of language is good, based on a number of observed dissociations in development, brain damage and brain recording (see Pinker 1994 for a survey of the most important evidence). The evidence for the individuability of different subsystems of language for grammar and parsing is what I am reexamining here.

Therefore, although I am sceptical about whether constraint-based models have succeeded in arguing against an indivisible syntactic processor, because work in these frameworks has focused on encapsulation rather than individuability, my goal of arguing that there is not a special purpose syntactic processing system is very much similar to the stated goal of much of this research.

3.7 Conclusion

In this chapter I have attempted to show that the structural contribution to ambiguity resolution may be in large part reducible to a single principle, Branch Right. Branch Right is an economy principle of the grammar that was motivated on independent grounds in Chapter 2. This suggests that the structural component of ambiguity resolution in parsing is nothing more than the application of the economy conditions of the grammar. This point will be discussed further in Chapter 5.

I should clarify a little further at this point how the main point of this chapter fits in with the overall aim of the thesis. I have argued here that a version of the local attachment preference has far greater empirical coverage than is usually assumed, and that it makes certain other structural ambiguity resolution strategies redundant. How crucial is this claim to the general claim that the parser is the grammar? In fact, it is possible for the main claim of this chapter to be wrong, but for the overall claim about the parser–grammar relation to still be correct. All that is required in order for the parser and the grammar to be identical is that there not be distinct structural complexity metrics for parsing and grammar. The claims made here about the generality of Branch Right satisfy this, but there are presumably many other ways of meeting this criterion. Nevertheless, the fact that this particular complexity metric has such broad coverage is of inherent interest, both because a local attachment preference is so widely assumed, and because it leads to a simpler view of structural complexity measures in parsing.

Appendix: Experimental Materials

The following example shows how to read the 4 conditions of experiment 1 from the each sentence below. The conditions are described in more detail in §3.2.

MA		the lawyer	—
MU	While I talked with	his lawyer	—
RA		the lawyer	—
RU		the lawyer	that

		him		and that made me rather nervous
John was watching		him	at the party	and that made me rather nervous
		—		I became rather nervous
		—		I became rather nervous

1. While I talked with the lawyer John was watching (him) at the party I became rather nervous/and that made me rather nervous.
2. When Mike arrived at the house I described (it) in great detail he was quite impressed/because he was interested in old building styles.
3. Until I trained the dog everybody was avoiding (it) like the plague I had very few visitors/because they were afraid of getting bitten.
4. After I watered the plant the housekeeper neglected (it) for many months it slowly recovered/and it died.
5. Because Rose praised the recipe I made (it) for her birthday it was worth all the effort/as a surprise.
6. Although the boy was afraid of the dog he annoyed (it) in the park he tried to remain calm/with his friends.
7. Because Joe liked the children he saw (them) on the weekends he found it hard to leave/whenever he could.
8. Although I liked the flowers the janitor removed (them) from the office I didn't get angry/while I was away.
9. As the president outlined the speech his advisor drafted (it) for the convention he realized how difficult his task would be/that was fast approaching.
10. As the king's army lost control of the castle the enemy was destroying (it) with huge cannons some reinforcements arrived/and closing in on where the king hid.
11. When the collector displayed the painting the expert identified (it) as a fraud the police were contacted/within minutes.
12. When the chef served the dessert the guests liked (it) quite a lot the treaty was quickly settled/and told him so.
13. Because the senator proposed the bill the speaker opposed (it) in the house he resolved to unseat the speaker in the election/and attacked the senator on TV.
14. Since the author's friends bought all the books the publisher distributed (them) to major booksellers the novel made the best-seller lists/and made enormous profits.
15. Because Fred was unhappy about the car he sold (it) to the dealer at half price he never bought the same make of car again/to get rid of it.

Parsing

16. Although the dissident eventually destroyed the letters he had kept (them) beneath the floorboards it pained him to do so/for months.
17. When the boys discovered the canoe their dad was fixing (it) in the garage they knew what they were getting for Christmas/as a Christmas present.
18. By the time the girls found the gifts their mother had wrapped (them) for the party they were quite late/and it was time to leave.
19. When the company board promoted the accountant John disliked (him) for his success John resigned/and his wealth.
20. As the residents looked at the building the crane was demolishing (it) with a wrecking-ball some of them couldn't help applauding/and leaving only rubble.
21. Although Gwen and Phil were proud of the vase they had obtained (it) at a garage sale it wasn't really valuable/sale for five dollars.
22. Because the police didn't trust the witness they had interrogated (him) for five hours they ran a lie detector test/before the trial.
23. Since the students had enjoyed the classes they completed (them) with no trouble they encouraged their friends to take the same classes/and got good grades.
24. Because the conference organizers liked the caterers they hired (them) for the banquet they asked them back the following year/in honor of the former president.

Chapter 4

Morphosyntax

4.1 Introduction: Composition and Decomposition

The previous two chapters have investigated some consequences of building syntactic structures from left-to-right, rather than in the more conventional bottom-to-top fashion. Chapter 2 showed that the syntactic units present in left-to-right derivations are sometimes rather different from the syntactic objects present in conventional derivations, and that this provides a different perspective on problems of contradictory constituency—*left-edge constituents* and *right-edge constituents* correspond on this approach to constituents at different points in a left-right derivation. Chapter 3 showed that the structural complexity metric Branch Right that is used in real-time parsing may be the same as a complexity metric used to account for grammaticality facts in Chapter 2, indicating that parsing and grammar may have more in common than is typically assumed.

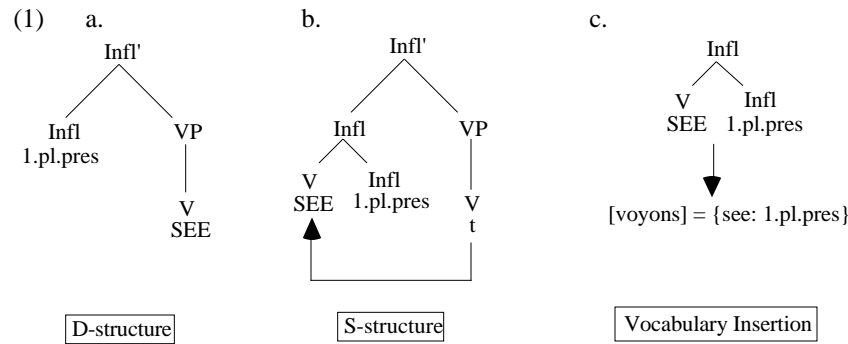
In this chapter I outline a derivational approach to morphosyntax and head movement which interfaces with the left-to-right structure building theory of Chapter 2, and thereby reverses the derivational ordering of morphological-phonological processes and syntactic processes that is assumed in a number of current theories. I present an account of local and non-local head movement and of clitic placement in the left-to-right approach, and a case study of the relation between intonational phrasing and syntactic constituency. A number of the results in this chapter depend specifically on the left-to-right character of the derivations I assume. Some other results that arise in the discussion of left-to-right morphosyntax are not in fact dependent on this ordering of operations. I will be explicit about which results do and do not require left-right derivations.

As an illustration of what I mean by ‘reversing’ the normal order of derivations, consider how an inflected French verb like *voyons*, ‘we see’ is derived, first in the transformational approach of Halle & Marantz 1993 (which builds on earlier proposals by Pranka 1983, Pesetsky 1985, Baker 1988, Pollock 1989, Chomsky 1991), and then in the inverted approach.

In standard transformational derivations an inflected verb arises out of the syntactic combination of lexical and inflectional heads which at the initial syntactic level of D-structure head their own separate syntactic projections (1a). In syntax these heads are just bundles of grammatical features, and are not yet associated with any phonological form. The heads are combined by upward (and usually leftward) movement of the verb to adjoin to the inflectional head, thereby

Morphosyntax

creating a complex syntactic head at S-structure (1b). This complex head is then shunted off to the morphological component of the grammar (1c). At this stage a phonological form is chosen which best matches the morphosyntactic features of the complex head, and this form is used to spell-out the syntactic features.⁹⁶



Government-Binding model of Morphosyntax
(Halle & Marantz 1993)

The sequence of operations shown in (1) is impossible in the left-to-right approach to syntactic derivations proposed in Chapter 2. Since rightward movement operations are the only movement operations permitted by Merge Right, head movement operations like (1b) are ruled out.⁹⁷ Given that most of the operations involved in head-composition in transformational models of inflection involve leftward movements just like (1b), Merge Right therefore forces us to adopt a rather different approach to morphosyntactic derivations from the model illustrated in (1).

The approach to morphosyntax that I explore in this chapter takes the inverted syntactic building and movement model of Chapter 2 one step further, and runs through the steps of (1) more or less in reverse. The derivation begins with an inflected phonological form like *voyons*, for which a best-match set of syntactic features are selected (2a).⁹⁸ This set of syntactic features is converted into a hierarchically organized syntactic head (2b) and input to a larger syntactic structure (2c). Subsequently, it is decomposed into its pieces by means of a rightward movement operation, which takes the verbal head from its position

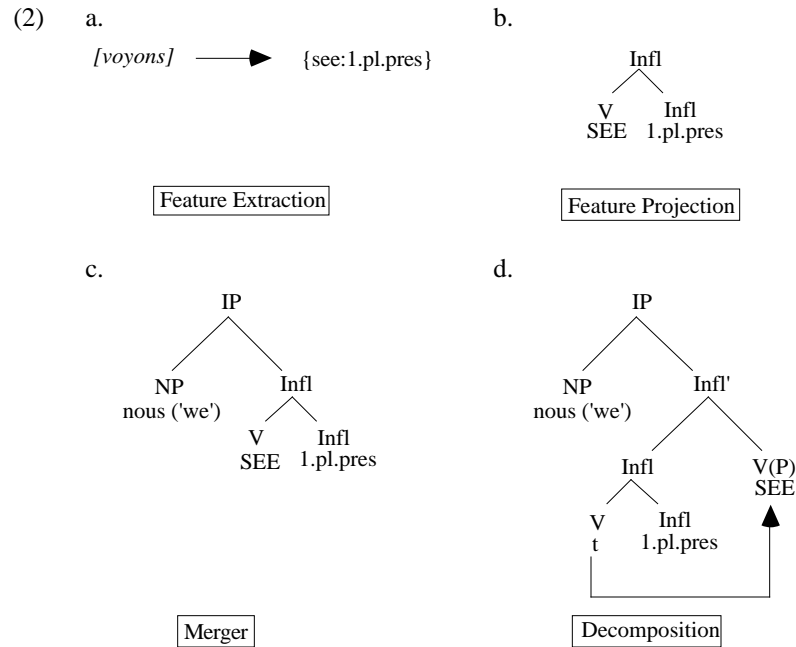
⁹⁶ This characterization of the mapping from syntax to morphology is an oversimplification, and abstracts away from a number of features of the Halle & Marantz 1993 *Distributed Morphology* model which are involved in the interfacing of phonological and syntactic forms.

⁹⁷ I assume that 'LF-movement' is also rightward, along the lines suggested by Davis & Alphonse 1992, Alphonse & Davis 1996.

⁹⁸ Throughout this chapter I will be assuming that words (or phonological words) are the units that are the 'inputs' to left-right derivations. However, little turns on this. In particular, I am not assuming that structure building internal to the word is different from structure building above the word level. Once the various morphosyntactic components of a word have been extracted from the phonological form, these pieces combine in a left-to-right fashion.

Morphosyntax

adjoined to Infl and puts it in a position in which it heads its own syntactic phrase (2d)



Decompositional morphosyntactic derivation

In what follows I will refer to the approaches in (1) and (2) as *compositional* and *decompositional* approaches respectively.⁹⁹ The different derivational orders adopted in (1) and (2) matters very little to the treatment of forms like the French inflected verb *voyons*.¹⁰⁰ However, I hope to show here that the reversal of the order of morphosyntactic derivations does not simply reproduce all of the processes and generalizations of the compositional approach in reverse. Changing the feeding relations among the different derivational stages has a number of positive consequences for how morphophonological processes are characterized, as this chapter endeavors to show.

The consequences of decompositional morphosyntax which are discussed in this chapter include the following:

⁹⁹ A decompositional approach to morphosyntax is proposed by Kitagawa 1986 in the context of the more traditional bottom-to-top approach to phrase structure. Kitagawa's approach to morphosyntax in Japanese is quite compatible with the approach presented here.

¹⁰⁰ This remark is true of verbs which undergo 'V-raising' in the sense of Emonds (1976, 1977) and Pollock (1989), but the derivational order does make a difference to cases involving 'Infl-lowering', such as we find in English. See §4.4.3 below for more discussion of this issue.

- In §4.2 and §4.3 I give arguments for replacing the Head Movement Constraint (Travis 1984, Baker 1988) with a requirement that chain links stand in a c-command relation. I show that this can explain the difference between contexts where head movement must be strictly local, and contexts where head movement need not be strictly local. Head decomposition processes must be strictly local, whereas head movement processes need not be strictly local.
- §4.4 shows how a number of clitic placement phenomena which have often been taken to defy syntactic analysis can be unified with more familiar head movement processes.¹⁰¹ This removes one important motivation for the existence of separate transformational operations in morphology and syntax.
- In §4.5 I look at phenomena which have been claimed to show that phonological processes refer to constituent structures which are different from those motivated by the syntax. I show how these claims need to be reconsidered in the light of the constituency present in left-to-right derivations, based on a case study of intonational phrasing in Tohono O'odham (Papago).

4.2 Movement and Projection of Heads

4.2.1 The C-command Condition on Movement

This section and the next are concerned with questions of locality in head movement. The main goal of these sections is to motivate an alternative to the widely adopted Head Movement Constraint, namely the C-command Condition on Movement (CCM), and to show both how it applies in a left-to-right decompositional framework, and how it accounts for where locality restrictions are and are not found in head movement.

I assume a general syntactic condition on head movement processes which requires that for any two positions linked by a movement operation, one of those positions c-commands the other (3). Contrary to standard assumptions I assume that it does not matter which of the two positions in the chain is the c-commander and which of the two is c-commanded.

- (3) *C-command Condition on Movement*
When two positions are linked by a movement operation, one of the two positions must c-command the other.

For the purposes of this condition I adopt a slightly non-standard definition of c-command. In addition to the standard definition of c-command (4a), I

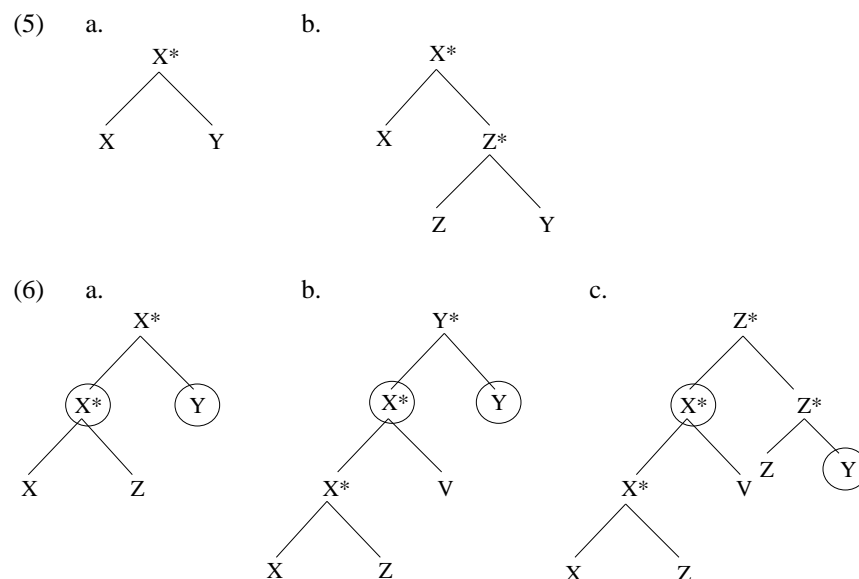
¹⁰¹ That clitic placement and other kinds of syntactic incorporation or affixation form a unified class was assumed in some earlier transformational approaches to head movement (e.g. Pranka 1983, Marantz 1984), but this has generally not been assumed since the work of Baker (1988), which assimilated incorporation and affixation to movement of entire heads rather than to clitic placement.

Morphosyntax

assume that a head X also c -commands a position Y if some projection of X c -commands Y (4b).¹⁰²

- (4) *C-command*
- a. X c -commands Y if every node that dominates X dominates Y , and X does not dominate Y . [i.e. X c -commands its sister and everything dominated by its sister.]
 - b. X c -commands Y if some projection of X c -commands Y .

This means that in addition to the standard cases of c -command between X and Y illustrated in (5a–b), X c -commands Y in the configurations shown in (6a–c), in which the node immediately dominating the head X does not dominate Y , but a projection of X c -commands Y . The relevant projections of X which c -command Y according to definition (4a)—thereby allowing the head to c -command Y according to definition (4b)—have been circled in (6a–c) for ease of reading.

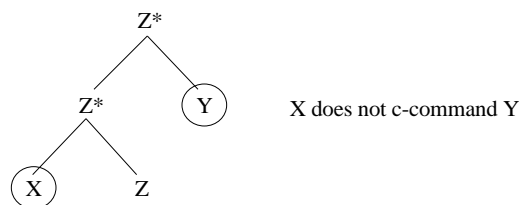


This modified version of c -command should not be confused with the often invoked relation *m-command*, which holds between X and Y when the first maximal projection node dominating X also dominates Y (and X and Y do not stand in a domination relation). (7) illustrates a configuration in which X m -commands Y , but X does *not* c -command Y under the modified definition assumed here. X itself is the highest projection of X , and the node immediately dominating X does not dominate Y .

¹⁰² As Jonathan Bobaljik (pc) points out, the recursive part of this definition (4b) may follow automatically if we assume that XP is nothing more than a projection of the features of X^0 , in which case the nodes are identical.

Morphosyntax

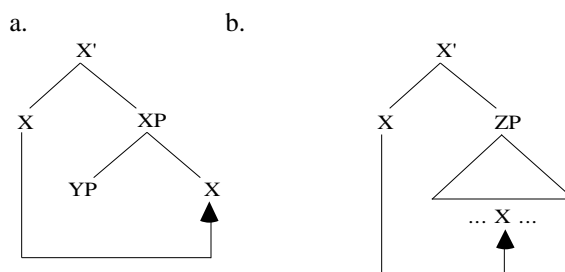
(7)



Now consider the range of possible head movement operations that satisfy the c-command condition on movement (CCM). The effect of the CCM is to impose a strict locality condition on movement operations which separate or join the subparts of complex heads. It does not impose the same strict locality condition on head movement operations involving entire X^0 s. Note that all of the movement operations here are rightward movements, as required by the condition Merge Right, which was motivated at length in Chapter 2.

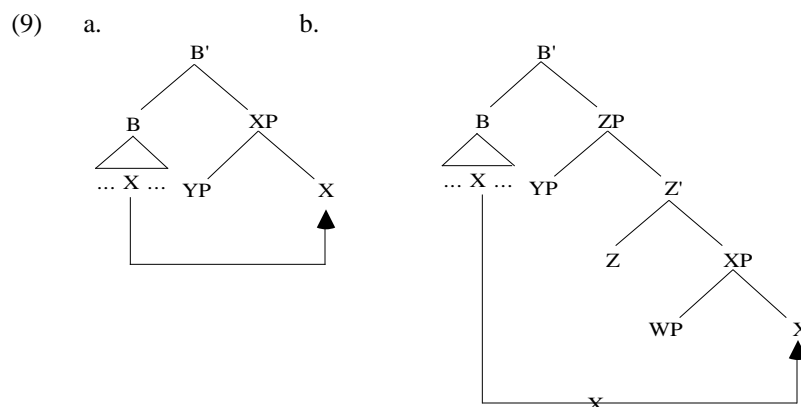
(8a–b) show how both local and non-local lowering of complete X^0 s is possible, either to the closest head position (8a), or across intervening head positions, as in (8b) in which at least one other head Z intervenes between the two positions occupied by X. X may in principle lower across an arbitrarily large number of intervening heads, because its initial position always c-commands its landing site.

(8)



(9a–b) show cases of head movement which parallel (8a–b) except that they involve the movement of a subpart of a complex X^0 . The most important difference between these examples and the examples involving movement of a complete X^0 is that the position in which the movement originates does not c-command the landing site of movement. This is because the head that moves originates in a position in which it is immediately dominated by a (zero-level) projection of a different terminal element, and therefore it is unable to c-command any node outside the complex X^0 . Given that the source of movement does not c-command the target position, the only kind of movement operation that will satisfy the c-command condition is one in which the target position c-commands the source position. The target c-commands the source if the movement is a strictly local lowering (9a), because the highest projection of the target c-commands the source, but not if the movement is a non-local lowering (9b), because of the intervening projection of Z.

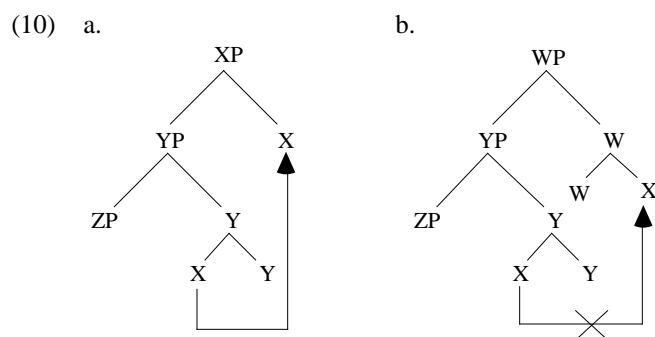
Morphosyntax



Note that I crucially do not assume that a category/segment distinction plays any role in the definition of c-command (contra May 1985, Chomsky 1986). If I did assume this distinction it would become possible for a head to c-command out of the complex head that it is contained in, and I would lose the account of locality conditions on head movement developed in this section and the next.

Therefore, as far as head lowering operations are concerned, those which move an entire X^0 are predicted to be able to move non-locally, whereas the decomposition operations which move a subpart of a complex X^0 are predicted to be strictly local (with respect to the head that is being decomposed).

Head 'raising' operations are predicted to show a similar contrast in locality requirements, in this case depending on whether the target of movement is an entire X^0 or is adjoined to an existing X^0 . Creation of new X^0 s should be possible non-locally, whereas adjunction to existing X^0 s should be strictly local. In the case of raising, though, it should not matter whether the source position is an entire X^0 or not. The examples in (10) illustrate. In (10a) the head X moves out of the head Y to a position which c-commands the source of movement. In (10b) X moves to a position adjoined to the head W , with the effect that neither end of the movement chain c-commands the other.



The upshot of this discussion is that if we assume the requirement in (3) that chain links stand in a c-command relation, then we predict that the locality requirement imposed by the Head Movement Constraint in many theories should

be descriptively true of just those processes which disassemble complex heads, but it should not be descriptively true of head movement operations in general.

- (11) *Locality of Head Movement* (corollary of CCM (3))
Operations which move entire X^0 s may be non-local, but operations involved in the assembly and decomposition of complex X^0 s must be strictly local with respect to X^0 and the landing site.

The goal of §4.3–§4.4 is to show that this prediction is accurate, and that left-to-right head movement respecting the c-command condition on movement makes it possible for processes that have often been claimed to defy syntactic analysis to be assimilated to normal head movement processes, while still providing the constraints that the Head Movement Constraint was designed to provide.

4.2.2 Head Projection

The following is a sketch of the conditions under which I assume that an element can and cannot be the head of a syntactic projection. The basic assumptions are:

Syntactic heads are collections of features, some of which can project syntactic phrases, others of which cannot.

Languages vary in which features they allow to project. Presumably all languages allow the verbal features of verbs to project, but there are other features which may or may not be allowed to project. For example, only some languages may allow a verb endowed with focus features to project the focus features to create a FocP projection. This assumption plays an important role in §4.3.2 in the discussion of local and non-local participle fronting.

Finally, features that cannot project can only occur as part of some larger head that can project.

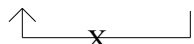
These assumptions are fairly standard, although they are not always made explicit.

4.3 (Non-)Locality in Head Movement

Since the work of Travis (1984) and Baker (1985b, 1988) it has standardly been assumed that head movement processes are all subject to a strict locality condition known as the *Head Movement Constraint* (12), which blocks head movement to all but the most locally c-commanding head position. This means, for example, that it should be impossible to move head X to adjoin to head Z in a configuration like (13), in which another head Y intervenes between X and Z.

Morphosyntax

- (12) *Head Movement Constraint* (Travis 1984, p. 131)
 An X^0 may only move into the Y^0 which properly governs it.

- (13) * Z + X_i Y t_i


The literature contains a number of cases of head movement processes which appear to violate the Head Movement Constraint. These ‘violations’ may be separated into locality violations and c-command violations. The locality violations involve head movements which target a c-commanding head position, but not the most locally c-commanding head position. (14a–b) illustrate this with examples of participle fronting across auxiliary verbs in Breton (Borsley, Rivero & Stephens 1996) and Bulgarian (Rivero 1996). In both examples, the sentence initial participle is separated from its underlying position by two other heads.

- (14) a. Kavet am eus bet al levr.
 found fs.pf pf.prt the book
 ‘I had found the book’ (Breton)
- b. ꞑetjal su^om bit knigata.
 read pf.1s pf.prt book.the
 ‘(According to someone,) I am reading the book.’ (Bulgarian)

There have been a number of attempts to modify the Head Movement Constraint in such a way that limited non-local head movement is allowed (e.g. Rivero 1991, 1994, Roberts 1992, 1994b).

Meanwhile, a number of other phenomena appear to violate that part of the HMC which requires movement to a c-commanding position. These cases, typically involving the positioning of clitic heads, have often been considered to be so different from more ‘normal’ head movement as not to warrant unification with the well-known instances of strictly local head movement.

(15a–b) show representative examples of heads which surface in positions which do not c-command their ‘underlying’ position, under most assumptions about where these heads originate syntactically. (15a) shows an example of Serbo-Croatian second position clitic placement in which the clitic cluster appears in the middle of the sentence initial NP (Browne 1974).¹⁰³ (15b) illustrates nominal clitics in Kwakwala, which suffix to the word that *precedes* the nominal that they agree with (Boas 1947, Levine 1980, Anderson 1984). The first line shows the phonological grouping of elements, the gloss shows the expected syntactic bracketing, based on which elements the clitics agree with.

¹⁰³ See §4.4.1 below for further discussion of clitic placement in Serbo-Croatian. There I discuss the possibility that examples like (15a) involve movement of the determiner as the result of ‘left-branch extraction,’ rather than positioning of the clitic cluster ‘inside’ the NP. The literature contains arguments showing that the left-branch extraction account of these examples faces serious difficulties.

Morphosyntax

- (15) a. Taj **mu-je** pjesnik dao autogram
this him-is poet give.ptc autogram
'This poet has given him an autograph.'
(Croatian: *avar & Wilder 1994)
- b. nɛp'idi-da gɛnanɛm=x.a gukw=sa t'isɛm (phonology)
throw-deic child [obj house] [obl rock] (syntax)
'The child hit the house with a rock by throwing.'
(Kwakwala: Levine 1980)

Given that examples like these appear to fall outside the bounds of normal syntactic head movement, it is often assumed that interfacing morphophonological structures like (15) with their syntactic representations requires either (i) a transformational morphological component, or (ii) a non-derivational theory of correspondence between morphophonological and syntactic relations. I argue here that neither of these conclusions is necessary in the left-to-right derivational theory proposed here: there is no distinction between syntactic and morphological transformational operations.

In this section I provide evidence that the Head Movement Constraint should be replaced in a left-to-right structure building theory with the C-command Condition on Movement (CCM) introduced in §4.2, and that the CCM's prediction of strictly local movement for head decomposition and non-local movement for movement of complete heads is accurate. §4.3.1 shows the role of the movement/decomposition contrast in clearing up some issues in the structure of complex VPs that I left open in Chapter 2, and shows how head movement processes interact with restrictions on argument structure. §4.3.2 shows the effects of the movement/decomposition contrast on 'long head movement' constructions across a variety of languages. §4.3.3 shows how the CCM preserves the generalizations that the HMC was designed to account for. Then in §4.4 I turn to cases of clitic placement like (15) and show how they may be handled by syntactic head movement in left-to-right derivations.

4.3.1 Movement and Decomposition inside VP

This section explores issues involving head movement inside VP, picking up some questions that were left unanswered in the discussion of extremely right-branching Cascade structures in Chapter 2, and argues that head movement within VP shows the limited locality properties predicted by the CCM. In addition to clarifying some questions about right-branching VP-structures, I show how it is possible to account for some restrictions on argument structure which Pesetsky (1995) attributes to the HMC, but without appealing to the HMC. I show that these restrictions on argument structure interact with Heavy Shift processes in a way which provides further support for the left-to-right ordering of derivations.

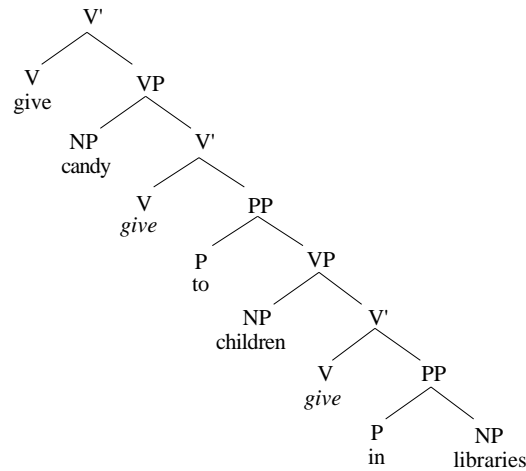
4.3.1.1 Two restrictions on causer arguments

Recall that I assumed that complex VP structures containing multiple arguments and adverbial modifiers have a final structure like (16), in which the V

Morphosyntax

head is represented by multiple copies and that non-final prepositional phrases like *to children* in (16) are ‘split’, with the effect that the NP complement of the preposition is not the sister of the preposition.

(16)



The following are some questions about these structures which were left unanswered in Chapter 2.

- How are thematic roles assigned to arguments? Do these structures place any constraints on which phrases receive which thematic roles?
- How can the verb be lowered across intervening P heads, without running afoul of widely assumed locality conditions on head movement?
- Structures like (16) incorporate properties of both Larson’s (1988) VP-Shell structures and Pesetsky’s (1995) Cascade structures. Pesetsky gives a series of arguments in favor of his structures over Larson’s, based on how they allow certain restrictions on argument structure to be reduced to the Head Movement Constraint. How are the empirical effects captured by Pesetsky to be captured in the hybrid structures assumed here?

Pesetsky’s arguments for his Cascade structures over Larson’s VP-Shells divide into two groups. The first set of arguments, involving the possibility of ‘PP-splitting’ effects in binding and coordination relations, have already been discussed in Chapter 2. The binding and coordination facts follow equally well under the modified Cascade structures adopted here, because the crucial property of Cascade structures for the account of binding/coordination is PP-splitting, which structures like (16) retain. The second set of arguments comes from restrictions on argument structure which Pesetsky argues can be accounted for by Cascade structures but not by VP-Shell structures.

Pesetsky discusses two restrictions on multiple complement structures. First, Oehrle (1976) observes a contrast between the range of interpretations

Morphosyntax

available the examples of double object and dative constructions in (17–20).¹⁰⁴ In addition to the normal reading of the double object construction in (17a), according to which Nixon is asserted to have performed an action, (17a) also allows an ‘idiomatic’ reading which asserts that Mailer wrote a book that he could not have written had it not been for Nixon. Oehrle identifies Nixon as the cause of Mailer having a book in this reading, rather than the agent. Most importantly, though, Oehrle points out that this causative reading is *not* available for the corresponding *to*-dative construction in (17b).¹⁰⁵

- (17) a. Nixon gave Mailer a book.
b. Nixon gave a book to Mailer.

Parallel contrasts between double object and dative constructions are shown in (18–20). Examples (19) and (20) show that when a subject is chosen that only allows the causative reading for the double object sentence, the corresponding dative construction becomes ungrammatical, as we would expect if the dative construction blocks causer subject readings.

- (18) a. Katya taught me Russian.
b. Katya taught Russian to me.
- (19) a. Talking to Bill for just a few seconds would have told Sue the answer.
b. * Talking to Bill for just a few seconds would have told the answer to Sue.
- (20) a. Hard work got Mary the prize.
b. * Hard work got the prize for Mary.

Pesetsky (1995: p. 60ff) demonstrates a restriction involving psychological predicates which is related to Oehrle’s observation, in that it involves a constraint on the expression of arguments bearing the Causer thematic role. Object experiencer predicates like *anger* and *worry* assign a Causer role to their subject (21), and subject experiencer predicates like *be angry* and *be worried* assign a Target or Subject Matter of Emotion (T/SM) role to a PP object (22). Given these possibilities, it is surprising that it is not possible to express both Causer and T/SM roles with the same predicate, as the examples in (23) show.

- (21) a. The article in the *Globe* angered Derek.
b. The blood clot worried Derek.

¹⁰⁴ The examples (17) and (18) are taken from Oehrle 1976 (p. 71); the examples in (19) and (20) are from Pesetsky 1995 (p. 194). See Green 1974 for discussion of related contrasts, such as (i–ii):

- (i) The thesis was giving John a headache.
(ii) * The thesis was giving a headache to John.

¹⁰⁵ See below for discussion of a class of environments in which the *to*-dative allows the causer-subject interpretation.

Morphosyntax

- (22) a. Derek was angry at the music columnist.
b. Derek was worried about his ability to play guitar.
- (23) a. * The article in the *Globe* angered Derek at the music columnist.
b. * The blood clot worried Derek about his ability to play guitar.

Moreover, the intended meanings for (23) are fully grammatical if a periphrastic causative construction is used instead of the object experiencer verbs used in (23).

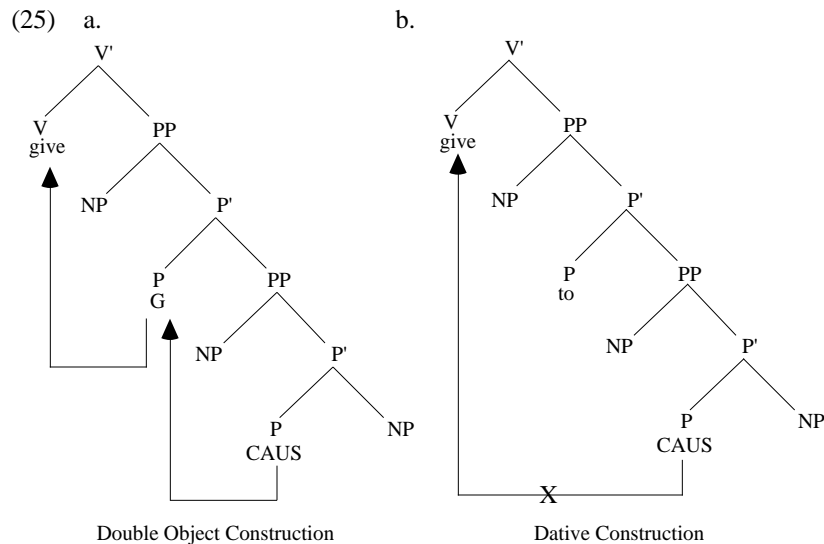
- (24) a. The article in the *Globe* made Derek angry at the music columnist.
b. The blood clot made Derek worried about his ability to play guitar.

Pesetsky points out that this restriction, which he dubs the *T/SM Restriction*, shares with Oehrle's observation the property that it involves a ban on subject arguments bearing Causer roles in constructions with a PP object. He shows that both restrictions can be ruled out as Head Movement Constraint violations on the assumption that the Causer role is assigned by a null head CAUS which originates in a position lower than the verb and must raise to adjoin to the lexical root of the verb in order to allow the verb to receive a causative reading.

In the absence of a PP object, the CAUS head is able to raise to adjoin to the lexical root head, in both object experiencer and double object constructions. For double object constructions, Pesetsky assumes that the second object is the sister of a null preposition *G*, rather than a copy of the verb, as it would be in Larson's VP-Shell structures. Pesetsky assumes that CAUS first adjoins to *G* on its way to *V*. (25a) illustrates the derivation of a double object construction: object experiencer constructions are derived in the same way, except that there is one less argument and no null head *G*.

In constructions with a PP object, however, CAUS is unable to raise to adjoin to the lexical verb because this would require long head movement across the overt *P* head. The illicit derivation is shown in (25b). In this way both Oehrle's observation and the *T/SM Restriction* are reduced to HMC violations.

Morphosyntax



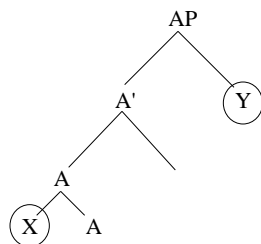
In order to show how I account for this I need to spell-out more explicitly how I assume that heads assign theta roles to their arguments. I assume that verbs which select for more than one internal argument consist of different pieces, which divide among themselves the task of assigning different theta roles. I assume that a head X can discharge one of its theta roles to position Y iff Y is the sister of a head containing X or the sister of a projection of a head containing X. I assume that the predication relation between adverbial phrases and the elements that they are predicated of must be established in the same structural configurations as theta role assignment.

- (26) *Configuration for Arguments and Predication*
 A head X may discharge a thematic role to a position Y or take position Y as a predicate iff Y is the sister of a head containing X or the sister of a projection of a head containing X.

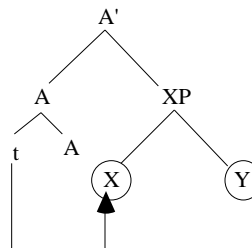
The principle in (26) allows an internally complex verbal head to assign theta roles to its arguments in either right-branching or left-branching VP structures (27), with no effect on interpretation. Following the assumption made in Chapters 2 and 3, though, I assume that when two structures are semantically equivalent the more right-branching alternative must be chosen. Therefore, I focus on right-branching VP-structures for the remainder of this discussion.

Morphosyntax

(27) a.



b.

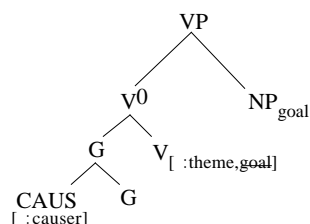


For concreteness, I assume the internal structure of multiple argument taking verbs that Pesetsky proposes, in which the verbal root originates in the highest head position in the VP and the additional argument assigning components of the verb (eg. G, CAUS) originate as heads in lower positions.¹⁰⁶ In left-to-right derivations, therefore, the verb begins as a complex like the one in (28), which illustrates a possible internal structure for the double object verb in sentences like *Nixon gave Mailer a book* (17a). The subscripts indicate the theta roles that are to be discharged. *G* is just a case assigner, CAUS is just a theta role assigner, and *V* has both case and theta role assigning properties.

(28) [v [G [CAUS_[: causer]] G] V_[: goal, theme]]

When the VP is first built and the Goal argument is attached as sister to the verb the goal theta role of *V* is assigned (29). I use a strike-out notation to indicate that a theta role has been discharged by a theta-assigner.

(29)



Subsequent to this, when the theme argument is added on the right of the structure, the entire complex head is lowered by copying the higher verbal head, creating a position in which the theme NP can attach to the structure in a maximally right-branching fashion. Since the V-copying operation involves an entire X^0 the CCM is automatically satisfied. At this point, the lower *V* head

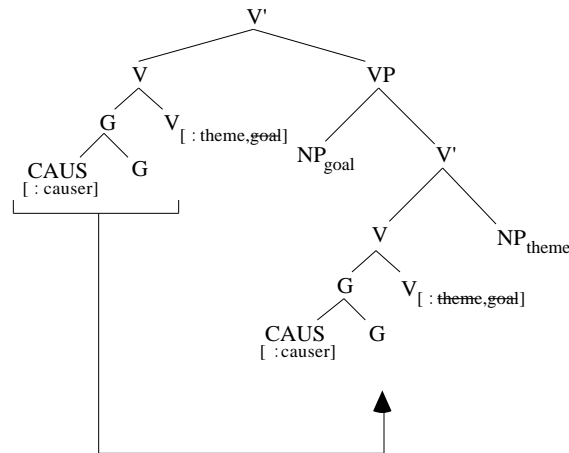
¹⁰⁶ For the current argument to go through, though, the relative embedding of the verbal root and the different argument assigning heads that it combines with does not matter. The only crucial point is Pesetsky's assumption that the causer argument is assigned in the lowest position of all the arguments.

Morphosyntax

assigns its theme role to the theme NP (30a). Case is assigned to the theme argument by the head G that is adjoined to the verbal root.¹⁰⁷

A similar local lowering of the CAUS head out of the complex head creates a position for the subject NP to be lowered from subject position and receive the Causer theta role (30b). Although the CAUS head is deeply embedded inside the complex verbal head, local lowering is possible, because it targets a position which can c-command into the complex verbal head.

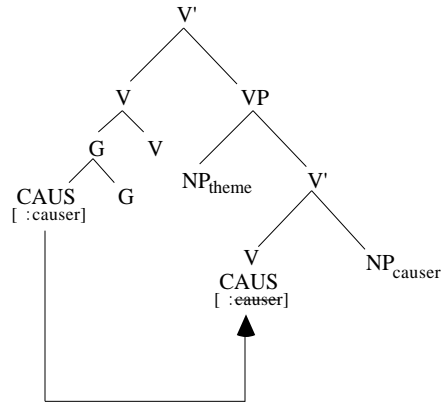
(30) a.



¹⁰⁷ I assume that the entire verbal complex lowers in this case because it is the verb rather than G that has the theme theta role to discharge. Another possibility, which would not affect any of the points of this section, would be to assume that G has both case and theta-role assigning properties, in which case just G+Caus could lower in this example. Note that under this approach the argument structure of V would have to be different in double object and dative constructions, whereas in the approach that I assume in the text the verb assigns the same theta roles in each case, with the one difference between the two constructions being that in the double object construction G is also present, to assign Case to the theme argument.

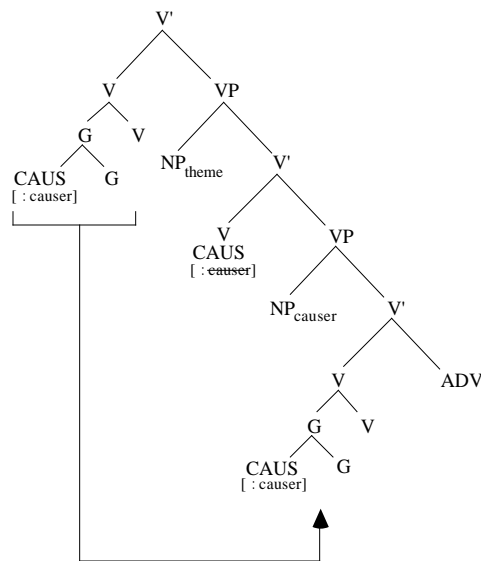
Morphosyntax

b.



With all of the theta roles successfully discharged, adverbial phrases can be added at the right of the structure. The adverbials can be attached as the sister of the verbal root head in a right-branching structure by lowering of the verbal root (31). Since this movement operation involves lowering of an entire X^0 it is able to undergo non-local movement across the intervening CAUS head.

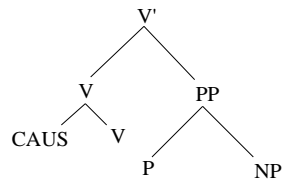
(31)



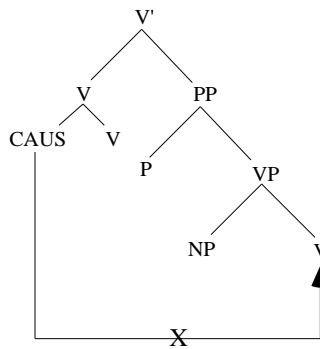
Now consider what happens if we try to derive the unavailable causative reading of the dative construction in the same way as we derived the causative reading for the double object construction. We can ignore for the moment the fact that the theme argument is discharged first in this construction and join the derivation at the point where the Goal PP argument has been attached to the

structure as sister of a the verb (in a lower layer of VP). This structure is shown in (32a). The only way of assigning the Causer role to the subject in this instance is to lower the CAUS head out of the [[CAUS] V] structure. However, this is ruled out by the CCM, as the structure in (32b) shows. The most right-branching attachment of the lowered CAUS head involves creating a PP-splitting structure, but in such a structure neither the source nor the target position of CAUS lowering c-commands the other, because of the intervening P projection.¹⁰⁸

(32) a.



b.



Oehrle's observation can therefore be captured in the left-to-right theory incorporating the CCM in much the same way as it is captured in Pesetsky's system, but without the strict locality requirement of the HMC, and therefore without entailing a ban on the non-local V-lowering that is required to create predication structures for adverbials. I follow Pesetsky in assuming that there is a formal parallel between the T/SM Restriction facts in (21–23) and Oehrle's observation, and therefore leave it as an exercise to the reader to verify that the T/SM Restriction is also captured in the left-to-right theory.

4.3.1.2 Heavy PP Shift voids the restrictions

This section shows an extension of the discussion in the previous section to an interaction between heavy shift and the causer subject restrictions. The empirical coverage of this section goes beyond a translation of Pesetsky's results into my approach, and provides support for the restricted form of PP-splitting that I assume.

Alec Marantz (pc) observes that dative constructions do not always disallow causer subject readings. There is systematic variation in the availability or not of causer-subject readings with dative constructions, depending on the heaviness

¹⁰⁸ Note that we have to exclude the possibility that the V+Caus complex lowers as a unit, in which case non-local movement is predicted to be allowed. One way of excluding this would be to require that V cannot lower and create a new V⁰ position unless it enters into a new syntactic relation in that position, e.g. adverbial predication.

Morphosyntax

of the PP. Although simple PPs do not allow causer-subject readings in datives, heavy PPs do (33–34).

- (33) a. * Hard work got the prize for anyone.
 b. Hard work got the prize for anyone willing to put in the time.
- (34) a. ?? That would give a headache to anybody.
 b. That would give a headache to anyone who listens to it for 5 minutes.

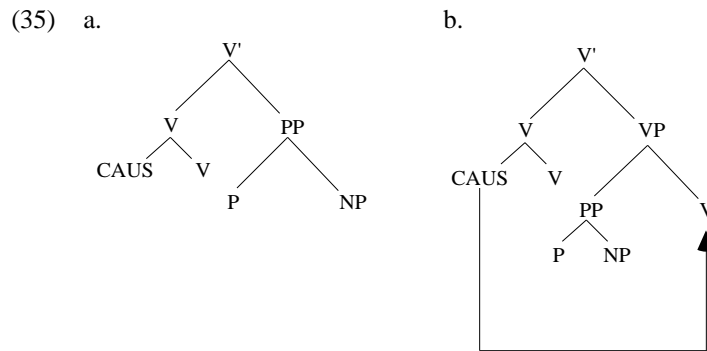
I think that this effect can be accounted for in terms of the left-to-right assembled right-branching structures proposed here.

First, recall that in Pesetsky's account of the impossibility of (33a, 34a) the overt preposition acts as a blocker for local head movement. This blocking effect crucially requires the PP-splitting property of Pesetsky's Cascade structures. If the PP was a constituent, then P^0 would not intervene between the verb and the Caus head. My account is identical to Pesetsky's in this respect.

Where my theory differs from Pesetsky's is that I do not assume a parallel representation in which the PP is a constituent. There is only one derivation for any sentence. Therefore, Heavy PP Shift and PP-splitting are in conflict, because one requires the PP to be a constituent, and the other requires it not to be. As discussed in §2.5.2 I assume that when Heavy PP Shift occurs, PP-splitting does not occur: this is the only way that PPs can move rightwards.

If PP-splitting does not occur, this has the consequence that the preposition no longer intervenes between the Caus head and the V position. In other words, the configuration which made (33a, 34a) bad no longer obtains, and therefore no longer acts as a blocker of V-lowering. This is why Heavy PP Shift in (33b, 34b) escapes Oehrle's restriction.

(35a–b) represent a minimal contrast with (32a–b), and show how the Caus head is able to undergo local lowering in the situation where the PP fails to undergo PP-splitting.



Note: structure shows PP prior to Heavy Shift

This account also predicts that Oehrle's restriction should *not* be voided when the PP undergoes leftward movement, as in (36). My intuition is that this prediction is correct, though this is not such an easy prediction to test, given its interaction with other restrictions on extraction from dative and double object constructions.

- (36) a. * Who did hard work get the prize for?
b. * Who would that give a headache to?

Therefore, since the left-to-right theory predicts PP-splitting effects to be blocked only when rightward movement occurs, the interaction of Oehrle's restriction with Heavy Shift constitutes a new argument for left-to-right assembly of phrase structure.

The next section examines the locality properties of verb fronting constructions in a variety of languages from the perspective of the CCM.

4.3.2 (Non-)locality in Participle Movement

4.3.2.1 Long Head Movement

Head movement processes are commonly assumed to include both operations which move complete syntactic heads and operations which combine stems and affixes to form complex syntactic heads. It is clear that most affixal heads are restricted to combine with a very small class of other heads. However, it is difficult to draw any inferences about restrictions on syntactic head movement processes based on this fact, because affixes are typically constrained by their (morphological) selectional properties in addition to whatever constraints apply to head movement. Therefore, given any impossible combination of affixes it is often hard to tell whether it is ruled out by constraints on head movement or whether it is ruled out by the (categorial) selectional properties of the particular heads involved.

There is one class of affixes which are an exception to this problem. These are the clitic heads found in many languages which are subject to the requirement that they be phonologically attached to some other head, but which are otherwise fairly unselective about what kind of head they attach to. Therefore, by focusing on clitics and their hosts as examples of syntactically complex heads we are able to compare operations which move entire heads and operations which (dis-)assemble complex heads, without the confounding factor of the selectional properties of many affixes. This section compares clitic-driven verb fronting with independently motivated verb fronting in a number of languages.¹⁰⁹

In periphrastic tenses in a number of languages participles may occupy a position higher than finite auxiliary verbs. Examples of this phenomenon have been reported in a number of southern and western Slavic languages, Rumanian, Albanian, Old Romance and Breton (Lema & Rivero 1989, 1991, *avar & Wilder 1994, Břkovic 1995, Embick & Izvorski 1995, Rivero 1996, Borsley et

¹⁰⁹ I use the term 'verb fronting' to refer to the fact that the verb occurs at the front of the clause in these constructions, and not to refer to any process whereby the verb *starts* in a non-initial position and then is moved to the front of the clause.

between the surface and underlying positions of the fronted participle, and these examples are therefore cases of movement across an intervening head.¹¹⁰

It is still possible to dispute the claim that (38–39) involve non-local head movement. Another possible analysis is that this participle fronting is an instance of ‘remnant VP topicalization’ such as is found in a number of Germanic V2 languages (Koster 1987, Webelhuth & den Besten 1987). These constructions have been analyzed as fronting to first position of an entire VP from which all of the arguments have been extracted (40a). The object may also be fronted with the participle (40b).

- (40) a. [_{VP} Gelesen t_i]_j habe ich das Buch_i t_j
 read.prt have.1s I the book
 ‘I have read the book.’
- b. [_{VP} Das Buch gelesen]_j habe ich schon t_j
 the book read.prt have.1s I already
 ‘I have already read the book.’ (German)

This possibility has been considered by a number of authors, and generally rejected. The best argument that I am aware of that participle fronting is not remnant VP topicalization is due to Lema & Rivero (1989), who point out that in Rumanian the three types of auxiliaries that license participle fronting—conditional, future and perfect-as-past—do not license fronting of a VP containing a direct object. On the other hand, auxiliaries like the modal *a putea* (‘can’) license fronting of a VP containing a direct object, but not of a participle alone. (41) shows the impossibility V+O fronting for future tense, and (42) shows that this is possible for modal sentences.¹¹¹

- (41) * [Citi cartea] Maria va.
 read book-the Mary will
 ‘Mary will read the book.’
- (42) [Citi cartea] **nu am** putut
 read book.the not have can
 ‘I have not been able to read the book.’ (Rumanian)

In order for the paradigm in (41–42) to be accounted for under a remnant VP topicalization account it would need to be the case that scrambling of the object

¹¹⁰ See Bošković 1995ab for a dissenting view on the status of such examples as involving long-distance participle fronting. Bošković argues that the Serbo-Croatian examples are derived without long head movement, but since this requires assuming that the participle undergoes excorporation in order to move to initial position in local steps, this analysis is extremely close to a long-distance head movement analysis.

¹¹¹ According to Rivero 1994 sentence (42) is only marginal for some speakers of Rumanian.

out of the VP is *obligatory* in conditional, future and perfect-as-past tenses, but *impossible* in modal sentences.¹¹²

Old Spanish also shows such a split. Future and conditional tenses allow fronting of a participle but not of an entire VP. Modal constructions allow entire VPs to be fronted but not verbs in isolation (Lema & Rivero 1991). Similarly, Embick & Izvorski (1995) report that in both Czech and Bulgarian the tenses that license participle fronting do not license fronting of V+NP sequences (43).¹¹³

- (43) * Psal list **jsem**.
 written letter be.1s
 'I have written the letter.' (Czech)

Therefore, I think that there is good reason to believe that cases of participle fronting like (38–39) do involve non-local head movement.¹¹⁴ Given that these frontings involve non-local head movement, we now need to ask whether the fronted participle is adjoined to another head or whether it heads its own X^0 . Recall that the prediction of the CCM is that long head lowering should be possible only if it moves an entire X^0 , and not if it removes a head from inside a complex X^0 . I aim to show that the prediction of the CCM is correct, and long-distance participle fronting is only possible when the fronted participle is able to be the head of a projection in the fronted position.

Two arguments support this claim. First, long-distance participle fronting for clitic auxiliary support of the kind shown in (38–39) above is only possible in those languages in which long-distance participle fronting is also possible when there is no clitic auxiliary to support. I suggest that this generalization holds because the participle can only head its own projection in the fronted position if it contains features such as focus features which are able to project. I assume that if a language allows a participle to project its focus features then the language allows the participle to do this in all environments, and not just in situations in which the participle precedes a clitic auxiliary. Therefore, the CCM predicts a correlation between the possibility of participle fronting in the absence of clitics and the possibility of non-local participle fronting (§4.3.2.2). On the other hand, when a language does not allow participle fronting

¹¹² As David Pesetsky and Jonathan Bobaljik (pc) point out, such a state of affairs is not inconceivable, given the different aspectual properties of different auxiliaries and the known interaction of object movement processes with aspect. Therefore, a better test of whether these cases involve remnant VP topicalization or not would involve whether postverbal adverbs can precede the auxiliary together with the participle in examples like (41–42).

¹¹³ While the complementarity of V^0 and VP fronting is a useful diagnostic, I do not have an explanation for it at present.

¹¹⁴ Other arguments have been given for distinguishing remnant VP topicalization from participle fronting. One argument is that participial auxiliaries may be fronted, but may not in general undergo VP-fronting. Another is that negation blocks LHM (in some languages, but not others: cf. Rivero 1991), whereas it *never* interferes with the availability of VP-fronting. See Lema & Rivero 1991, Rivero 1994 for details.

independently of clitic support, only the more local kind of participle fronting shown above in (16) is possible.

The second argument is based on the demonstration that long-distance participle fronting in Serbo-Croatian is blocked in precisely the environment where it is prevented from being the head of its own projection in its fronted position (§4.3.2.3).

4.3.2.2 Long-Distance Participle Fronting Crosslinguistically

A number of Slavic languages and non-Slavic languages of the Balkan region (eg. Albanian, Rumanian) contain clitic auxiliary verbs which must appear in second position in the sentence.¹¹⁵ The examples in (44–45) show representative examples from Croatian (Šavar & Wilder 1994). A subject or other initial XP can precede the clitic and act as its host (44a–b). If there is no initial XP the participle can front and act as host for the auxiliary (45). However, the auxiliary cannot stand ‘unsupported’ in sentence initial position (46a), nor can it (in general) be preceded by *both* an XP and a fronted participle (46b–c).

- (44) a. Ivan **ga je** često čitao
Ivan it be.3s often read.ptc
‘Ivan often read it.’
- b. Često **ga je** Ivan čitao
often it be.3s Ivan read.ptc
- (45) Čitao **ga je** Ivan često
read.ptc it be.3s Ivan often
- (46) a. * **Ga je** Ivan često čitao
b. * Ivan često **ga je** čitao
c. * Čitao često **ga je** Ivan (Croatian)

These languages typically also have a series of non-clitic auxiliaries which do not require support and are able to stand alone at the start of the sentence. This is shown for the Croatian emphatic auxiliary *jesam* and the negative auxiliary *nisam* in (47).

- (47) a. **Jesam** čitao knjigu.
pos.be.1s read.ptc book
‘I HAVE read the book.’

¹¹⁵ The appropriate definition for what counts as ‘second position’ for purposes of these clitic constructions is a somewhat elusive matter. The exact determinants of what counts as second position are not important for the current argument, but see below §4.4.1 for a more detailed discussion of second position properties.

Morphosyntax

- b. **Nisam** čitao knjigu.
not.be.1s read.ptc book
'I didn't read the book.'
(Croatian)

The properties in (44–46) are shared across a number of different languages. Variation across languages is found, however, in whether long-distance participle fronting is possible. In compound tenses in which a finite auxiliary combines with two participles, some languages only allow the participle closest to the finite auxiliary to front in order to act as its host—the dialect of Croatian studied by *avar & Wilder is an example of this (cf. Wider & *avar 1994, *avar & Wilder 1994), as shown in (48). When the subject occurs sentence-initially no participle needs to front (48a). In the absence of an overt subject a participle auxiliary may front (48b). (48c) shows that the lower participle cannot be fronted, although main verb participles can front in simpler tenses containing only one participle.

- (48) a. Marija i Ivan **su** bili čitali knjigu *Locality Obeyed*
M. and I. be.3p be.ptc read.ptc book
'Maria and Ivan had read the book.'
- b. Bili **su** čitali knjigu.
be.ptc be.3p read.ptc book
- c. * Čitali **su** bili knjigu. (Croatian)

A number of other languages, meanwhile, do not impose this locality restriction on participle fronting in complex tenses. Examples of such languages are Bulgarian (Embick & Izvorski 1995) and the variety of Serbo-Croatian discussed in Bošković 1995ab (49–50).^{116,117}

- (49) a. Pročela **si** bila knjigata *Locality Violated*
read.fem be.2s been.fem book.def
'You (allegedly) have read the book.'
- b. Bila **si** pročela knjigata. (Bulgarian)
- (50) a. Čekali **ste** bili Marijinu prijateljicu *Locality Violated*
waited are been Marija's friend
'You had been waiting for Marija's friend.'
- b. Bili **ste** čekali Marijinu prijateljicu (*check*) (Serbo-Croatian)

¹¹⁶ Macedonian is an example of a third class of language, in which only the main verb participle can be fronted (see Tomic 1996 for discussion).

¹¹⁷ However, it should be pointed out that there is a constraint on non-local participles: the participle that fronts must be a clausemate of the auxiliary that it supports. Fronting of participles from embedded clauses is impossible, to my knowledge.

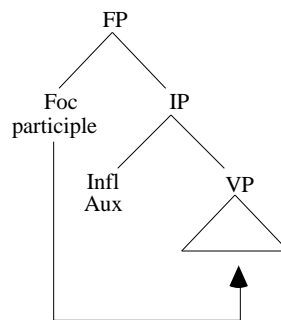
Morphosyntax

The cases that we have discussed so far all involve participle fronting that is *forced* by the requirements of a clitic auxiliary. While participle fronting is sometimes *necessary* in order to support a clitic auxiliary, some languages also allow participle fronting when the finite auxiliary is taken from the non-clitic series and therefore does not need a host. Examples of this ‘optional’ participle fronting from Bulgarian and Serbo-Croatian are shown in (51–52). Examples are taken from Embick & Izvorski 1995.

- (51) a. **Bihte** bili arestuvani ot policiata
 would.2pl been arrested by police.def
 ‘You would have been arrested by the police.’
- b. Arestuvani **bihte** bili ot policiata (Bulgarian)
- (52) a. **Beja^{re}** sreo Petra
 was met Peter
 ‘He had met Peter.’
- b. Sreo **beja^{re}** Petra (Serbo-Croatian)

I suggest that this optional participle fronting places the participle at the front of the clause as the head of its own projection. This XP is a projection of the focus features of the verb, and the focus features subcategorize for an IP sister containing the underlying position of the focused element. The configuration is shown in (53). The most important consequence of this is that since the fronted participle heads its own projection the CCM will allow it to lower across the non-clitic auxiliary to its underlying position inside VP.

(53)



Individual languages may choose to allow or disallow the kind of participle fronting in (53), and they may choose to restrict the construction to broader or narrower classes of compound tenses. Formally, the difference among languages and construction types lies in whether focus features attached to the verb are allowed to project, but I have nothing to say about what determines these choices (but see Rivero 1991 for some interesting suggestions).

I make the following assumptions about participle fronting for clitic support:

Morphosyntax

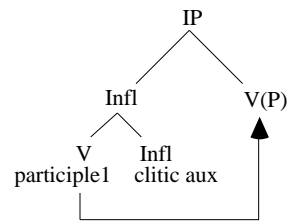
- The underlying position of auxiliaries is Infl.
- Clitic auxiliaries require that they be attached to the right of an element upon which they can be phonologically dependent.

Both of these assumptions are standard, and are presented for the sake of completeness. The next two assumptions are more particular to my analysis, and play an important role in my account of cross-linguistic variation.

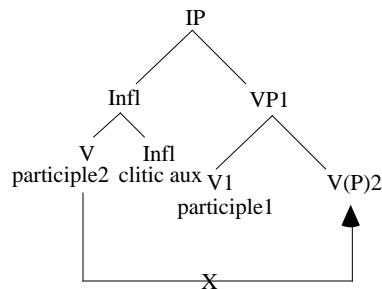
- If the clitic's host occupies an independently motivated position (e.g. Focus⁰), then the clitic adjoins to the host in that position.
- If the host has no reason for being where it is, other than for hosting the clitic, then the host adjoins to the clitic.

Given these assumptions and the CCM, there are predicted to be two ways in which participles can host clitic auxiliaries. In the first of these, the participle is fronted solely in order to host the clitic, and therefore enters the syntactic derivation adjoined to the Infl node that the clitic occupies. The adjoined participle will only be able to lower to its underlying position if that position is the head position closest to Infl, because the target must c-command the source. This movement is illustrated in (54a). If the participle were to move from its adjoined position directly to a lower head position the CCM would be violated (54b). Therefore, when the participle is fronted only to host the clitic auxiliary, there is a strict locality condition on fronting such that the participle that fronts must be the closest participle to the clitic auxiliary.

(54) a.



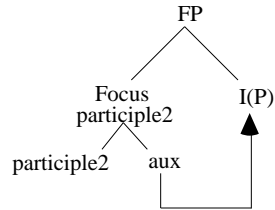
b.



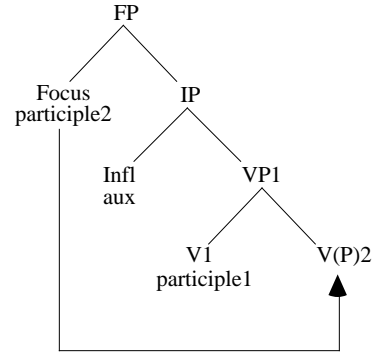
The second possibility is that the participle independently fronts to a position higher than the clitic auxiliary and that the auxiliary adjoins to the participle, rather than vice versa. (55) shows how long-distance participle fronting is possible in this scenario. Decomposition proceeds by first locally lowering the auxiliary from its position adjoined to the participle (55a). Subsequent to this, the participle is able to lower across intervening heads to reach its underlying position (55b), since it is an X^0 that heads its own projection, and therefore c-commands any position inside its sister.

Morphosyntax

(55) a.



b.



In order for this account to be viable, it must be the case that long-distance participle fronting for the hosting of clitic auxiliaries (e.g. 38–39) is only possible in those languages which independently license fronting of a participle to sentence-initial position. This prediction appears to be true. The table in (56) shows the cross-linguistic distribution of long-distance participle fronting for hosting clitics and the distribution of participle fronting with non-clitic auxiliaries.

(56)

	Fronting w/ non-clitic	Non-local fronting	Source
Bulgarian	Yes	Yes	E&I ¹¹⁸ , Rivero 1991
Macedonian	Yes	Yes	E&I, Tomic 1996
Serbo-Croatian	Yes	Yes	Borković 1995
Slovene	Yes	Yes	E&I
Polish	Yes*	Yes	Borsley & Rivero 1994
Rumanian	Yes	Yes	Rivero 1991, 1994
Breton	n/a*	Yes	Borsley et al. 1996
Croatian	No	No	Wilder & *avar 1994
Old Spanish	No*	No	Lema & Rivero 1991
Czech	n/a ¹¹⁹	No	Rivero 1991
19th cent. Eur. Port.	??	No	Lema & Rivero 1989
Albanian	??	??	Demiraj 1986

¹¹⁸ The table is based on a table in Embick & Izvorski 1995 (E&I), but it differs from their table in a number of respects. It omits certain of their categories, but adds some languages. Also, certain of my classifications differ from theirs: these are discussed in the appendix.

¹¹⁹ Czech does not have non-clitic auxiliaries, and therefore the prediction of the CCM is untestable in this language. (*check*)

The table largely supports the prediction of the CCM. Languages that allow long-distance participle fronting for clitic support independently allow participle fronting as a focusing process.¹²⁰ The cells in the table that are marked with an asterisk require some discussion, since they are cases where my classification differs from at least some reports in the literature. I discuss these cases (Polish, Old Spanish, Breton) in more detail in the appendix to this chapter.

Summarizing the cross-linguistic survey, long-distance participle-fronting for clitic support appears to be contingent on the availability of independent long-distance participle-fronting. I suggest that this is because fronting to a position in which the participle heads its own projection is necessary in order to allow unbounded head movement, consistent with the predictions of the CCM. The second argument that local head movement is forced by decomposition processes comes from an interesting twist in the data just discussed.

4.3.2.3 Filled Comp blocks LHM in Serbo-Croatian.

In the previous section I provided a cross-linguistic argument for restricting strict locality effects in head movement to processes which (dis)assemble complex heads. Non-local head movement is only possible when it moves an entire X⁰ element. In this section I present a similar argument about locality in head movement, but this time based on a contrast in locality of head movement internal to Serbo-Croatian.

The argument draws primarily on data presented in Rivero 1993, and is based on the distribution of the particle *-li*, which is used to mark *yes/no* questions in Serbo-Croatian. *Li* has second position clitic properties similar to the clitic auxiliaries that we have been discussing. It can be supported by a fronted verb (58a) or by a sentence initial XP (58b), and it cannot appear sentence initially (59a) or in third position in the sentence (59b).

- (57) Marko studira medicinu
 Marko study.pres.3s medicine
 ‘Marko is studying medicine.’ (Serbo-Croatian)
- (58) a. Studira **li** Marko medicinu?
 ‘Is Marko studying medicine?’
- b. Marija **li ti ga** dade?
 Maria Q 2s.dat 3s.acc gave.3s
 ‘Was it Maria who gave it to you?’
- (59) a. * **Li** studira Marko medicinu?
 b. * Studira Marko **li** medicinu?

¹²⁰ The focus effect of optional participle fronting has been remarked upon by some authors, but I have not verified this for all of the languages discussed, and therefore this interpretive property remains a hypothesis to be confirmed or disproven.

Morphosyntax

As already pointed out above, Serbo-Croatian allows long-distance participle fronting, both to support a clitic auxiliary and when the auxiliary does not require a host (60a–b). (60a) shows an example of a compound tense with a clitic auxiliary and two participles, in which the lower of the two participles has been fronted to host the auxiliary. (60b) is an example fronting across a non-clitic auxiliary from Bošković 1995b.

- (60) a. **Čekali ste** bili Marijinu prijateljicu
 waited are been Marija's friend
 'You had been waiting for Marija's friend.'
- b. Istukao **bejaše** Petra
 beaten was Peter
 'He had beaten Peter.' (Serbo-Croatian)

Given the combined facts that finite verbs can front to support *-li* and that long-distance participle fronting is possible in Serbo-Croatian, it is surprising that *-li* cannot be supported by fronting of participles across auxiliaries (62a). In periphrastic tenses in which there is no initial XP to support *-li* the non-clitic full form of the auxiliary must be fronted to host *-li* (62b).

- (61) **Čitao sam** knjigu
 read have.pres.1s book
 'I have read a book.'
- (62) a. * **Čitao li sam** knjigu.
- b. **Jesam li** čitao knjigu
 have.pres.1s Q read book
 'Have I read a book?' (Serbo-Croatian)

This gap in the paradigm is not so surprising, however, when we consider the account given above of how long head movement is possible. Operations that lower an entire X^0 are possible across a potentially unbounded distance, whereas operations that pull a head from out of a complex X^0 are required to be strictly local, in order to satisfy the CCM.

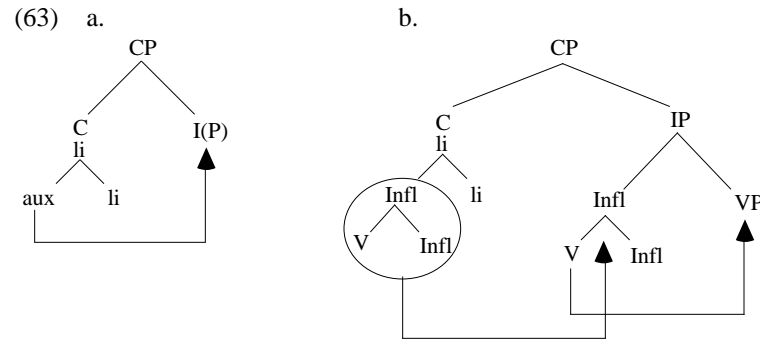
A standard assumption about Serbo-Croatian *-li* is that it heads the C projection, possibly marking it with the feature [+Q]. A consequence of this is that a fronted verb cannot be the head of the C projection that *-li* already heads. Therefore, the verb can only attach to C by adjunction to *-li*, as in (63).¹²¹

If the verb is adjoined to *-li*, then it will only be able to satisfy the CCM if it undergoes strictly local lowering. This presents no problem if either (a) the fronted verb is a finite auxiliary whose underlying position is the closest head position to *-li* (63a), or (b) the fronted verb is a finite verb which can lower to

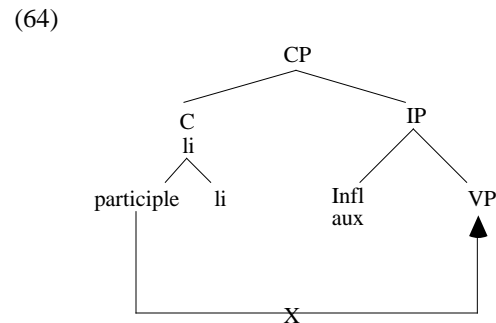
¹²¹ I assume that it is also not an option for the verb to move to a Focus position above CP, for the reason that Focus requires an IP complement. What I do not rule out is that the verb fronts to a Focus position below C (see §A.1 for a reason to assume this). However, such a movement would fail to satisfy the need of *-li* for a host.

Morphosyntax

its underlying position in a series of strictly local steps, peeling off its inflections as it goes (63b).



If, on the other hand, the fronted verb is a participle, then there will be no way for the participle to reach its underlying position from a source position adjoined to *-li*, without violating the CCM, as (64) shows. The position adjoined to *-li* does not c-command into the VP, because of the intervening zero-level projection of *-li*, and the VP does not c-command into the complex C head, because of the intervening IP material. It is for this reason that the presence of *-li* blocks participle fronting.



In sum, Serbo-Croatian licenses participles as the head of CP, which makes long-distance participle fronting possible, as we have seen in (60). But in the one environment in which a fronted verb is forced to adjoin to another head in C^0 , participle fronting is blocked. The blocking effect of *-li* on long head movement in Serbo-Croatian therefore provides support for the claim that locality conditions on head movement are a specific property of operations which decompose complex X^0 s.

However, Bulgarian appears to raise a complication for this account of the Serbo-Croatian facts. Bulgarian *yes/no* questions are also formed with a participle *-li* which has second position clitic properties. Like Serbo-Croatian, Bulgarian allows long-distance participle fronting in declarative sentences (65) and allows *-li* to be supported by a fronted finite verb (66).

Morphosyntax

- (65) Pročela **si** bila knjigata
 read.fem be.2s been.fem book.def
 ‘You (allegedly) have read the book.’
- (66) Vidjaxme **li** knjigata?
 see.aorist.1p Q book.def
 ‘Did we see the book?’ (Bulgarian)

In contrast to Serbo-Croatian, though, Bulgarian *does* allow fronting of a participle across a clitic auxiliary in order to support *-li*, as (67) shows.¹²² How is this possible, given that it is impossible in Serbo-Croatian?

- (67) a. Vidjal **si go**
 seen have.pres.2s him
 ‘You have seen him.’
- b. Vidjal **li si go?**
 ‘Have you seen him?’ (Bulgarian)

At first sight, Bulgarian examples like (67b) appear to pose a serious challenge to my account of the impossibility of fronting participles to host *-li* in Serbo-Croatian. For Serbo-Croatian I argued that the UG ban on anything but strictly local head movement in head (de)composition blocks participles from hosting *-li*. But if Bulgarian demonstrates that the possibility of participles hosting *-li* is a matter of cross-linguistic variation, then I cannot very well appeal to a UG mechanism to rule out Serbo-Croatian examples like (62a).

However, an observation due to Penčev (1993: cited in Dimitrova-Vulchanova & Hellan 1996) provides a simple explanation for Bulgarian examples like (67b). Penčev points out that Bulgarian allows VP-fronting in *yes/no* questions (68), although VP-fronting is not normally possible.

- (68) [Srēna recenzenta si] **li** Ivan.
 met.ptc opponent his Q Ivan
 ‘Was it meeting his opponent that Ivan did?’

The availability of VP-fronting in Bulgarian *yes/no* questions raises the possibility that examples like (67b) do not involve long head movement, but instead involve VP topicalization to a specifier position, and therefore do not undermine my account of locality effects in Serbo-Croatian *yes/no* questions.

There is another reason not to abandon our account of locality in head movement in the light of (67b). Another property of Bulgarian *yes/no* questions shows locality properties related to what we have seen in Serbo-Croatian.

The proclitics *te* (future) and *ne* (neg) may appear sentence initially, but need not do so. In *yes/no* questions containing *te* and an initial XP, *-li* precedes *te* (69a), and a subject may intervene between *-li* and *te* (69b). Examples are taken from Tomic 1996.

¹²² Macedonian raises a similar issue, in that it allows fronted participles to host clitic *-li*. See Tomic 1996 for discussion of the relevant facts.

Morphosyntax

- (69) a. Penka **li** **ʔte ti** dade knigata?
 Penka Q fut 2s.dat give.3s book.def.fem.sg
 ‘Is it Penka who will give you the book?’
- b. Utre **li** Penka **ʔte ti** dade knigata?
 tomorrow Q Penka fut 2s.dat give.3s book.def.fem.sg
 ‘Is it tomorrow that Penka will give you the book?’ (Bulgarian)

I assume that the ordering of *-li* before *te* in (69a–b) reflects the fact that *-li* is a syntactically higher head than *te*.

We have already seen above that verbs may precede *-li*. Participle verbs can also precede *te* (70).¹²³

- (70) Proʔel **ʔte su**°m ja knigata.
 read.ptc fut be.1s it.acc book.def
 ‘I will have read the book (by then).’

However, interesting changes in word order occur in sentences that contain both *-li* and *te*, and not containing a sentence initial XP. In these situations the verb cannot front across *-li* and *te* in order to host *-li*, contrary to what we might expect from the examples in (67–70). In these situations the relative order of *-li* and *te* reverses, and *-li* now appears to the right of the finite verb, which in turn is to the right of *te* or *ne*, as the examples in (71–72) illustrate. The examples in (71) are from Rivero 1993; (72) is from Dimitrova-Vulchanova & Hellan 1996.

- (71) a. **ʔte** go viʔdaʔ
 will him see.pres.2p
 ‘You(pl) will see him.’
- b. **ʔte** go viʔdaʔ **li**?
 ‘Will you see him?’ (Rivero 1993)
- (72) a. Toj **ne** beʔe **li** rabotil tam?
 he neg had Q worked there
 ‘Hadn’t he worked there?’
- b. **Ne** raboti **li** toj?
 heg works Q he
 ‘Doesn’t he work?’ (Dimitrova-Vulchanova & Hellan 1996)

¹²³ However, non-participial forms of the verb may not front across *te* in Bulgarian, as Dimitrova-Vulchanova & Hellan (1996) observe.

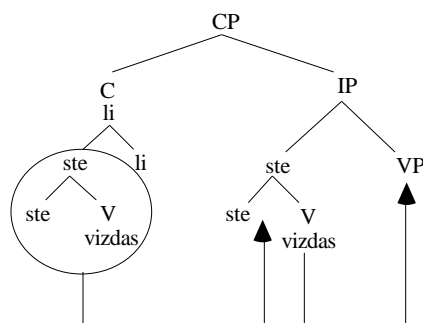
(i) * Proceta **ʔte** ja knigata.
 read fut.3s it.acc book.def
 ‘I will read the book.’

Morphosyntax

The examples in (69–72) show that the verb may precede either *li* or **te* when either appears alone, but not when they both appear in the same clause. If we assume that the verb adjoins to *-li* in Bulgarian in the same way as verbs in Serbo-Croatian, then these facts may be explained in just the same way as the Serbo-Croatian facts, as a violation of the locality requirements imposed by the CCM. If a verb were to precede *te* or *ne* in a yes/no question, non-local lowering would be required for the verb to reach its underlying position. But since the surface position of the verb would be adjoined to *-li*, this is impossible.

Regarding how the orderings in (71–72) are derived, I follow Tomic 1996 and Rudin 1996 in assuming that they involve a situation in which a complex head containing *te* or *ne*, any pronominal clitics and the verb precedes *-li*.¹²⁴ In other words, Bulgarian allows the verb to piggy back on *te* in order to move to a position adjoined to *-li*. Since this complex head can be decomposed by means of exclusively local steps of head movement, the CCM is satisfied. A sample derivation for (71b) is shown in (73) (the pronominal clitic *go* has been omitted from the diagram for the sake of exposition).

(73)



This analysis of locality of head movement in Bulgarian *yes/no* questions is currently rather tentative. In particular, more justification needs to be given for the claim that the orderings in (71–72) are derived by fronting of a larger-than-normal head, rather than by lowering of *-li*, as some authors have argued (e.g. Rivero 1993, Dimitrova-Vulchanova & Hellan 1996). For example, we should expect that the complex head that I claim exists in (71–72) should not be interruptible by XPs.¹²⁵

Summary. Summarizing this section, I have examined the locality restrictions on head movement that hold independent of any category selection restrictions of

¹²⁴ See also Izvorski, King & Rudin 1996 for further arguments against lowering of *-li* in Bulgarian.

¹²⁵ Although I have treated *te* and *ne* here as if their behavior is entirely parallel, this is not strictly true. Verbs can front across *te* when it is not in a *yes/no* question with *-li*, whereas *ne* blocks verb raising in all environments, whether or not *-li* is also present. I leave this difference unexplained for the time being, but see Izvorski et al. 1996 for discussion of differences between *te* and *ne*.

particular affixes, focusing on the various kinds of movement which satisfy the requirements of clitic auxiliaries and clitic question markers for a host. I argued that there is a locality contrast between head movement operations which (de)compose complex heads (strictly local) and operations which move entire heads (longer movement possible). This generalization is predicted by the c-command condition on movement (CCM) introduced in §4.2.

4.3.3 Local Head Movement

Having shown that there are good reasons to replace the Head Movement Constraint with the more liberal CCM, it is important to show that the kinds of restriction which originally motivated the HMC are still captured in this theory. Although I obviously cannot survey the full range of effects which have been thought to be handled by the HMC, I can show how a couple of classic generalizations are handled by the CCM. These are restrictions on incorporation (Baker 1985, 1988) and restrictions on second position phenomena in Germanic (Travis 1984).

Baker (1985, 1988) argues that incorporation phenomena involve strictly local head movement. This means that in a configuration like (74) the head Y should be able to incorporate into head X, but head Z should not be able to incorporate into X skipping Y, because Y is a closer head to X. This is attributed to the HMC (which for Baker is a corollary of the empty category principle of Chomsky 1981, 1986).

(74) X [_{YP} ... Y ... [_{ZP} ... Z ...

As an example of what this generalization predicts, consider the examples from Niuean in (75–76) (Seiter 1980, Baker 1988). Direct objects may undergo Noun Incorporation (75b), but the object of a preposition cannot be incorporated into a verb (76b).

- (75) a. Volu nakai he tau faʔnau e fua niu?
grate Q erg.pl.children abs.fruit coconut
'Are the children grating the fruit of the coconut?'
- b. Volu niu nakai e tau faʔnau?
grate coconut Q abs.pl.children
'Are the children grating coconut?'
- (76) a. Fano a ia ke he tapu he aho tapu.
go abs.he to church on day Sunday
'He goes to church on Sundays.'
- b. * Fano tapu a ia (ke he) he aho tapu.
go church abs.he to on day Sunday
'He church-goes (to) on Sundays.'
- (Niuean)

The restriction shown in (75–76) follows equally well if head movement is restricted by the CCM rather than the HMC. This is because the separation of

Morphosyntax

Another motivation for the HMC, pointed out by Travis, was the need to ensure that in Germanic verb second constructions the appropriate verb appears in second position. For example, consider the German examples with compound tenses in (78). The direct object has been topicalized to first position and a verb must appear in second position. Under standard assumptions about verb second constructions these two positions are the specifier and head of the CP projection respectively. The relevant generalization is that only the (highest) finite verb may appear in second position (79a). If one of the non-finite verbs fronts, the sentence is ungrammatical (79b–c). This is not because the finite verb cannot appear sentence finally, as the embedded clause in (79d) shows.

- (79) a. Die Bücher will er gelesen haben.
the books want.3s he read.ptc have.inf
'He wants to have read the books.'
- b. * Die Bücher haben er gelesen will.
c. * Die Bücher gelesen er haben will.
- d. ... daß er die Bücher gelesen haben will. (German)

I assume that the locality effect in German (and similar effects in English, involving auxiliary fronting in questions) is due to the fact that fronted verbs in German do not head their own projection, but instead are adjoined to C^0 . Non-local verb fronting is contingent on the verb being able to head its own sentence-initial projection, as we have seen in detail above. For the languages that allow this I suggested that this is because the language allows the focus features of a focused verb to project. This appears not to be an option in German, and therefore a fronted verb must be adjoined some other projecting head, e.g. C^0 . Since second position verbs are adjoined to C^0 they will only satisfy the CCM if they undergo local lowering.

Conclusion. In this section I have argued that head movement is not required to be strictly local in the sense of the Head Movement Constraint (Travis 1984, Baker 1985, 1988). Instead, I claimed that strict locality is a property of head movement operations which (dis)assemble complex heads, and that movement of entire heads is not required to be local. This contrast in locality conditions follows from the requirement that positions related by movement stand in a c-command relation to one another.

This analysis provides a common account of some syntactic effects which at first sight seem to bear no relation to one another. For example, this section gives the same characterization to (i) the possibility of fronting finite verbs but not participles in Serbo-Croatian *yes/no* questions, and (ii) the possibility of causer-subject readings of dative constructions only when Heavy PP Shift occurs.

I should point out that although I have been showing in this section how local and non-local head movement is instantiated in the Merge Right system, it would be relatively straightforward to translate the main generalizations of these sections into more standard approaches, in which movement proceeds in a leftward and upward direction, and head movement assembles rather than

disassembles complex heads. In an upward movement approach any theory which allowed only strictly local movement for adjunction but non-local movement for substitution operations would capture most of the contrasts discussed here. [The interaction of Heavy PP Shift with restrictions on causer-subjects in §4.3.1 is one exception to this.]

This account of locality in movement and incorporation is interesting in its own right, but it also serves as a preliminary to the discussion of effects of ordering in §4.4, which shows how the left-to-right theory provides a uniform syntactic account of head movement and clitic placement.

4.4 Effects of Ordering

This section documents three arguments for preferring an approach to morphosyntax which builds syntactic structures from left-to-right and (consequently) decomposes complex heads rather than assembling them, as is more standardly assumed. It builds on the account of head movement developed in sections §4.2–§4.3.

In §4.4.1 I discuss issues in the placement of second position clitics, primarily in Serbo-Croatian, which have been argued to require the morphological component of the grammar to be transformational in nature, involving transformations distinct from syntactic head movement. I argue that once the syntax is viewed as pulling apart rather than putting together complex heads, the transformational operations required to handle the clitic placement facts fall under the range of possibilities allowed by the CCM, and therefore reduce the need for a separate transformational morphology.¹²⁶

In §4.4.2–3 I discuss two cases of apparent linear asymmetries in the typology of head movement, both of which receive a natural explanation in the current approach, which allows only rightward movement operations. §4.4.2 looks at the range of cross-linguistic variation in syntax–phonology mismatches in clitic placement, and shows that a paradigmatic gap in the paradigm of these mismatches independently noticed by Hale & Selkirk (1987) and Marantz (1988) is precisely what cannot be generated by rightward movement. §4.4.3 then looks at the adjacency condition on the lowering of heads below their underlying position noted by Bobaljik (1994, 1995) and Lasnik (1995), and shows that the adjacency requirement is an automatic consequence of the Merge Right approach to structure building.

4.4.1 Serbo-Croatian Clitics

4.4.1.1 Interfacing Phonological and Syntactic Structures

This section discusses some properties of the placement of second position clitics which have often been taken to show that the position of syntactic heads is determined at least in part by non-syntactic head movement in the

¹²⁶ Note that I am not trying to question the existence of autonomous morphological constraints (e.g. in the form of templatic morphology), only the existence of a morphological component with *transformational* operations different from the syntax.

morphophonological component of the grammar. The data in this section is drawn primarily from Serbo-Croatian, but the analysis presented should extend fairly straightforwardly to the examples from other languages discussed in the next section, since they are less complex than the Serbo-Croatian facts.¹²⁷

We first review the basic facts of clitic ordering and placement, and the argument that the characterization of clitic placement cannot be reduced to syntactic mechanisms alone. Then I show how clitic-related movement falls under the range of options allowed by the CCM.

Serbo-Croatian clitics appear together as a cluster in the ‘second position’ of the sentence. The clitic elements include pronominals, some auxiliary verbs and the interrogative particle *-li*. The relative ordering of the clitics is generally described in terms of a template like (80).¹²⁸ Our main interest here will not be in the relative ordering of the clitics in the clitic cluster, but in the characterization of ‘second position’, where the cluster appears in the sentence.

(80) *li*-Aux-Dat-Acc/Gen-*se*(refl)-*je*(aux.3sg)

It is a (fairly) reliable generalization that the clitic cluster appears in second position. The range of elements that can act as the clitics’ host in first position, however, is somewhat heterogeneous, and can include arguments, adverbials, adjectives, *wh*-phrases, and fronted participles, as the examples in (81) show.

- (81) a. Marija **mi- ga -je** dala.
 Maria me.dat-it.acc-is given.fem.sg
 ‘It was Maria that gave it to me.’
- b. Dala **mi- ga- je** Marija.
 given.fem.sg me.dat-it.acc-is Maria
 ‘Maria has given it to me.’
- c. Juj[°]e(?) **mi- ga- je** dala.
 yesterday me.dat-it.acc-is given.fem.sg
 ‘It was yesterday that she gave it to me.’
- d. Lepa **si mi**
 beautiful.fem.sg are-me.dat
 ‘You look very beautiful to me.’ (Tomić 1996)
- e. Z^što **li-mu-ga-je** poklonila?
 why Q-him-it-Aux presented
 ‘Why did she present it to him?’ (Schütze 1994)

¹²⁷ This does not necessarily mean that the phenomena are less complex, only that in the case of Serbo-Croatian we are fortunate that the facts of clitic placement have been studied in considerable detail.

¹²⁸ The template in (80) is slightly simplified. See *avar & Wilder 1994 for more comprehensive templates for clitic clusters in Croatian.

Morphosyntax

If this were all there is to say about second position clitic placement in Serbo-Croatian, it would be rather easy to formulate a syntactic clitic placement rule. An appealing first hypothesis, which a number of people have pursued, is to assume that the clitic cluster occupies the head of a CP projection, giving a superficial similarity to accounts of verb second in Germanic. The first position material either moves to the specifier of CP if it is an XP or left adjoins to the head of CP if it is a head. Provided that an additional constraint is added to block situations in which the clitic cluster is simultaneously preceded by elements in both the specifier *and* the head of CP, it can be ensured that the cluster appears in second rather than third position.

There is, however, a good deal more to be said about the definition of second position for Serbo-Croatian clitics. What has attracted most interest to Serbo-Croatian clitics is the fact that the clitic cluster can often appear internal to another phrase, such that the material in first position appears not to constitute a well-defined syntactic constituent.¹²⁹ (82a–b) show that a clitic cluster that can follow the initial NP can also occur immediately following the determiner. (93a–b) shows that a clitic auxiliary can either follow an entire PP or it can intervene between a prenominal adjective and the head noun. The preposition–adjective sequence that assumes the role of ‘first position’ in (93b) is not normally assumed to be a syntactic constituent.

- (82) a. Taj pjesnik **mu-je** dao autogram
this poet him-is give.ptc autograph
‘This poet has given him an autograph.’
- b. Taj **mu-je** pjesnik dao autogram.
(Croatian: Wilder & *avar 1994)
- (83) a. U zelonoj kući **je** stanovao.
in green house is stayed
‘He stayed in the green house.’
- b. U zelenoj **je** kući stanovao. (Croatian: *avar & Wilder 1994)

Facts like these have led a number of people to believe that second position must be defined in non-syntactic terms, and should be defined in terms that refer to phonological units. The most common assumption is that the Serbo-Croatian clitic cluster must immediately follow the first Prosodic Word (PrWd) of the sentence. Nevertheless, as a number of different authors have recently observed, the facts in (82–83) do not necessarily rule out a syntactic characterization of second position, given that Serbo-Croatian independently allows splitting of XPs by fronting subparts of those NPs (*avar & Wilder 1994, Progovac 1993, Schütze 1994). Examples of such ‘left-branch extractions’, which are not possible in English, are shown in (84–85).

¹²⁹ Note that the remarks about clitic placement that follow are not true of all dialects of Serbo-Croatian. There is a good deal of disagreement between individuals and dialects. However, I think that the generalizations that I present here are not artifacts of combining judgements from different speakers/dialects.

Morphosyntax

- (84) a. Ivan **je** kupio zeleno auto.
 Ivan be.3s buy.ptc green car
 'Ivan bought a green car.'
- b. Zeleno **je** Ivan kupio auto.
- (85) a. Ivan **je** razbio tatino auto.
 Ivan be.3s ruin.ptc fathers car
 'Ivan has ruined his father's car.'
- b. Tatino **je** Ivan razbio auto. (Croatian: *avar & Wilder 1994a)

*avar & Wilder (1994) and Progovac (1993) argue that it is possible to give a syntactic characterization of second position clitic placement by assuming that the XP-splitting cases like (82–83) are derived by placing the clitic cluster in C⁰ and then fronting just a subpart of the XP to SpecCP, in the manner of the left-branch extractions in (84b) and (85b). This is a promising approach to rescuing a syntactic characterization of Serbo-Croatian clitic placement, but it predicts an *exact* correlation between the strings that can undergo left-branch extraction and the strings that can count as 'first position' for clitic placement, a prediction which does not hold up.

Percus (1993) and Schütze (1994) show the following non-correlations between clitic placement and left-branch extractions. In PPs which contain multiple prenominal adjectives the clitic cluster is preferentially placed following the *first* adjective—which completes the first prosodic word—and cannot be placed after subsequent adjectives (86).

- (86) a. U velikoj **je** sobi klavir.
 in big aux room piano
 'In the big room is the piano.'
- b. U ovoj **je** velikoj sobi klavir.
 in this aux big room piano
 'The piano is in this big room.'
- c. ?* U ovoj velikoj **je** sobi klavir.

Left-branch extractions from similar NPs, on the other hand, show exactly the opposite possibilities. The preposition and *all* of the adjectives can be extracted, stranding the head noun, but the preposition and just the first adjective cannot be extracted, stranding the noun and the remainder of the adjectives (87).

- (87) a. U veliku Jovan ulazi sobu.
 in big Jovan enters room
 'Jovan enters the big room.'
- b. * U ovu Jovan ulazi veliku sobu.

Morphosyntax

c. ?? U ovu veliku Jovan ulazi sobu. (Schütze 1994: 400–401)

I leave aside for the moment the question of what characterizes the set of possible left-branch extractions in Serbo-Croatian (but see *avar & Wilder 1994, Schütze 1994 for discussion of some possibilities), but I think that the paradigm in (86–87) seriously undermines any attempt to reduce clitic placement to the syntax of left-branch extraction. I therefore assume that clitic positioning in these cases is determined by phonological units such as Prosodic Word.¹³⁰ I should stress, though, that I do not mean to imply by this that clitic cluster placement can be fully explained in terms of phonological units. A generalization like ‘position the clitic cluster immediately following the first prosodic word’ is useful in accounting for certain syntactically recalcitrant examples like (86), but it both overgenerates and undergenerates in a number of ways, as a number of authors have pointed out (Halpern 1992, Progovac 1992, Schütze 1994, *avar & Wilder 1994). I return to such cases below.

Given that the position of the clitic cluster needs to refer to positions which are well-defined phonologically but apparently not well-defined syntactically, the problem arises of how to interface the syntactic and phonological representations of Serbo-Croatian sentences in order to guarantee correct word order.

In what follows I compare some different approaches to the phonology–syntax interface based on their characterization of Serbo-Croatian clitic placement, and argue that the derivations of the left-to-right theory fare best.

The approach adopted by Halpern 1992, Percus 1993 and Schütze 1994 is to assume that the syntax feeds the morphological/phonological component of the grammar with representations in which the clitic cluster may or may not already be appropriately positioned prosodically. If the syntactic positioning satisfies prosodic constraints, then nothing further need be done by the morphological component. If, on the other hand, the syntax supplies a representation in which the clitic cluster is inappropriately positioned—for example, sentence initially—then the morphological transformation *Prosodic Inversion* (PI) may be invoked to reorder the clitic cluster and the following PrWd, creating a morphologically well-formed output.

- (88) *Prosodic Inversion* (Halpern 1992, p. 81)
Prosodic adjunction of clitics: for a directional clitic X which must attach to a [PrWd] to its left (respectively right):
- a. if there is a σ , Y, comprised of material which is syntactically immediately to the left (right) of X, then adjoin X to the right (left) of Y.
 - b. else attach X to the right (left) edge of the σ composed of syntactic material immediately to its right (left).

As an illustration of how this would work, consider an example in which the clitic cluster separates a prenominal modifier from the head noun in the sequence P–Adj–Clitic–N, as in (83b, 86a) above. The output of the syntax is a

¹³⁰ Of course, this in turn begs the question of how the size of Prosodic Words is determined, a question which I have nothing interesting to say about here.

Morphosyntax

wiped out and replaced with phonological structure—this means that phonological structures retain much of the structure of syntactic structures—and second, that restructuring operations (prosodic mapping/rebracketing) and reordering operations (prosodic inversion/directional affixation) may be interleaved.

What both the Prosodic Inversion and the Rebracketing analyses of clitic placement have in common is the assumption that there are structure mismatches between syntax and morphology/phonology, which must be repaired, either by morphophonological transformations such as PI or by rebracketing. These transformations are different from syntactic transformations, and therefore there are movement transformations in both morphology and syntax.

Whereas the majority of theories of head movement since Travis 1984 and Baker 1985, 1988 have assumed that incorporation/affixation and head movement form a natural class to the exclusion of clitic placement operations, which are handled by a separate component of the grammar, Marantz assumes a slightly different typology. He assumes that clitic placement operations and incorporation/affixation form a natural class, and that these are distinct from head movement. What I am attempting to do in this chapter is avoid both of these distinctions.

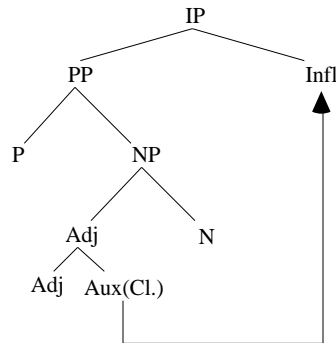
The need for distinct syntactic and morphological transformations does not arise if we adopt the model forced by Merge Right. As we have seen in the previous section, in the Merge Right approach the morphological component of the grammar feeds the syntax rather than vice versa. With respect to clitic placement, this means that the derivation begins with a morphologically well-formed input, and must proceed from there to a well-formed underlying syntactic structure. This means that the syntactic operations that move clitics are those which move the clitics towards their underlying positions, rather than operations that move them to their surface positions, as is more standardly assumed.

An interesting consequence of this reversal of the derivation of clitic placement is that the positions targeted by clitic movement become a syntactically well-defined set of positions, which falls together with the class of possible targets of head movement discussed in §4.2–4.3.

As an illustration, consider the example in (91), repeated from (86a) above. This is the kind of example which provided the strongest evidence against exclusively syntactic placement of clitics. We join the derivation at the stage at which the sentence initial PP containing the clitic has been constructed. At this point, the clitic can move rightwards out of the PP, creating a new Infl position in which it heads its own projection. In this position it c-commands the position that it moved from, and therefore satisfies the CCM. Other cases of clitic ‘un-placement’ could be handled in just the same manner.

- (91) a. U velikoj **je** sobi klavir.
in big aux room piano
'In the big room is the piano.'
- b.

Morphosyntax



The most important feature of (91b) for my purposes is the fact that the process which maps the clitic's surface position into its syntactically well-defined position is formally identical to the head movement operations discussed in §4.3. If this claim is correct, then we arrive at a more inclusive notion of head movement than is usually considered possible, and an important part of the motivation for a transformational morphology is removed. In §4.4.2 I show how other cases of phonology-syntax mismatches in clitic placement can be handled in a similar manner by syntactic head movement.

An obvious question which arises at this point is: couldn't this all be captured equally well in a traditional bottom-up derivational theory, enriched with the CCM or something like it? In such an account, clitics could move anywhere inside a sentence initial XP, provided that they satisfy the CCM. One difference between such an account and the analysis that I have given in terms of left-to-right derivations is that there would be a need to add a phonological filter on specifically this kind of head movement, something which does not need to be added in the left-right story, because the phonological filter is satisfied in the input to syntax.

In the next subsection I provide a stronger reason for the left-right approach to clitic (un)placement, based on the demonstration that the surface structure configurations in which clitics appear must be visible to the rest of the syntactic derivation, which they are not if clitic placement is the very last part of overt syntax.

4.4.1.2 Determinants of Clitic Placement

The preceding discussion shows that in a left-right approach to structure building, the morphophonological and syntactic representations of clitic placement can be interfaced derivationally using exactly the same transformational operations that are responsible for syntactic head movement. This is, however, just one part of the problem of clitic placement in Serbo-Croatian. I have shown how a phonologically sanctioned clitic placement may be turned into a syntactically sanctioned structure, but I have not said anything about what determines where clitics appear. Therefore, this section gives a sketch of what I see as the central problem raised by Serbo-Croatian clitics, and what the left-to-right approach to structure building has to offer on the matter.

Morphosyntax

The main difficulty in characterizing clitic placement in Serbo-Croatian involves ensuring the appropriate contribution of *both* syntactic and phonological factors. As has often been pointed out in the literature, the term ‘second position clitics’ grossly underdetermines the range of positions where clitics may (and may not) appear in Serbo-Croatian.

Previous literature has amply demonstrated that the range of possible clitic positions cannot be described by either a single syntactic generalization or a single phonological generalization. We have already seen that the examples which most strongly motivate a phonological account of clitic placement involve positions which defy syntactic analysis, but are easily described in terms of placing the clitics after the first PrWd (pending, of course, a full account of what constitutes a PrWd). The relevant examples are repeated in (92).

- (92) a. U velikoj **je** sobi klavir.
 in big aux room piano
 ‘In the big room is the piano.’
- b. U ovoj **je** velikoj sobi klavir.
 in this aux big room piano
 ‘The piano is in this big room.’
- c. ?* U ovoj velikoj **je** sobi klavir.

Meanwhile, there are well-documented cases of clitic placements which conform to a fairly straightforward syntactic generalization, e.g., ‘place the clitic as head of CP’, but which wildly violate the prosodic requirement shown in (92). (93a) shows a case in which the clitic cluster appears in third (or more) position, as a result of topicalizations, but it is still possibly in C⁰. In (93b) the clitic cluster is preceded by just one phrase, but that phrase is internally complex and presumably contains more than one prosodic words.

- (93) a. Ove godine taj pesnik napisao **mi je** knjigu.
 this year that poet write.ptc me AUX book
 ‘That poet wrote me a book this year.’ (Schütze 1994: 377)
- b. Roditelji uspešnih studenata **su se** razšli.
 parents successful.gen students.gen AUX REFL dispersed
 ‘The parents of the successful students dispersed.’ (Progovac 1993)

Therefore, while there is some truth to both the phonological account of clitic placement and the syntactic account, but neither is adequate on its own. Far better empirical coverage may be attained by assuming that any given clitic placement must satisfy *either* the phonological *or* the syntactic requirement. Halpern (1992), Percus (1993) and Schütze (1994) all adopt essentially this characterization of the conditions on clitic placement, with the additional assumption that this disjunctive requirement is ordered by the feeding relation between syntax and morphology/phonology. The output of the syntax is first submitted to the syntactic condition on clitic placement. If this results in successful clitic placement, then nothing further happens. Only if the syntactic

condition is not satisfied does Prosodic Inversion apply after prosodic mapping has occurred.

Accounts in which syntactic and phonological requirements are disjunctively ordered achieve broad empirical coverage, but they make one clear prediction which appears to be incorrect. The prediction is that if the syntax does not successfully place the clitic cluster and therefore the task is left to the phonology, then *only* phonological properties should be relevant to the determination of clitic placement. There should be no residue of syntactic constraints applying to the phonological clitic placements.¹³¹

However, there do appear to be a number of examples which satisfy the disjunctive ordering condition, but which are nonetheless impossible. These are the examples that Halpern dubs *fortresses*. Some examples are given in (94–97).

(94a) shows that a prenominal possessor may be separated from a head noun by the clitic cluster when the possessor precedes the head noun. When the possessor follows the head noun, though, the clitic cluster cannot intervene (94b).¹³²

(94) a. Anina **mi ga je** sestra poklonila.
 Ana.gen me it AUX sister given
 ‘Ana’s sister has given it to me.’ (Progovac 1993: 7)

b.%* Prijatelji **su** moje sestre upravo stigli.
 friends have my.gen sister.gen just arrived
 ‘My sister’s friends have just arrived.’ (Halpern 1992: 94)

(95) shows further examples in which a head noun may not be separated from following NP-internal material by the clitic cluster.¹³³ (96) shows that the same fact holds for nouns modified by relative clauses.

(95) a.%* Studenti **su** iz Beograda upravo stigli.
 students AUX from Beograd just arrived
 ‘Students from Beograd have just arrived.’ (Halpern 1992)

b. * Roditelji **su se** uspešnih studenata razšli.
 parents AUX REFL successful.gen students.gen dispersed
 ‘The parents of the successful students dispersed.’ (Progovac 1993)

¹³¹ This does not apply to theories which assume that syntactic structure is not wiped out at the input to phonological structure or for which syntactic and phonological representations are assumed to exist in parallel, e.g. Marantz 1988a, 1989, Zec & Inkelas 1990.

¹³² See Schütze 1994 for a more detailed description of the acceptability status of these violations of fortresses.

¹³³ I assume for purposes of discussion here that coordinate structures like (95c) consist of a ternary branching NP node which dominates NP CONJ NP. In order for this description to apply to (95d) we may assume that names are multiply headed compounds.

Morphosyntax

c.%* Sestra **je** i njen muž doći u utorak.
 sister will and her husband come in Tuesday
 ‘My sister and her husband will come on Tuesday.’

d.%* Lav **je** Tolstoj veliki ruski pisac.
 Leo AUX Tolstoy great Russian writer
 ‘Leo Tolstoy is a great Russian writer.’ (Halpern 1992)

(96) a. Djevojka, koju Ivan voli, **je** fina.
 girl who.acc Ivan like.3s be.3s pretty
 ‘The girl that Ivan loves is pretty.’

b. * Djevojka koju **je** Ivan voli fina.

c. * Djevojka koju **je** Ivan voli fina. (*avar & Wilder 1994: 35–36)

Finally (97) shows that a preposition cannot be separated from its complement by a clitic cluster. Sentences (A) and (B) represent a mini-dialogue which creates a felicitous context for the use of a stressed preposition in the answer. On the assumption that a stressed preposition should count as a PrWd, the preposition ought to be a possible clitic host, and yet it cannot host the clitic.

(97) A: Jesi li bio U kući?
 be.2s Q be.ptc IN house
 ‘Were you IN the house?’

B: Ne, PRED kućom sam bio. / * PRED sam kućom bio.
 no, IN-FRONT house be.1s be.ptc
 ‘No, I was IN FRONT of the house.’ (*avar & Wilder 1994: 36)

All of the starred examples in (94–97) involve splitting of an XP by the clitic cluster. I therefore assume that they do not satisfy the garden-variety syntactic condition on clitic placement (i.e. move to C⁰).¹³⁴ However, the

¹³⁴ A comment is in order on the relevance of left-branch extractions to the examples in (94–97). Progovac (1993) and *avar & Wilder (1994) argue that examples like these are impossible because the pre-clitic material is not independently extractable. These authors take such a correlation to be an argument in favor of a pure syntax account of clitic placement. [Schütze 1994 takes issue with this, based on the fact that the left-branch extractions are quite impossible, whereas the clitic placements are only marginal for some speakers.]

However, even if these clitic placements do correlate with impossible left-branch extractions, this is insufficient to explain why the clitic placements are bad. We have seen above that there are examples of clitic placements which are possible which *cannot* be derived by left-branch extraction, and which have a phonological characterization. Such examples show that even if we assume that *some* clitic placements are derived by means of syntactic left-branch extraction, we must also allow for satisfaction of a phonological requirement to legitimize a clitic placement.

examples all (appear to) satisfy the phonological condition on clitic placement. So they ought to be well-formed according to theories which assume disjunctive ordering of syntactic and phonological constraints. [And according to my theory as well, given what I have said so far.]

Interestingly, these examples have a systematic syntactic characterization: the clitic cluster cannot separate the head of an XP from material inside XP that follows the head. This suggests that syntactic and phonological conditions on clitic placement should be *conjunctively* applied, at least in part. Partially conjunctive application of syntactic and phonological requirements is not possible in a theory in which the syntactic and phonological components of grammar manipulate different representations but are derivationally related. Partially conjunctive constraints are easily applied, however, in an approach in which phonological and syntactic representations are distinct but copresent, as proposed by Zec & Inkelas (1990). Therefore, fortresses might be taken as an argument for parallel representations over derivations. However, I do not think this conclusion is necessary.

In the approach being developed here it is possible to apply the partially conjunctive syntactic and phonological constraints that seem to be required in order to account for Serbo-Croatian clitic placement, but without dropping the claim that phonological and syntactic structures are derivationally related. This is because of two properties of left-right derivations. First, although I assume that phonological representations feed syntactic representations derivationally, this does not mean that a *complete* phonological representation for a sentence is first built and then mapped into a complete syntactic representation for the sentence. Rather, I assume that as (phonological) words are input to a derivation they are immediately mapped into their syntactic representation and that both the phonological and syntactic representations are added to any already built structure to their left. Second, as shown in §4.4.1.1 I assume that phonologically defined and syntactically defined structures are not independent. Phonologically placed clitics occupy positions in syntactic structures, and therefore they have the potential to interfere with syntactic relations.¹³⁵

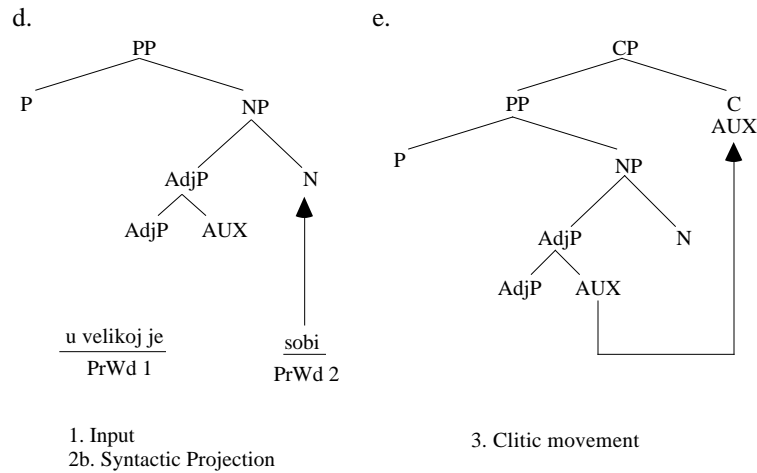
In what follows I illustrate the derivation of some key examples of possible and impossible clitic placements in Serbo-Croatian in left-right derivations. I assume the following conditions on clitic placement in Serbo-Croatian.

- (98) a. Prosodic Words (PrWd) are input to the derivation of a sentence in a strictly left-right fashion.

In the light of such examples, then, we need to ask why satisfaction of the phonological constraint is not enough in the examples in (94–97).

¹³⁵ Both of these assumptions about the phonology–syntax relation have many implications which I have barely begun to explore. Also, they contradict standard assumptions about both the ordering and the degree of ‘encapsulation’ of phonology and syntax (Zwicky & Pullum 1986, Pullum & Zwicky 1988 and many of the papers in Inkelas & Zec 1990). Nevertheless, I think that they are probably necessary properties of a ‘realistic’ grammar (in the technical sense of Bresnan 1978) which is the basis of comprehension/production as well as grammaticality judgements. See Chapter 5 for rather more discussion of these issues.

Morphosyntax



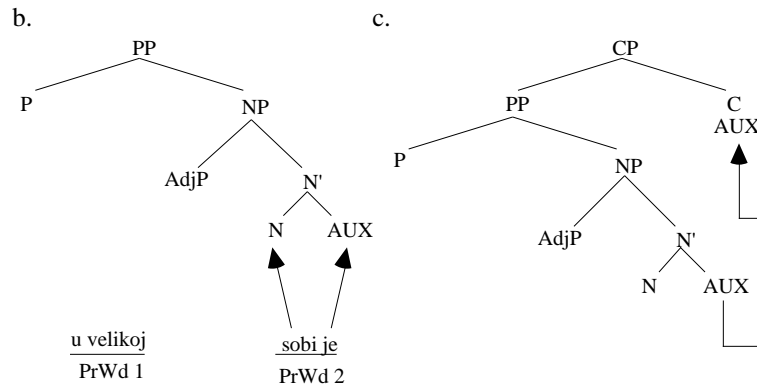
The initial input to the derivation is a prosodic word (99b), of which the clitic is the final piece.¹³⁷ This satisfies the clitic placement requirement. In (99c) the parts of the PrWd are projected onto syntactic structure, and in (99d) the second PrWd is input and projected, completing the sentence initial PP. After that, the clitic auxiliary is moved from its position inside the PP to a position where it is the head of CP (99e). If the underlying position of the auxiliary is some other position, then the auxiliary may subsequently be moved there by normal head lowering operations.

The derivation in (99) exemplifies prosodically defined clitic placement. (100) illustrates a syntactically defined clitic placement. (100) is just like (99) except that the clitic auxiliary and the head noun have been inverted, so that the auxiliary is now introduced at the end of the second PrWd. We join the derivation at the point at which the second PrWd is input to the derivation.

- (100) a. U velikoj sobi **je** klavir.
in big aux room piano
‘In the big room is the piano.’

¹³⁷ In fact, all that matters to the current account is that each PrWd corresponds to a syntactic constituent. This requirement is straightforwardly imposed if PrWds are the input to syntactic derivations, but it could also be imposed if the input to the derivation consists of smaller units.

Morphosyntax



The derivation in (100) satisfies the clitic placement requirement when the PrWd *sobi je* is syntactically projected and attached to the rest of the PP. At this point the clitic is right-peripheral in the first syntactic constituent of the sentence.¹³⁸

Having seen how well-formed clitic placements are derived, we now turn to a pair of impossible examples. The first case is a variant on (99a) and (100a) in which the determiner *ovoj* ‘this’ has been added before the adjective. I assume that the string *u ovoj* constitutes the first PrWd and that the string *velikoj je* is the second PrWd.¹³⁹

- (101) a. ?* U ovoj velikoj **je** sobi klavir.
 in this big aux room piano
 ‘The piano is in this big room.’

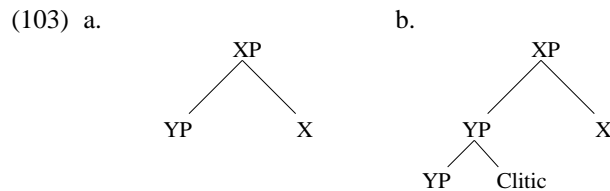
¹³⁸ Another possibility would be for the clitic to project directly onto C⁰. However, this would violate the requirement that PrWds correspond to syntactic constituents.

¹³⁹ A note on the status of the N_{empty} node in (101b). I assume that empty nodes may be created at the right edge of a structure in situations where preceding syntactic elements cannot directly merge with one another, because they do not bear any direct syntactic relation (e.g. feature-checking) relation to one another.

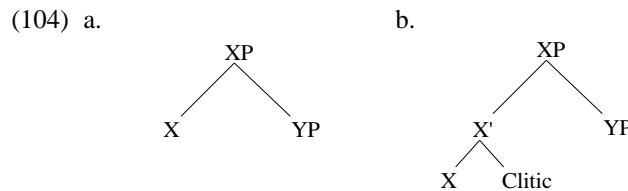
Morphosyntax

(102b) satisfies the condition on clitic placement, because the clitic auxiliary is right peripheral in the first PrWd of the sentence. However, the structure that is built violates the condition in (98d), which prevents the clitic from appearing between the head of XP and any post-head material inside XP.

As it stands condition (98d) is somewhat less than explanatory. In particular, it offers no insight into why this particular position rather than some other position is unavailable to clitics (e.g. immediately pre-head, fourth word, etc.). However, I think there is a reason why this particular restriction holds. Recall that I assumed that when clitics are merged at the right of another element, the host rather than the clitic projects (98e). Consider the different effects that this has when a clitic appears before a head and when it appears after a head. (103) shows that if a clitic is adjoined to YP preceding the head X then the sisterhood relation that holds between X and YP in the absence of the clitic (103a) still holds when the clitic is present (103b).¹⁴⁰



If YP follows X, however, then the sisterhood relation between X and YP that holds when a clitic is not present (104a) no longer holds when a clitic is right-adjoined to X (104b), because the clitic right adjoins to X.



If we further assume that the sisterhood relations that are blocked by clitics in (94–97) hold as grammatical requirements, either because the head selects the following XP or because the following XP modifier is predicated of the head, then these examples are ruled out without appealing to the stipulation in (98d), and therefore the condition in (98d) can be dropped.¹⁴¹

¹⁴⁰ Marantz (1988, 1989) argues that Merger under Adjacency is blocked in situations in which merger creates a relation that did not exist prior to merger. These examples of clitic placement look like the opposite of this situation, in that the clitic is blocked from breaking up a relation that would otherwise hold between the elements on either side of the clitic.

¹⁴¹ Note that this account of fortresses commits me to the assumption that N(P) modifiers are sisters of the head noun in Serbo-Croatian. This assumption differs from what I assume about the position of relative clauses in English in Chapter 3. This could belie a deep problem, or it could reflect a cross-linguistic difference.

The fact that the clitic is eventually moved to another position where it does not intervene between the head and its complement or modifier is not sufficient to restore the needed sisterhood relation, because the surface position of the clitic is syntactically represented, and the movements that the clitic undergoes are normal syntactic movements, which I assume to leave a trace/copy.

Note that this kind of account is not possible in a theory in which the clitic placements in (94–97) are only phonologically represented (for example, as the output of Prosodic Inversion).

The restriction that this section describes could be described in a standard syntactic approach in which syntactic processes feed morphological processes, provided that syntactic structures are not lost at the input to the morphological component. However, if it is the case that fortresses are impossible because they violate syntactic requirements, then we should expect fortress-positions of clitics to be represented in the syntax. But in a theory in which syntactic structures feed phonological structures, and which also allows post-syntactic clitic movement operations, it should be possible to move a clitic into a fortress position *after* the syntax, and thereby violate no syntactic requirement. In the left-right derivations on the other hand, the structures in which clitics appear are part of the *input* to the rest of the syntactic derivation, and for this reason there is every reason to expect them to interfere with syntactic relations.

In sum, this section has shown some advantages of the Merge Right/decomposition approach to head movement for the treatment of clitic placement in Serbo-Croatian. By inverting the derivational ordering of syntactic and phonological processes, it allows the empirical domain of syntactic head movement to be extended to cases of clitic movement which have generally resisted a syntactic characterization. In order to do this I needed to assume that phonologically determined clitic placements are represented syntactically rather than just phonologically. I showed that this makes it easier to explain some cases of ‘fortresses’ to clitic placement which are problematic for theories which assume disjunctive ordering of syntactic and phonological conditions on clitic placement.

In §4.4.2 I continue to explore the status of clitic placement in left-to-right derivations, but from a cross-linguistic perspective.

4.4.2 A Linear Asymmetry in Clitic Placement

In §4.4.1 I showed how apparent phonology-syntax mismatches in the placement of clitics in Serbo-Croatian may be captured in terms of rightward syntactic head movement, rather than by the separate morphophonological transformations that are sometimes invoked to explain these effects. In this section I look at a broader range of cases of clitic placement from a number of languages, and argue that a typological gap noticed by Hale & Selkirk (1987) and Marantz (1988a) corresponds exactly to the kind of clitic placement that cannot be ‘undone’ by the rightward head movement operations allowed by Merge Right and discussed in this chapter so far. The examples considered here are drawn primarily from Klavans 1985 and Marantz 1988a.

The simpler cases of phonology–syntax mismatches in clitic placement involve situations in which the clitic creates phonological constituents which are unexpected on syntactic grounds, but the syntactically expected ordering of

Morphosyntax

elements is preserved. Two well known examples of this from English are shown in (105–106). The possessive suffix *-s* should be the sister of an entire NP on syntactic grounds, but phonologically it just attaches to the final word of the NP, regardless of what that word is (105). The reduced future tense auxiliary *'ll* attaches phonologically to the final word of the subject NP, although it is syntactically an independent head (106).

- (105) a. [[The guy over there] 's] boat (syntactic bracketing)
 b. [[The guy over there's]] boat (phonological bracketing)
- (106) a. [[I] [will [go to the food trucks]]] (syntactic bracketing)
 b. [[I'll] [go to the food trucks]] (phonological bracketing)
 c. [[The people coming to the talk'll] [go to the food trucks first]]

Some related cases of clitic placement from other languages are shown in (107–109). The Nganhcara indirect object agreement marker presumably combines syntactically with the verb, but phonologically it suffixes to the preceding NP, which can bear any of various thematic roles (Smith & Johnson 1979, cf. Klavans 1985).¹⁴²

- (107) a. nhila pama-ng nhingu pukpe-wu ku÷a **wa=ngu**
 he.nom man-erg him.dat child-dat dog give=dat.3s
 'The man gave the dog to the child.'
- b. nhila pama-ng nhingu pukpe-wu **ku÷a=ngu** wa:
 c. nhila pama-ng ku÷a nhingu **pukpe-wu=ngu** wa:
 d. nhila pama-ng ku÷a pukpe-wu **nhingu=ngu** wa:
 e. ku÷a nhingu pukpe-wu nhila **pama-ng=ngu** wa:
 f. ku÷a nhingu pukpe-wu pama-ng **nhila=ngu** wa:
 (Nganhcara: Smith & Johnson 1979)

Kwakwala object clitics show a similar distribution: they are phonologically attached to the NP that precedes the NP that they mark (Boas 1947, Levine 1980, Anderson 1984). The same situation occurs in Yagua (109: Payne & Payne 1990). The first line shows the phonological bracketing, the gloss shows the syntactic bracketing.

- (108) nɛp'idi-da gɛnanɛm=x.a gukw=sa t'isɛm (phonology)
 throw-deic child [obj house] [obl rock] (syntax)
 'The child hit the house with a rock by throwing.'
 (Kwakwala: Levine 1980)

¹⁴² Note that this characterization of the Nganhcara clitic placement facts corresponds to the one given by Marantz 1988, which differs from the one presented in Klavans 1985. Klavans assumes that the indirect object agreement marker combines syntactically with the whole clause, rather than with the verb, which makes it appear to be a penultimate position clitic. See Marantz 1988a for discussion.

Morphosyntax

- (109) Sa-saq̣y Alchico=ni'i' Rospita=rà p̣q̣'q̣ (phonology)
 agr-give Alchico [3s.obj Rospita] [inan.obj bread] (syntax)
 'Alchico gives bread to Rospita.' (Yagua: Payne & Payne 1990)

More complex cases of phonology–syntax mismatches involve situations where the linear order as well as the phonological bracketing deviates from what is expected on syntactic/semantic grounds alone. The Serbo-Croatian examples discussed in §4.4.1 are one example of this, and a possibly related case is found in Tohono O'odham (Papago), as shown in Pranka 1983 and Hale & Selkirk 1987. O'odham auxiliary verbs exhibit a similar second position requirement to Serbo-Croatian clitics. This requirement may be satisfied by an entire XP or X⁰ in initial position, as the examples in (110) show.

- (110) a. Huan 'o g wisilo ceposid.
 John aux det calf brand.imp
 'John is branding the calf.'
- b. Wisilo 'o ceposid g Huan.
 calf aux brand.imp det John
 'John is branding the calf.'
- c. Ceposid 'o g Huan g wisilo.

However, the auxiliary may also split up sequences which cannot be separated by anything other than an auxiliary, as we saw in Serbo-Croatian. For example, the sentences in (111) show that the negative particle *pi* always immediately precedes the verb, regardless of the position of the verb, i.e. *pi-V* is an inseparable sequence. The one exception to this, shown in (112) is when *pi* and the verb are separated by the auxiliary.¹⁴³

- (111) a. Huan 'o g wisilo pi ceposid.
 John aux det calf neg brand.imp
 'John is not branding the calf.'
- b. Wisilo 'o pi ceposid g Huan.
- (112) Pi 'o ceposid g Huan g wisilo.
 neg aux brand.imp det John det calf
 'John is not branding the calf.' (Pranka 1983)

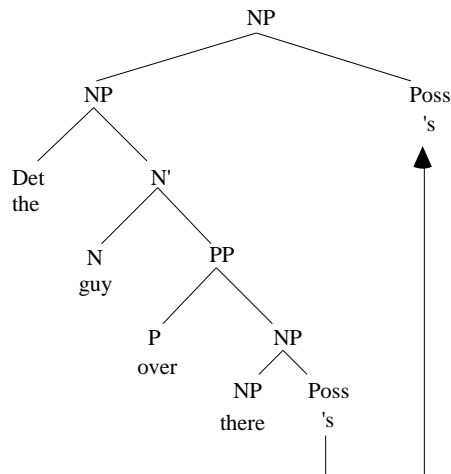
A property that all of these examples of phonology-syntax mismatches in cliticization share is that the 'misplaced' clitics can be moved from their phonologically defined position to their appropriate syntactic position by

¹⁴³ This is a slight simplification, because there are additional particles in O'odham which can intervene between *pi* and the verb, e.g. the locative particle *'am*. However, this does not undermine the point about the auxiliary. The order *pi-'am-verb* is rigidly fixed, wherever the verb appears in the sentence, and it can only be interrupted by an auxiliary in second position.

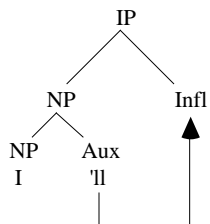
Morphosyntax

rightward head movement. (113) and (114) show the relevant derivational steps for English possessive-*s* and auxiliary *-ll*, but the same operations can be used to map any of the clitic placements in this section onto their underlying syntactic position.

(113)



(114)



It is not difficult to conceive of cases of clitic placement which could only be undone by leftward head movement, but such cases are strikingly absent from the survey above. For example, we do not find examples where case or agreement markers on NPs cliticize phonologically onto the *following* NP, which would be the mirror image of the Kwakwala and Yagua examples above. Likewise, we do not find a mirror image of English *-ll* in which an auxiliary cliticizes onto the leftmost word of an arbitrarily complex NP. This gap in the typology of clitic placement phenomena has been observed by Hale & Selkirk (1987) and Marantz (1988a), in the context of frameworks which do not predict such a gap.¹⁴⁴ If phonology-syntax mismatches in clitic placement are derived

¹⁴⁴ Note: the Irish preposition-determiner combinations discussed in Pranka 1983 would be a counterexample to this, if it were the case that the P cliticizes onto the

in the way proposed here, then the absence of such examples is expected, as there is no way of deriving them syntactically. If this generalization holds up, then it provides an argument in favor of the rightward head movement approach to clitic (un)placement.¹⁴⁵

The next section documents a further linear asymmetry in head movement which can be explained by the approach to structure building proposed here.

4.4.3 The Adjacency Penalty on Head Lowering

Lapsing into traditional bottom-to-top building terminology for a moment, the most cursory glance at the literature on head movement confirms that head raising operations are cross-linguistically far more common than head lowering operations. The contrast between the two is so strong that it has been considered feasible to claim that head lowering is, in fact, non-existent (cf. Chomsky 1993).

The claim that head lowering does not exist is faced with a small number of problem cases, which have been discussed at great length in the literature. The most notable of these comes from English, and involves the fact that inflected verbs appear to surface in the underlying position of the verb, whereas inflected verbs in many other languages appear in the underlying position of inflection. The arguments for this view are found in many other places so I will not repeat them here (cf. Emonds 1976, 1977, Pollock 1989).¹⁴⁶ Other cases have been presented from Irish (McCloskey 1996) and Bambara (Koopman 1992).

Det, rather than vice versa. I see no evidence either way in Pranka's discussion of these cases.

The Irish P-Det combinations are special in that the determiner does not have to be the head of the DP complement of the preposition. The determiner may be part of the subject of the phrase that the preposition selects.

There are good reasons to believe that P-Det combinations in French, Spanish etc. (*du, del*) involve attachment of the preposition to the determiner, rather than vice versa, but these are situations in which the DP is the complement of the P. As Jonathan Bobaljik points out, a French configuration parallel to Pranka's Irish examples might be instances of P [DP & DP] coordinations. P-det combinations are apparently marginal in such contexts, but it is difficult to disentangle this from the fact that the coordination is already marginal, even before P-det combinations are taken into consideration.

¹⁴⁵ Pranka 1983 takes evidence from intonational phrasing in O'odham to argue that the position of the auxiliary in examples like (112) is derived by moving the auxiliary into the NEG+Verb cluster from a left peripheral position. If correct, this argument is a counterexample to my generalization about the typology of clitic placements. In §4.5 below I discuss the syntax of intonation in O'odham in more detail, and argue that auxiliary placement in this language is not a counterexample to the generalization.

¹⁴⁶ Note that a number of recent analyses of verb positioning in English have assumed that the verb does undergo limited head movement (Pesetsky 1989, Johnson 1991, Koizumi 1993, 1995). However, these analyses all share with more traditional analyses the assumption that main verbs in English do not move as far as the underlying position of inflection.

Morphosyntax

A number of different approaches to this problem have been suggested, including lowering as a ‘language particular process’ (Chomsky 1991), post-syntactic merger of the verb and inflection under adjacency (Halle & Marantz 1993, Bobaljik 1994, 1995), and a ‘checking’ approach to head movement according to which words enter the syntax fully inflected and undergo head movement just for purposes of checking the well-formedness of the inflections that they carry (Chomsky 1993, Lasnik 1995b). Under this last approach English verbs raise to Infl to have their inflectional features checked in the mapping to LF.

An important observation, due to Bobaljik (1994, 1995), is that many of the purported instances of head lowering are subject to the additional requirement that the head that lowers be linearly adjacent to the position that it lowers to.¹⁴⁷ Such requirements do not, in general, apply to head raising operations. In particular, this requirement does not apply to the raising of verbs to Infl or C. Lasnik 1995a comments on the unusualness of the adjacency requirement on lowering, and speculates that this is the penalty that UG imposes on so marked a process as head lowering.

I think that left-to-right syntactic derivations provide a possible explanation for why the adjacency requirement on head lowering operations should hold. I illustrate with the case of English verbal inflection. My treatment of English inflection is most closely related to the analysis in Halle & Marantz 1993 and Bobaljik 1994, 1995, both of which are revivals of the analysis given in Chomsky 1957. I lay no claim to originality in the details of the analysis. Rather, the interest of this example here is that the left-to-right approach to structure building provides a possible explanation for *why* Infl lowering in English should have the properties that it has.

Recall that in all of the examples of decomposition of complex heads discussed in this chapter so far decomposition proceeds by moving one head *rightwards* out of another head. This is the only option permitted by Merge Right. A decomposition operation which required leftward movement would not be possible.

Now suppose that an English inflected main verb was inserted in the V position in a negated sentence such as **John not runs*. Restoring the inflectional head to its underlying position in this instance would require leftward movement to a position preceding negation. This is impossible. Therefore, in order to avoid such a situation, English allows the alternative of *do*-support, which inserts the inflection as a separate head from the verb, attached to the dummy verb *do*, when attaching the inflection to the verb would require an illicit leftward movement.

How then is it possible for inflection to *ever* appear attached to the main verb in English? This is where adjacency becomes relevant. There is exactly one situation in which inflection would need to undergo non-string-vacuous leftward movement in order to get from the position of the verb to its underlying position: that situation is when Infl and V are string adjacent. When the positions V and Infl are string adjacent the inflected verb may be directly

¹⁴⁷ It appears that certain elements, such as adverbials, do not count in the determination of linear adjacency.

projected from the phonological representation into a structure in which both V and Infl occupy their underlying positions.

This sketch of an analysis is essentially identical to that given by Halle & Marantz and by Bobaljik, except with the reverse ordering of derivational steps. What I hope to have added to their analysis is an explanation for why the extra requirement of adjacency holds in English and a handful of other cases.

4.5 The Syntax of Intonation in O'dham

Unlike in English, the intonation of sentences in Tohono O'dham (Papago¹⁴⁸) is fully predictable from the surface ordering of words. This makes it possible to investigate fairly directly the structural properties that intonational phrasing is sensitive to. The received view of intonational phrasing in O'dham is that there is a mismatch between phonological phrases and syntactic phrases (Hale 1975, Hale et al. 1977, Pranka 1983, Hale & Selkirk 1987).

In this section I discuss two features of intonational phrasing in O'dham which are relevant to the left-to-right approach to structure building. First, I reconsider the claim that phonological constituents do not correspond to syntactic constituents, in the light of the different constituents that are available in the left-to-right construction of phrase markers. I argue that phonological phrases are constituents under this approach. Second, intonation phrasing in O'dham has been used by Pranka (1983) to argue for a particular syntactic analysis of second position clitic placement in O'dham, which if correct is a counterexample to my generalization in §4.4.2 that the phonology–syntax mapping for clitics can always be derived by rightward movement. I argue that the O'dham clitic placement facts are quite consistent with my generalization about phonology–syntax mismatches in clitic placement from §4.4.2.

4.5.1 Basic Facts

Intonational phrases in O'dham follow the regular pattern (L)HL. The distribution of high and low tones is predictable from a combination of stress and syntax. For words in isolation, tones are determined in the following manner (cf. Hale 1975).

- (115) O'DHAM TONE ASSIGNMENT (Word level)
- a. The first and last stressed vowel, and all intervening vowels, are assigned an H.
 - b. Vowels following the last stress are assigned an L.
 - c. The final vowel of the word is assigned an L.

Stress almost always falls on the first vowel of the stem of a word. Prefixes and suffixes do not receive primary stress. Some examples of how these rules derive word-level tone assignments are shown in (116). As the

¹⁴⁸ Tohono O'dham is a language of southern Arizona. In recent years the group has switched to using the name *Tohono O'dham* 'desert people', rather than the earlier name *Papago*, which meant 'bean eaters' in the language of a neighboring group.

Morphosyntax

examples show, the same (L)HL pattern is followed, regardless of the length of the word.

- (116) a. ó'odham, 'person'
 H L L
 b. gátwid, 'to shoot'
 H L
 c. ha-jéwed.-ga, 'their-land-alienable'
 L H L L
 d. pápalóodi, 'windmill, kite'
 HHHLL
 e. tóki-béhe°dam-máagina, 'cotton-picker machine'
 H H HH H HL L L

The same principles are applied to derive tone assignment in phrasal units, when the phrase is head final. The noun *Húsi* 'Joe' is assigned the tone pattern HL when spoken in isolation, but when it is embedded inside a larger NP the tone pattern changes, as (117) illustrate. The second syllable of the word no longer receives a low tone when it is not phrase final.

- (117) a. Húsi, 'Joe'
 HL
 b. Húsi 'ógog, 'Joe's father'
 HHHL
 c. Húsi 'ógog kíi, 'Joe's father's house'
 HH HH HL

It is not just NPs that observe the (L)HL intonation pattern. The same is true of PPs and clauses (118).

- (118) a. g Húsi kíi (HHHL) NP
 art Joe house
 'Joe's house.'
- b. 'am múisa wéco (LHHHHL) PP
 loc table under
 'under the table.'
- c. No g wákial cíkpan? (LHHHHL) IP
 inter.aux.3.imperf art cowboy work.imperf
 'Is the cowboy working?'

When the phrase is head final these syntactic phrases correspond to a single intonational phrase. O'odham also allows the arguments and specifiers to appear following the head, in which case the intonational phrasing is different. When the phrases appear to the right of the head they constitute a separate intonational phrase, as the counterparts to (118a–c) in (119) show.

- (119) a. g kíi-j g Húsi (HL) (HL) NP

Morphosyntax

art house.3s art Joe
 'Joe's house.'

- b. 'am wéco g mísa (LHL) (HLL) PP
 loc under art table
 'under the table.'
- c. No cíkpan g wákial? (LHL) (HLL) IP
 inter.aux.3.imperf work.imperf art cowboy
 'Is the cowboy working?'

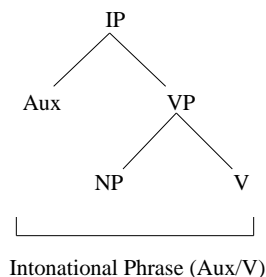
I suggest the following characterization of intonational phrase domains in O'odham. The boundaries of an intonational phrase are determined by a lexical head (N, V, P) and a higher functional head associated with that lexical head (Det, Infl, locative marker respectively). Intonational phrases extend from the functional head to the lexical head, plus they are subject to the requirement that they correspond to syntactic constituents. In other words, intonational phrases correspond to the smallest constituent containing both the functional head and the lexical head.

- (120) INTONATIONAL PHRASING (O'odham)
 For any pair of heads consisting of a functional head (Det, Infl, Loc) and its associated lexical head (N, V, P), the smallest constituent containing both the functional and the lexical head constitutes an intonational phrase.

First I will show how this characterization accounts for the alternations in (118–119), and then we turn to the issues raised by auxiliary placement and its interaction with intonation in O'odham.

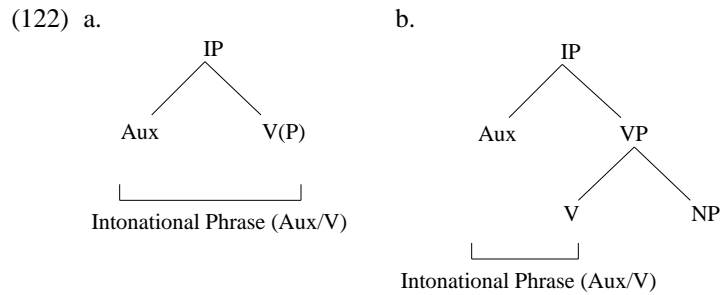
Consider the examples in (118c) and (119c) involving the intonational domain of an auxiliary/verb combination. We may assume that the relative structuring of the words in (118c) is as in (121), in which the auxiliary occupies the head of IP and the subject is a daughter of VP. In fact, it does not matter what the node labels are: however the words are structured, the smallest constituent containing both the auxiliary and the verb will always contain the subject NP.

- (121)



Morphosyntax

The phonological phrasing of (119c), on the other hand, depends on an intermediate stage in the derivation of the sentence. (122a–b) show the structure of the sentence before and after the subject NP is added to the structure. Before the subject is added there is already a constituent containing the auxiliary and the verb—this is the extent of the intonational phrase. If the subject is subsequently introduced as a daughter of VP (122b), then the smallest constituent containing the auxiliary and the verb will be larger than the constituent that determined the intonational phrasing, but this has no effect on the intonational phrasing. Since the subject NP is not included inside the intonational domain of the verb and auxiliary, it forms its own intonational phrase.



In this way it is possible to characterize the extent of intonational domains in O'dham in terms of the 'minimal constituent' containing a given pair of elements. Exactly the same effects would be achieved if 'minimal constituent' was replaced with 'earliest constituent in the derivation to contain the two elements'. This would still pick out the IP in (122a) as the correct intonational phrase, because it exists before the structure in (122b) is built.

4.5.2 Clitic Auxiliaries and Intonation Phrases

As we have seen already, finiteness in O'dham is marked by an auxiliary. The positioning of this auxiliary is described (correctly) as being 'second position', but as with the Serbo-Croatian clitics discussed in §4.4.1 the greater part of the analytical work involves characterizing what can count as 'first position'. First position may be filled by arguments, adjuncts, verbs, PPs, possessors, particles, parts of complex *wh*-phrases, and floated quantifiers. Some examples are given in (123–124), together with their intonational phrasing.

- (123) a. (L HHH HHH H L) *Single Int. Phrase*
 ... ma-t g wákial g wísilo cépos.
 comp-aux.3.perf art cowboy art calf brand.perf
 '... that the cowboy branded the calf.'
- b. (HH HL)
 D óo p ñéid?
 who aux.2s see.imperf
 'Who do you see?'

Morphosyntax

- (124) a. (HLL) (L HHH HL) *Separate pre-Aux Int. Phrase*
 Wákial 'at g wísilo cépos.
 cowboy aux.3.perf art calf brand.perf
 'The cowboy branded the calf.'
- b. (HLL) (L H L) (H LL)
 Wísilo 'at cépos g wákial
 calf aux brand.perf art cowboy
- c. (H L) (L H LL) (H LL)
 Cépos 'at g wákial g wísilo.
 brand.perf aux art cowboy art calf

In all of the examples in (123–124) the auxiliary appears in second position in the sentence. Importantly, however, the examples in (123) differ from those in (124) with respect to their intonational phrasing. In (123) the first position material forms part of the same intonation phrase as the auxiliary and the verb, whereas in (124) the first position material forms an intonational phrase of its own. Accounting for this contrast is the problem that we will be concerned with for most of the rest of this section.

4.5.2.1 XP and X⁰ as First Position

The table in (125) lists the intonational phrasing of all of the first position elements discussed in Hale & Selkirk 1987.

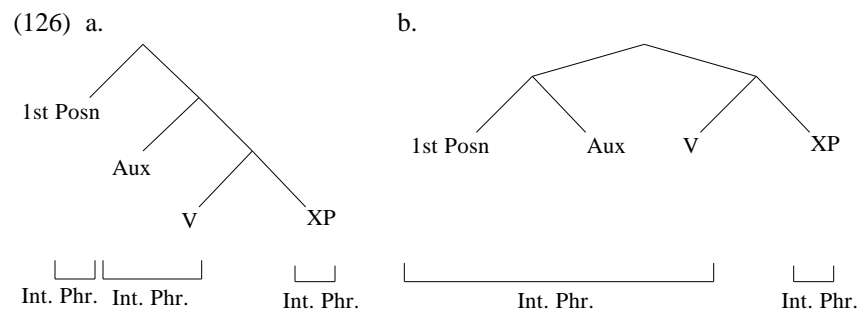
(125)

	SEPARATE INTONATION PHRASE FROM AUX		SAME INTONATION PHRASE AS AUX
a.	Subject NP	i.	Particles (Neg etc.)
b.	Object NP	j.	Complementizers
c.	VP-adjuncts	k.	possessive particles
d.	<i>then</i> -proform	l.	floated quantifiers
e.	pronouns	m.	<i>wh</i> -words
f.	possessor NPs	n.	left-branch of <i>wh</i> -phrase
g.	verbs		
h.	modifier of predicate nominal		

Consider how this set of facts may be accounted for under the assumption that the smallest constituent containing the Aux and the V forms an intonational phrase. In this approach we can characterize the configurations that give rise to the two intonation patterns in category neutral terms, as shown in (126a–b). (126a) shows a situation in which the first position material is the sister of a constituent which contains both the verb and the auxiliary. In this case the first position material is not included in the intonational phrase of the Aux/V, because it is not part of the smallest constituent containing the Aux and the V. (126b) shows a situation in which the auxiliary and the verb do not form a

Morphosyntax

constituent to the exclusion of the first position material. In this case the smallest constituent containing Aux and V also includes the first position material, and therefore there is no separate intonational phrase for the first position material.



Notice that according to the analysis I am proposing for intonation phrasing in O'dham, the determination of whether or not the first position material is part of the Aux/V intonational phrase does not depend on the category label or projection type (X^0 vs. XP) of the first position material, merely on whether the auxiliary and the verb form a constituent to the exclusion of the first position material. Here my analysis diverges from Hale & Selkirk (1987), who attempt to explain the contrast in (125) in terms of the X^0 /XP distinction.

For example, one possible realization of the independent-intonation configuration in (126a) would be a situation in which the first position material is an XP occupying the specifier of IP and the auxiliary is the head of IP. This could account for the intonational phrasing when the initial material is an argument XP (125ab) a VP-adjunct (125c), or a demonstrative or temporal proform (125de, cf. 127).

- (127) a. (HL) (LH L)
 'Hég 'o cémaj.
 that aux.3s small
 'That (one) is small.'
- b. (H L) (L L HL) (L H H HH H L)
 'E' d a 'ant o cípk [ma-nt hékid o 'i wám]
 then aux.1s fut work.perf [comp-aux.1s when fut incep wake.perf
 'Then I will work, when I wake up.'

However, not all examples look like they involve fronting of an XP to SpecIP. Examples with a verb in initial position (e.g. 128) are less likely candidates for an analysis in which the initial position material is in SpecIP. There is no evidence that the fronted verb in sentences like (128) is actually a fronted remnant VP, from which all of the arguments have been extracted. If (128) involved a fronted remnant VP this would predict the possibility of fronting the verb together with its arguments. As far as I know, O'dham does not allow the verb and its argument(s) to precede the auxiliary simultaneously.

Morphosyntax

- (128) (H L) (L H LL) (H LL)
 Cépos 'at g wákial g wísilo.
 brand.perf aux art cowboy art calf
 'The cowboy branded the calf.'

However, if the verb in (128) occupies a head position above IP (e.g. Comp or Focus), then it still falls outside the smallest constituent containing the Aux and the underlying position of the verb, and therefore it forms an independent intonational phrase.¹⁴⁹

4.5.2.2 XP-internal Auxiliaries in *Wh*-Questions

Turning to the fillers of the first position that do *not* form an independent intonational phrase (second column in (125)), we find that these also do not obviously form a homogeneous class in terms of their categorial status. Some are canonical instances of heads, but other members of this class are elements which in other languages often show XP behavior. For example, the interrogative particle in (129a), the negative particle in (129b) and the complementizer in (129c) are all likely candidates for X⁰ elements, which are either adjoined to the Aux (130a) or which the Aux has adjoined to (130b).

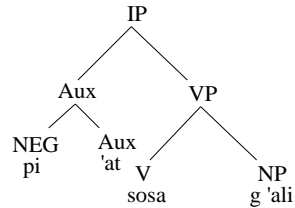
- (129) a. (L HH H H L)
 Na-t g 'áli pi s ós a?
 inter-aux art child neg cry.perf
 'Did the child not cry?'
 b. (L L H L) (H L)
 Pi 'at s ós a g 'áli.
 neg aux cry.perf art child
 'The child did not cry.'
 c. (L H H H L)
 Ku-p háscu táccu?
 comp-aux.2s what want
 'What do you want?'

¹⁴⁹ A concern about this analysis is that these fronted verbs are claimed not to receive any special focus interpretation (Ken Hale, pc). If this is the case, then I predict that they should not be able to head their own sentence initial projection, and therefore should not be able to undergo head movement across the Aux.

An alternative analysis of these cases could assume that the first intonation phrase in (128) begins with the fronted verb and *ends* with the auxiliary.

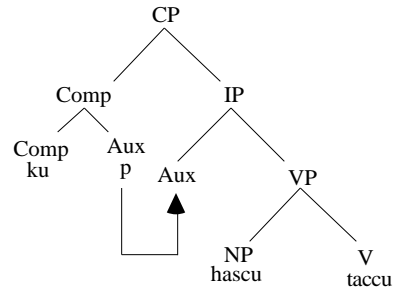
Morphosyntax

(130) a.



= (129b)

b.



= (129c)

The most surprising cases of first position elements that do not form an independent intonational phrase are the *wh*-words (125m–n). This includes *wh*-words which form a complete *wh*-phrase and subparts of a *wh*-phrase (131), when a possessive NP is questioned (132).

(131) a. (HH HL)
 D óo p ñéid?
 who aux.2s see.imperf
 ‘Who do you see?’

b. (HH H HL)
 Báa pt o híi?
 where aux.2s.perf fut go.perf
 ‘Where are you going to go?’

(132) (HH H H HL)
 D óo p gógs-ga ñéid?
 who aux.2s dog-poss see.imperf
 ‘Whose dog do/did you see?’

In more familiar languages *wh*-movement shows properties of phrasal movement rather than head movement, so an initial expectation would be that fronted *wh*-phrases in O’odham should also show properties of XPs. But most XPs form independent intonational phrases when they occur in sentence initial position, unlike the *wh*-words. Hale & Selkirk (1987) treat the cases of sentence initial *wh*-phrases as the result of head movement, which enables them to maintain the generalization that first position XPs form independent intonational phrases, whereas first position X⁰s do not. I suggest a different approach to these facts, based on some novel data which indicate that examples like (132) do not involve movement of the possessor *wh*-phrase away from the possessed NP, but instead involve placement of the auxiliary inside the complex NP, in a manner similar to what we have seen in Serbo-Croatian in §4.4.1.¹⁵⁰

¹⁵⁰ I am grateful to Ken Hale and Albert Alvarez for providing me with this data.

Morphosyntax

The possessor *wh*-word and the possessed NP may be separated by the auxiliary, but not by any other material. This implies that O'dham does not allow 'left-branch' extractions of the form *whose did you drive car*, which are possible in a number of languages.

As we have seen above, O'dham allows NP arguments to either precede or follow the verb. This optionality is not available when the possessor *wh*-word fills the pre-Aux position. In this instance the possessed NP must precede the verb as the contrast between (132) and (133) shows.

- (133) * D óo p ñéid gógs-ga?
 who aux.2s see.imperf dog-poss
 'Whose dog do/did you see?'

When the verb has two overt preverbal arguments, the possessed NP must appear adjacent to the Aux (134).

- (134) a. D óo pt gógs-ga g cúukug máa?
 who aux.2s.T dog-poss art meat give.perf
 'Whose dog did you give meat to?'
- b. * D óo pt g cúukug gógs-ga máa?

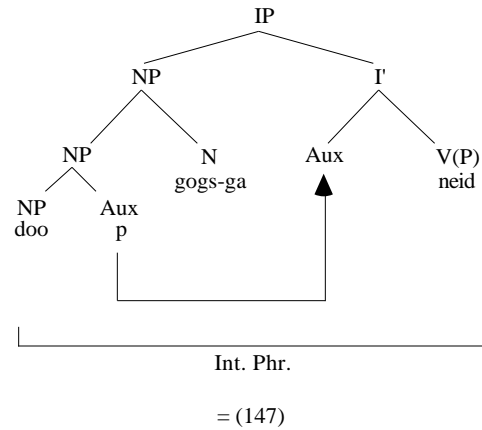
When the possessor *wh*-phrase is part of an argument in an embedded clause, the possessed NP must still occur immediately following the Aux, and cannot remain in the embedded clause (135).

- (135) a. D óo p gógs-ga hab 'elíd mápt méa?
 who aux.2s dog-poss thus think c.aux.2s.T kill.perf
 'Whose dog do you think you killed?'
- b. * D óo p hab gógs-ga 'elíd mápt méa?
- c. * D óo p hab 'elíd gógs-ga mápt méa?
- d. * D óo p hab 'elíd mápt méa gógs-ga?

The examples in (132–135) indicate that the entire *wh*-phrase (i.e. *whose dog*) is a syntactic unit occurring in sentence initial position and that the Aux in examples like these is positioned internal to the *wh*-phrase, in the same way as clitics are positioned XP-internally in Serbo-Croatian, and that it is subsequently moved out of the *wh*-phrase to become the head of the IP projection. Given that the surface position of the Aux is internal to the *wh*-phrase, the smallest constituent containing the Aux and the verb also includes the *wh*-phrase, as (136) shows. The structure in (136) shows the entire *wh*-phrase as the specifier of IP, but the same prediction about intonation phrasing would obtain if the *wh*-phrase were in SpecCP.

- (136)

Morphosyntax



This account of the XP-internal Aux placements makes it possible to capture the intonation phrasing of possessor-NP questions without being forced to assume that the possessor-NP undergoes head movement. In this respect it fares better than the account of Hale & Selkirk (1987), who are forced to leave the X^0 status of *wh*-phrases as an unexplained property.

However, this account leaves open a couple of questions for which I do not have an answer at this point. First, why must the Aux follow the possessor *wh*-word rather than the entire NP, given that the Aux has the choice of appearing either internal to or following a possessive NP when it does not contain a *wh*-word (137)? Notice that in both cases the pre-Aux material forms an independent intonation phrase.

- (137) a. (HH H L)(L HL)
 Húan gógs-ga 'añ ñéid.
 John dog-poss aux.1s see.imperf
 'I see/saw John's dog.'
- b. (HL) (L H H HL)
 Húan 'añ g gógs-ga-j ñéid.
 John aux.1s art dog-poss-3s see.imperf
 'I see/saw John's dog.' (Hale & Selkirk 1987)

A possible reason for the difference between (132) and (137b) is that (137b) is the result of independent movement of the possessor NP and the possessed NP, i.e. these cases are genuine 'left-branch extractions.' In support of this, notice that the possessed NP contains an article and an agreement suffix in (137b) and not in (137a), whereas the same possessed NP appearing in the same linear position in (132) lacks the article and the agreement suffix. This implies that the two examples with possessor-NPs are syntactically different.

Second, how does my account of *wh*-questions with possessive NPs generalize to examples like (131) containing a simple *wh*-phrase, in which the Aux follows the entire *wh*-phrase but the *wh*-phrase is still a part of the Aux/V intonation phrase? In order to account for the intonation phrasing of simple *wh*-

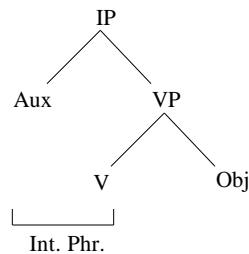
initial questions I must assume that the Aux is required to be adjoined to the *wh*-phrase when it enters the syntax, but it is not clear why this should be the case.

I leave both of these questions open for the time being.

4.5.3 Comparison with Hale & Selkirk 1987

Let us compare this analysis of intonation phrasing in O'odham with the analysis presented in Hale & Selkirk 1987 (henceforth H&S). H&S assume that intonational phrases do not correspond to phrase structure constituents in O'odham. This conclusion is hard to avoid, given their syntactic assumptions, and given that intonation phrases have boundaries between heads and their post-head complements, but also include pre-head material from separate syntactic phrases, as in the configuration in (138).

(138)



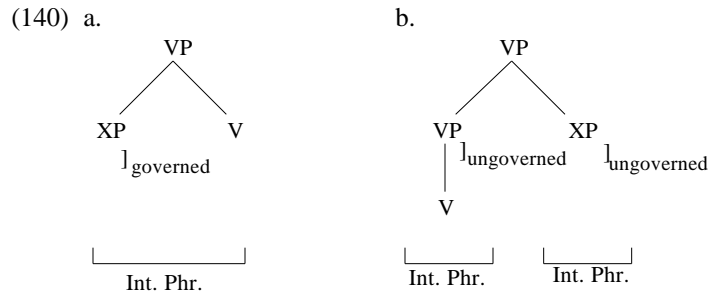
H&S instead adopt an approach in which syntactic configurations determine the *boundaries* of phonological phrases. For O'odham they assume that intonation phrase boundaries are determined by the rule in (139), which states where the right-hand edges of intonation phrases fall.

(139) PAPAGO PHRASING PARAMETER (Hale & Selkirk 1987)
 $]_{X^{\max}}$, where X^{\max} is not lexically governed.

H&S assume furthermore that heads govern material to their left in O'odham, but do not govern material to their right. The effect of this is that when an XP appears before the head that selects it (e.g. in the order NP V) the righthand boundary of the XP is lexically governed and therefore does *not* delimit an intonational phrase. On the other hand, if the head and the XP appear in the opposite order (i.e. Y XP) then XP will not be lexically governed, and therefore its right edge delimits an intonational phrase.

In order to account for the fact that the verb marks the right-hand edge of an intonation phrase, H&S assume in addition that postverbal arguments are VP-adjuncts. This ensures that the verb always occurs at the right-hand edge of a VP constituent, and since VP is not lexically governed its right-hand edge marks an intonation phrase boundary. The two relevant configurations are shown in (140a–b)

Morphosyntax



This set of assumptions captures the head-initial/head-final contrast in O'dham intonation phrasing demonstrated above. Now consider the predictions this makes about where the left-edge of the Aux/V intonation phrase should occur. For H&S the grammar of O'dham just specifies where the right-hand edges of intonation phrases fall, and therefore their theory predicts that the intonation phrasing of the pre-Aux material depends on whether its right edge is a lexically ungoverned XP-boundary or not. Given their additional assumption that neither the head nor the specifier of IP is lexically governed, this prediction reduces to the prediction that the pre-Aux material will form an independent intonation phrase if it is an XP and will be part of the Aux/V intonation phrase if it is a head.

The XP/X⁰ distinction works well for most of the types of pre-Aux material listed in the table in (125), but it leads to a couple of improbable conclusions. First, fronted verbs are predicted to be fronted VP-constituents, since they form an independent intonational phrase. I am unaware of any independent evidence that fronted verbs in O'dham are fronted remnant VPs. Second, cases where the Aux occurs in the middle of a complex *wh*-phrase must be analyzed in such a way that the pre-Aux *wh*-possessor is a head rather than an XP. However, given that *wh*-possessors are interchangeable with possessor NPs when not in pre-Aux position, there is good reason to analyze these words as XPs.

Furthermore, in addition to the question of whether the *wh*-possessor forms an independent intonation phrase, it is also relevant to ask whether the right-hand edge of the complex *wh*-phrase that the Aux splits up should be an intonation phrase boundary under the Hale & Selkirk analysis. We have seen that if an expression like *whose dog* is extracted from an embedded clause in O'dham this phrase appears in sentence-initial position, interrupted by the Aux. The relevant example is repeated in (141).

- (141) D óo p gógs-ga hab 'elíd mápt méa?
 who aux.2s dog-poss thus think c.aux.2s.T kill.perf
 'Whose dog do you think you killed?'

If we assume that the fronted *wh*-phrase in this example does not occupy a position inside the matrix VP (such positions are presumably reserved for the arguments of the matrix verb), but is instead in a position in the matrix IP or CP, then the *wh*-phrase is an XP which is not lexically governed, and therefore H&S predict that an intonation phrase boundary should fall at the right edge of the preverbal NP *gógs-ga* in (141). This prediction is incorrect.

For H&S intonation phrasing is entirely determined by the position of boundary-delimiting categories. Since intonation phrases cannot correspond to constituents, given H&S's (quite standard) syntactic assumptions, the status of an intonation phrase as a syntactic constituent or not is irrelevant. What I have tried to show in this section is that under the left-right approach to the derivation of phrase markers proposed in Chapter 2 it is possible to equate intonation phrases with syntactic constituents. This fact alone does not solve the problem of where intonation phrase boundaries fall in O'dham—the task of specifying which kinds of categories determine the size of intonation phrases remains. The advantage of this approach is that it is possible to refer to both boundaries *and* constituency relations in defining intonation phrases. In the case of O'dham I have argued that this distinction is useful in particular for the explanation of cases where the left-edge of an intonation phrase does not coincide with the category (i.e. Aux) that determines the positioning of that edge.

4.5.4 Clitic (un-)placement again

In the light of the analysis of intonation phrasing and auxiliary placement in O'dham presented in the preceding sections, we are now in a position to return to the treatment of auxiliary placement given by Pranka (1983). Pranka uses O'dham intonation phrasing facts like those discussed above to argue for an analysis of clitic positioning which contradicts the generalization proposed in §4.4.2 above.

Pranka assumes that the VP defines an intonational phrase in O'dham.¹⁵¹ She therefore assumes that when the pre-Aux material forms an independent intonation phrase it must have moved out of the VP. In cases where the pre-Aux material is part of the same intonational phrase as the verb, Pranka assumes that the Aux has lowered into the VP. In other words, the surface position of the Aux in these cases is assumed to be to the right of its underlying position. Pranka's rules for deriving the surface position of the Aux are given in (142). The first rule derives the independent intonation when the pre-Aux material is an XP, and the second rule derives the situation in which the pre-Aux material is a head in the same intonational phrase as the verb.

(142) *O'dham Aux placement rules (Pranka 1983)*

a.	[IP	Aux	[VP	XP	Y
		1		2	3
	2	1		0	3
b.	[IP	Aux	[VP	X	Y
		1		2	3
		0		2	1 3

¹⁵¹ I am adapting Pranka's terminology slightly to be consistent with the rest of the discussion here. What I am calling VP Pranka refers to as V"; what I call IP Pranka refers to as V".

Morphosyntax

If Pranka's analysis is correct, then the derivation of O'dham auxiliary placement is a problem for the left-to-right approach to structure building, because the *overt* occurrence of the auxiliary is to the right of its underlying position. This would require leftward movement of the Aux to reinstate it in its syntactically defined position. In the left-to-right approach such a movement is impossible.

Under the analysis of O'dham intonation provided above, however, there is no need for the Aux to ever appear in a surface position to the right of its underlying position. Cases which Pranka analyzed as involving movement of the Aux to a position inside VP could equally well be analyzed as involving positioning of an underlying VP element outside VP. Of course, if it is possible to simply restate any movement of Aux to the right of X as a movement of X to the left of Aux, then choosing between the two alternatives is not a particularly interesting exercise. From Pranka's perspective, however, there was a good reason to pursue the Aux-lowering analysis, because this made it possible to make the generalization that the intonational phrase containing the verb had its left-hand boundary at the left edge of VP.

As we have seen here, though, it is unlikely that we will be able to maintain the claim that the verb's intonational phrase corresponds to a fixed constituent boundary. For cases like (141) in which the Aux/V intonation phrase includes the whole of a complex *wh*-phrase that has been extracted from an embedded clause, the intonation phrase needs to extend at least as far as the left-edge of a matrix clause specifier above VP (e.g. SpecIP or SpecCP). On the other hand, for simple sentences in which the subject NP precedes the Aux and forms an independent intonation phrase, the intonation phrase needs to extend only as far as the left-edge of the Aux. In order to accommodate both of these facts in an approach like Pranka's, in which the left-edge of the Aux/V intonation phrase corresponds to a fixed position in the phrase marker, we would be forced to assume that sentence initial subject NPs always move to a surface position which is structurally *higher* than the surface position of a *wh*-phrase that has been extracted from an embedded clause. I know of no independent justification for this assumption.

Therefore, we have good reason to question the assumption about intonation phrasing which was the primary motivation for Pranka's lowering analysis of Aux placement in O'dham. In addition, I have provided arguments that the Aux sometimes appears to the *left* of its underlying position when the pre-Aux material does not form an independent intonation phrase (e.g. when it appears in the middle of a complex *wh*-phrase). Therefore, I do not think that O'dham Aux placement represents a challenge to my generalization about clitic (un-)placement from §4.4.2

To conclude this section: my analysis of intonation phrasing in O'dham draws a closer parallel between syntactic constituency and phonological constituency than is often considered to be possible. This raises the obvious question of whether claimed instances of phonology–syntax mismatches in other languages will also turn out to be non-mismatches in the left-to-right approach to structure building. This is an interesting topic which awaits further investigation.

4.6 Conclusion

This chapter has provided an extension of the left-to-right approach to syntax to the interface of syntax with phonology and morphology. The most basic consequence of Merge Right is to reverse the derivational ordering of syntactic and morphophonological processes, such that morphology and phonology now feed syntactic derivations. I demonstrated a series of things to be gained from this kind of derivation, covering locality in head movement, directionality in clitic placement and head movement, and the relationship between syntactic and phonological constituency.

The next chapter draws together the separate lines of inquiry from Chapters 2–4 and uses the results from these chapters to return to the issues raised in Chapter 1.

Appendix: More on Participle Fronting

In this appendix I discuss three languages for which my characterization of the participle fronting facts in §4.3.2 is at odds with one or more reports in the literature, and for which I therefore need to provide further justification.

A.1 Polish

In the table in (56) I claim that Polish allows participle fronting when no clitic auxiliary requires a host. There are two challenges to my characterization of Polish. First, Embick & Izvorski (1995) claim that Polish only allows participle fronting in sentences with clitic auxiliaries. Second, Borsley & Rivero (1994) argue that Polish participle fronting always necessarily involves incorporation of the fronted participle into the projection headed by the auxiliary. These claims, if true, undermine the prediction of the CCM that long-distance head movement (which Polish allows) should only be possible when the fronted head is the head of its own syntactic projection.

Facts presented by Borsley & Rivero 1994 seem to show fairly clearly that Polish does allow participle fronting which is not driven by the need of a clitic auxiliary to have a host. The Polish conditional auxiliary *by* may appear sentence initially (143a), implying that it is not a clitic, but it may also follow an overt subject (143b) and it may follow a fronted participle (143c).¹⁵²

¹⁵² Apparently sentences like (143a) in which *by* appears sentence initially are not accepted by all speakers of Polish. Booij & Rubach (1987) report that it is acceptable for some speakers, whereas Embick & Izvorski (1995) report that their informants disprefer this order. This may mean that *by* is only marginal as a non-clitic.

Morphosyntax

- (143) a. **By** widzia° tę książkę.
 cond.2s seen.masc.s this book
 ‘You would see this book.’
- b. Ty **by** widzia° tę książkę.
 you cond.2s seen.masc.s this book
- c. Ty widzia° **by** tę książkę.
 you seen.masc.s cond.2s this book

For other languages, I have taken the possibility of participle fronting with non-clitic auxiliaries (as in Polish (143c)) to indicate that fronted participles may head their own Focus projection. However, this is challenged by Borsley & Rivero’s (1994) arguments that Polish participles in sentences like (143c) are in fact incorporated into the auxiliary. They argue that Polish differs in this respect from other languages with participle fronting, in which they assume that the participle uniformly undergoes long head movement across the auxiliary to the head of CP.

Borsley & Rivero give two arguments for the claim that fronted participles incorporate into the auxiliary in Polish. The first argument is that participle fronting is possible in both root and embedded contexts in Polish, but only in root clauses in a number of other languages, such as Serbo-Croatian. Representative examples are given in (144–145). For other languages, Rivero has argued that the absence of participle fronting in non-root environments is due to the fact that participles front to C, but cannot do so when the embedded C position is already filled by a complementizer (cf. Lema & Rivero 1989, Rivero 1991, Rivero 1994). Given that Polish allows fronting in non-root contexts, Borsley & Rivero argue that the fronting must therefore target Infl rather than C, and in this respect differs from the other Slavic languages.

- (144) a. Wie, z’ę lustro **by**my kupili
 know.3s that mirror cond.1p bought
 ‘He knows that we would buy a mirror.’
- b. Wie, z’ę lustro kupili **by**my (Polish)
- (145) a. Ivan kaže, da **je** čitala Marija knjigu.
 Ivan said that is read Marija book
 ‘Ivan said that Maria has read the book.’
- b. * Ivan kaže da čitala **je** Marija knjigu. (Serbo-Croatian)

However, as Embick & Izvorski (1995) point out, in evaluating the possibility of embedded participle fronting, it is important to control for the potentially confounding effect of the second-position requirements of many auxiliaries. Some complementizers are potential hosts for second position clitics, in which case the order *complementizer-participle-clitic* puts the clitic in third position in the clause. Embick & Izvorski show that once this factor is controlled for, embedded participle fronting is possible, as their examples in

Morphosyntax

(146–147) show. (146) is a Serbo-Croatian counterpart to (145b), except that the clitic auxiliary has been replaced by a non-clitic auxiliary. (147) shows that Bulgarian, in which clitic auxiliaries are not restricted to second position in the clause, also allows embedded participle fronting.^{153,154}

- (146) On tvrđi da istukao **beja**•e Jovan Petrovog prijatelja.
 he claims that beaten was Jovan Peter’s friend
 ‘He claims that Jovan had beaten Peter’s friend.’ (Serbo-Croatian)
- (147) Razbrah•e pro•el e knjigata.
 understood that read had book.def
 ‘I understood that he had read the book.’ (Bulgarian)

Therefore, we can explain the distribution of root and non-root clause participle fronting without appeal to the distinction between long head movement (root only) and incorporation (root and non-root) that Borsley & Rivero assume.

The second argument that Borsley & Rivero give in favor of their incorporation analysis of Polish participle fronting also involves a cross-linguistic contrast among participle fronting languages. In Polish, clitic pronouns may not intervene between a fronted participle and the auxiliary, although a pronoun may intervene between the participle and the auxiliary when the participle is not fronted (148). In Bulgarian, on the other hand, clitic pronouns may intervene between fronted participles and clitic auxiliaries (149).

- (148) a. Bym to zrobil
 cond.1s it done
 ‘I would do it.’
- b. Zrobil bym to.
 c. * Zrobil to bym. (Polish)

¹⁵³ To be fair to Borsley & Rivero, they are aware of the potential confound of second position requirements, and identify the clitics of Bulgarian, which are not restricted to second position, as the relevant test cases. They report the Bulgarian sentence in (i) as bad, in support of their claim.

(i) (*) Znam•e pro•eli sme knjigata.
 know.1s that read pres.1s book.def
 ‘I know that we read the book.’ (Bulgarian: Borsley & Rivero 1994, p.384)

However, Embick & Izvorski (1995) claim that their corresponding example (147) is accepted by all of their informants. I am unsure of what underlies this factual disagreement.

¹⁵⁴ Notice that if the fronted participle in (146) is in a head position above IP (e.g. Focus), then I have to assume that the complementizer *da* can take a FocusP as its sister rather than an IP. This amounts to a form of ‘CP-recursion’. Although I have to assume that CP can dominate FP in Serbo-Croatian, I assume that FP cannot dominate CP: if this were a possibility, then we would lose the account of why participle fronting is blocked in the environment of *-li* in Serbo-Croatian.

Morphosyntax

- (149) a. Toj **g o e** viǰdal.
he him pres.3s seen
'He saw/has seen him.'
- b. Viǰdal **g o e**. (Bulgarian)

However, I am reluctant to attribute the contrast in (148–149) to a contrast in the syntactic position of the fronted participle. We find exactly the same contrast internal to Croatian, depending on whether the clitic auxiliary is first person (Polish pattern) or third person (Bulgarian pattern), as (150–151) show (examples from *avar & Wilder 1994).

- (150) a. **ǰ**itao sam ga **ǰ**esto.
read.ptc be.1s it often
'I have often read it.'
- b. * **ǰ**itao ga sam **ǰ**esto.
- (151) a. * **ǰ**itao je ga Ivan **ǰ**esto.
read.ptc be.3s it Ivan often
'Ivan has often read it.'
- b. **ǰ**itao ga je Ivan **ǰ**esto. (Croatian)

The facts in (150–151) parallel the contrast between Polish and Bulgarian. Extending Borsley & Rivero's argumentation, therefore, we might then assume that in Croatian, participle fronting involves incorporation into the auxiliary when the auxiliary is first person singular but long head movement across the auxiliary when the auxiliary is third person singular. However, I am unaware of any other evidence that would support such a syntactic distinction. The standard approach to the contrast in (150–151) is to assume that the Croatian clitic cluster is ordered by a morphological template which happens to place *je* at the end of the cluster and all other clitic auxiliary forms close to the beginning of the cluster. I therefore assume that the contrast between Polish and Bulgarian in (148–149) is also the result of morphological idiosyncrasies of how each language orders its various clitics, and does not reflect a syntactic contrast in participle fronting constructions.

Therefore, I think that there is good reason to doubt both of Borsley & Rivero's (1994) arguments for syntactically distinguishing participle fronting in Polish from participle fronting in the other languages listed in (56).

A.2 Old Spanish

In the table in (56) I list Old Spanish as conforming to the generalization that long-distance participle fronting is possible, provided that fronting is also allowed in non-clitic contexts. However, Rivero 1996 cites Old Spanish as a counterexample to this kind of generalization. Lema & Rivero 1991 argue that Old Spanish (from the period lasting roughly from the 12th to the 15th century)

Morphosyntax

has both clitic auxiliary driven participle fronting and ‘stylistic’ verb fronting constructions, but that they are quite distinct processes which should not be conflated.

Given that Old Spanish apparently did not allow long-distance participle fronting, Lema & Rivero’s argument is potentially very damaging to my generalization. Therefore their arguments deserve some scrutiny.

Old Spanish future and conditional tenses exhibit clitic driven verb fronting. This includes fronting of main verb infinitivals in simple analytic futures and conditionals, as in (152), and fronting of infinitival auxiliaries in more complex analytic tenses involving three verbs (153).

- (152) a. Dar-te hé un exemplo
give-you fut.1s an example
‘I will give you an example.’
- b. E si fuere en el ángulo de occidente, aver-lo-há en su senectut
and if cond.3s in the angle of occident have-it-fut.3s in his old age
‘And if it was in the angle of occident, he will have it in his old age.’
- (153) a. Ca en yermo o en poblado poder nos hán alcançar
since in field or in village can us fut.3p reach
‘Since they will be able to reach us in the open fields or in an urban setting.’
- b. E seer uos án perdonados pecados u<uest>ros
and be you.dat fut.3p forgiven sins your
‘And your sins will be forgiven.’ (Old Spanish)

Notice that in (153) it is the higher of the two dependent verbs which fronts to support the clitic. Lema & Rivero claim that this is in fact the only possibility in such tenses, and that the lower verb can never front. If this is the case, then Old Spanish is like Croatian in not allowing the kind of long-distance verb fronting found in languages like Bulgarian and Rumanian. Lema & Rivero’s claim is actually difficult to evaluate for a couple of reasons.

First, since they are dealing with historical data they are reliant on corpora, which allows inferences about ungrammaticality to be made only indirectly. Of course this problem can be dealt with by looking for systematic gaps in construction types found in a corpus, but Lema & Rivero unfortunately give no indication of whether the absence of long-distance fronting is (statistically) surprising, given independently known properties of their corpus.¹⁵⁵

Second, most of the examples that Lema & Rivero give of three verb constructions appear to be biclausal constructions containing the main verb

¹⁵⁵ The relevant considerations would be things like (i) how many potential contexts for long-distance fronting are found in the corpus, and (ii) in another language which allows both local and long-distance participle fronting, what percentage of the time is the local or the non-local option chosen in an (even slightly) comparable corpus to the Old Spanish database?

complement of a future auxiliary and the infinitival complement of that main verb. These environments are not really comparable to the long-distance fronting environments that we have seen in other languages.¹⁵⁶ These caveats notwithstanding, let us assume that Lema & Rivero are correct and that Old Spanish only allows local fronting. The CCM account of locality in verb fronting therefore predicts that Old Spanish will not allow verb fronting in non-clitic contexts. Lema & Rivero give evidence which Rivero (1996) takes to disconfirm this prediction.

Old Spanish shows inversions of the finite auxiliary and the participial main verb in perfect tenses, as in (154). According to Lema & Rivero, these frontings are not clitic driven.¹⁵⁷

- (154) a. Confondido me has la vida
 confused me have.2s the life
 ‘You have confused my life.’
- b. Esto oído lo he
 this heard it have.1s
 ‘I have heard this.’ (Old Spanish)

The status of this construction is uncertain, as Lema & Rivero concede, but even apart from this, I do not think that it bears on the generalization claimed in (56). All of the languages which showed a “yes” in the “fronting with non-clitic auxiliary” column show participle fronting with clitic auxiliaries and participle fronting with non-clitic auxiliaries. The Old Spanish constructions in (152–153) on the one hand and in (154) on the other hand also differ in that infinitival verb forms are being fronted in (152–153) and participles are being fronted in (154), and therefore the availability of verb focusing for participles does not necessarily imply that the same option should be available for infinitivals.

Moreover, it is not entirely clear that examples like (154) do not involve VP-topicalization. Lema & Rivero give a number of reasons for distinguishing sentences like (154) from VP-topicalization, but they have much in common and their differences are for the most part easy to explain.¹⁵⁸

¹⁵⁶ It is not irrelevant that Lema & Rivero (1991) look at such constructions, because their principle concern in the paper is to contrast ‘long head movement’ in Old Spanish with VP-topicalization in Old Spanish, a construction which fronts a verb together with its arguments, and which can place at the front of the clause a verb which is underlyingly in an embedded clause. In this respect VP-topicalization is a more unbounded process than the verb fronting shown in (152).

¹⁵⁷ Lema & Rivero suggest tentatively that this fronting should be assimilated to Icelandic Stylistic Fronting, which also raises a participle across a perfect auxiliary (Maling 1980, Platzack 1987, Sigurðsson 1989). However, their main concern in their paper is to distinguish inversion in perfect tenses from VP topicalization and long head movement.

¹⁵⁸ L&R’s VP-topicalization constructions all involve modal verbs which embed infinitival complements. The lack of infinitival complements in the perfect tense construction in (154) may account for the fact that L&R find no instances of non-local fronting in this construction.

Therefore, contrary to the claim in Rivero 1996 I do not think that the varieties of verb fronting found in Old Spanish undermine the generalization in table (56).

A.3 Breton

One potentially serious counterexample to the generalization in table (56) remains: according to Rivero 1996 Breton shows non-local participle fronting like Bulgarian and Rumanian, but does not show participle fronting independent of clitic auxiliaries. However, this is not a counterexample to my generalization, because Breton lacks non-clitic auxiliaries which would make it possible to properly test the generalization. The relevant facts from Breton are the following. The examples in this section are all taken from Borsley, Rivero & Stephens 1996, although my characterization of the data differs from theirs in some respects.

Three kinds of periphrastic tense construction in Breton exhibit verb-fronting across a finite auxiliary. The three forms of the verb that can front are the past participle (155a), the passive participle (156a) and the ‘verb-noun’ that appears in present tenses in Breton (157a). The two participial forms front alone, and cannot be fronted together with a direct object (155b, 156b), but the verb-noun may front together with its direct object (157b).¹⁵⁹

- (155) a. Lennet en deus Yann al levr.
 read 3s.m has Yann the book
 ‘Yann has read the book.’
- b. * Lennet al levr en deus Yann.

L&R cite as another difference the fact that VP-topicalization allows fronting of verbs together with their arguments, whereas none of the perfect inversion examples show this. However, inspection of their examples shows another possible explanation: all instances of fronting verbs with their arguments involve fronting of infinitival verbs. All cases of participle fronting, whether in modal constructions or in the simple perfect tense, involve fronting of the verb alone.

L&R also claim that VP-topicalization and perfect inversion differ in that topicalization across negation is possible, but perfect inversion is not attested with negation. Here, unfortunately, it is again difficult to evaluate the status of this paradigmatic gap, given the absence of any information about how many tokens we might expect to find in the corpus. Also, the fact that the perfects are monoclausal structures and the VP-topicalizations involve biclausal modal structures introduces an additional variable which might underlie the difference in whether negation can be crossed by fronting. V(P)s may be blocked from fronting when their *own* clause is negated.

¹⁵⁹ The fact that the verb-noun may front either with or without its object appears to be a counterexample to Rivero’s observation that long head movement and VP-fronting constructions are normally in complementary distribution: if an auxiliary allows one, then it will not allow the other. However, as Borsley et al. (1996) point out, there is evidence that Breton has two different (homophonous) auxiliaries *ober* ‘do’, and that one of these allows V-fronting and the other VP-fronting (cf. Stephens 1982, Hewitt 1990).

Morphosyntax

- (156) a. Lennet e oa al levr gant Yann.
read prt was the book by Yann
'The book was read by Yann.'
- b. * Lennet gant Yann e oa al levr.
- (157) a. Lenn a ra Yann al levr.
read prt does Yann the book
'Yann reads the book.'
- b. * Lenn al levr Yann a ra. (Breton)

Meanwhile, the Breton progressive tense does not allow fronting of the progressive verb on its own (158a), but does allow fronting of an entire VP (158b).

- (158) a. * O lenn emañ Yann al levr.
prog read is Yann the book
'Yann is reading the book.'
- b. O lenn al levr emañ Yann. (Breton)

The most likely candidate for what drives verb-fronting in examples like (155–157) is the fact that finite verbs cannot appear sentence initially in Breton. A finite verb can be made non-initial by being preceded by a topicalized XP (159a) or by negation (159b). Topicalization and negation both block verb-fronting across the auxiliary (160), implying that verb-fronting occurs only when required in order to prevent the finite verb from appearing in sentence initial position.

- (159) a. Al levr en deus lennet Tom.
the book 3s.m has read Tom
'Tom has read the book.'
- b. N'en deus ket lennet Tom al levr.
neg 3s.m has read Tom the book
'Tom has not read the book.'
- (160) a. * Al levr lennet en deus Tom.
- b. * Lennet n'en deus ket Tom al levr. (Breton)

Thus, Breton verb-fronting across auxiliaries shows the properties of verb-fronting to support clitic auxiliaries that we have seen in many other languages.

In past perfect and past passive constructions, which contain two participles, fronting of either the auxiliary participle or the main verb participle is possible (161–162), as in Bulgarian. Therefore, Breton shows long-distance participle fronting.

Morphosyntax

- (161) a. Bet am eus kavet al levr.
had have.2s found the book
'I have found the book.'
- b. Kavet am eus bet al levr.
- (162) a. Bet eo lennet al levr gant Yann.
been is read the book by Yann
'The book has been read by Yann.'
- b. Lennet eo bet al levr gant Yann. (Breton)

The prediction that I made above regarding the distribution of long-distance participle fronting was that if a language shows long-distance participle fronting with clitic auxiliaries, then it will also allow long-distance participle fronting with non-clitic versions of those auxiliaries. Breton is not a counterexample to this claim, for the unfortunate reason that it does not have non-clitic versions of the relevant auxiliaries which would make the prediction testable. Therefore I do not find support for Rivero's (1996) use of Breton as a counterexample to Embick & Izvorski's (1995) similar claim about long-distance participle fronting in Bulgarian.¹⁶⁰

The Breton facts are compatible with the generalization about where long-distance participle-fronting is available, but notice that the absence of non-clitic auxiliaries in Breton makes a difference to the kind of information that a child learning Breton requires in order to learn that Breton allows long-distance participle fronting. In languages like Bulgarian and Serbo-Croatian, in which participle fronting occurs in the absence of clitic auxiliaries, the learner only needs to observe participle fronting with non-clitic auxiliaries in order to know that long-distance participle fronting is possible. The child learning Breton, on the other hand, can only acquire the possibility of sentences like (162b) by direct exposure.

¹⁶⁰ Breton does have one finite auxiliary that may appear sentence initially, the form of the copula that takes a progressive or PP complement.

- (i) Emañ Anna o lenn al levr.
is Anna prog read the book
'Anna is reading the book.'
- (ii) Emañ Yann war an hent.
is Yann on the road
'Yann is on the road.'

This auxiliary does not allow verbs to front across it, but does allow fronting of a VP, as we have seen in (158a–b) above. I do not have much to say about why there should be complementarity of verb-fronting and VP-fronting, but this complementarity is probably the cause of the impossibility of verb fronting with this auxiliary.

Morphosyntax

Chapter 5

The Parser and the Grammar

5.1 Putting the Pieces Together

In Chapter 1 I raised some issues about the general architecture of the language faculty, in particular the question of whether there are distinct syntactic systems for parsing and for the representation of grammatical knowledge. I suggested there that this distinction is not necessary, but delayed further discussion of the matter until after I had laid out in more detail what I think the parser and the grammar are like. In the light of the evidence presented in Chapters 2–4 I am now in a better position to justify the claim that there is no need to distinguish the grammar from the parser, and that there are a good many reasons for not distinguishing them. The goal of this chapter is to spell-out more carefully what this claim entails, to explain why it has generally been considered untenable, and to explain why I think the standard arguments for separating the parser and the grammar no longer apply.

Let me first briefly review the conclusions reached in Chapters 2–4 and how they are relevant to the question of the parser–grammar relation.

Chapter 2 considered a range of facts, mostly from English, which under standard syntactic assumptions make it look like the results of different constituency tests yield contradictory results. I argued that the contradictions go away, and that we gain some understanding of why the different tests give the results they do, if we assume that syntactic structures are assembled in a strictly left-to-right fashion. I assumed that structure building is constrained by the condition Merge Right, and subject to the structural economy condition Branch Right. These conclusions about how syntactic structures are built were reached based entirely on standard syntactic considerations involving grammaticality judgements, but it is hard to ignore the fact that Merge Right forces derivations to proceed in the same order that sentences are heard and produced. It is almost certain that the parser builds structures from left-to-right. Therefore, the conclusion that the grammar builds structures in the same order as the parser removes one of the most obvious differences between the parser and most standard models of grammar (which either involve bottom-to-top derivations or static characterizations of well-formed phrase markers). Of course, Merge Right and Branch Right do not themselves constitute a complete theory of syntax. The normal questions about what are the correct syntactic primitives, what are the correct underlying structures and what kinds of operations relate different

The Parser and the Grammar

structures apply here as in any other theory. All that is different is the order of the operations and certain properties of partial phrase markers.

In Chapter 3 the attention shifted to parsing, in particular to the topic of structural ambiguity resolution, which has been the focus of the majority of experimental work in sentence processing over the last 20 years. I argued that a wide range of structural biases in ambiguity resolution can be accounted for by a parsing strategy which favors the building of right-branching structures. This principle—Branch Right—is exactly the same principle that was claimed to choose between semantically equivalent derivations in Chapter 2. In Chapter 2 Branch Right chose among different structures that yielded the same meaning; in Chapter 3 Branch Right chose among structures which yielded different interpretations, but which were all equally compatible with what the perceiver knew about the intended meaning.

Historically, the existence of parsing ‘strategies’ which are required in order to resolve structural ambiguities has been taken as one kind of *prima facie* evidence that there is a parser, distinct from the grammar, because structural ambiguity resolution just isn’t the kind of thing that the grammar does. The conclusions of Chapter 3 undermine this argument, given that the complexity metric of the parser that resolves structural ambiguities turns out to be an independently motivated economy condition of the grammar.¹⁶¹

Chapter 4 returned to topics from the traditional domain of grammar, and extended the investigation of left-right derivations begun in Chapter 2, here focusing on issues at the interface of syntax with morphology and phonology. I showed that strict left-to-right derivations force a reordering of the feeding relationship between morphology and syntax that is assumed in much recent work. In the current model, morphophonological analysis of a word precedes its participation in syntax. I show that this has a number of positive consequences for topics such as clitic placement, head movement, and phrasal phonology. Besides extending the support for left-right derivations as a property of grammar, the relevance of this chapter to the parser–grammar issue is that it shows that morphophonological analysis of words precedes the syntactic operations that apply to their component parts. This ordering of processes is almost certainly also a property of the system responsible for natural language parsing.

In sum, Chapters 2–4 point out a number of striking parallels between the parser and the grammar. An obvious possibility that these parallels suggest is that the parser and the grammar are not just very similar, they are in fact the same thing. In other words, there is no parser (or put another way, there is no grammar: the choice does not really matter). In the remainder of this chapter I explore this possibility, what its attractions and potential difficulties are, and in particular why it has been doubted in the past. In §5.2 I describe the basic properties of such a model. §5.3 discusses some of the most well-known arguments against the model, including a reevaluation of the status of the *Derivational Theory of Complexity* (DTC). Discussion of the DTC is continued in detail in the appendix. §5.4 discusses some remaining questions that the

¹⁶¹ See Gorrell 1995 for a proposal which draws connections between notions of *economy of representation* in syntax and *minimal attachment* in parsing in a manner similar to what I am doing here.

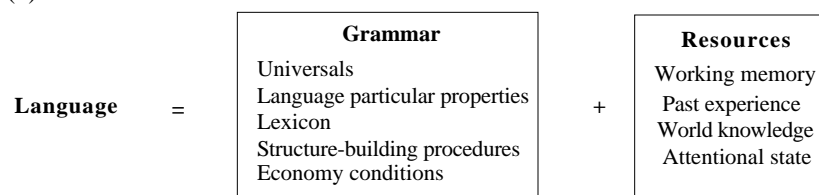
The Parser and the Grammar

model raises, and §5.5 briefly discusses the status of the competence–performance distinction, and what I assume about it.

5.2 The Parser is the Grammar

Consider the following general model of language knowledge and use. There is a grammar, which specifies mappings between sounds and meanings,¹⁶² and a finite set of resources, which corresponds to the memory and processing resources available to run the procedures that the grammar specifies. I suspect that it would be impossible to do without either of these components, and therefore these are probably the minimal ingredients of any model of human linguistic capacities. The model is sketched in (1). For want of a better name, I refer to this as the *PIG model* of language (Parser Is Grammar).

(1)



The PIG Model

Under this view, grammaticality is just parsability in the limit. A sentence is grammatical if it can be generated under conditions of unlimited resources.

(2) **GRAMMATICAL:** a sound/meaning pair is grammatical if the grammar can generate a correspondence for that pair, given potentially unbounded resources.

Meanwhile, a sentence is parsable with resources R if it can be generated using the grammar plus the resources R. The term ‘resources’ is left intentionally vague, and should range over at least memory, time, expectations and world knowledge, perceptual skills and motivation and attention.

(3) **PARSABLE:** a sound–meaning pair is parsable with resources R if the grammar can generate a correspondence for that pair using only the resources R.

Therefore, although we can draw a clear distinction between the *processes* of parsing and assigning grammaticality to sentences, and although there are clear differences between what is parsable and what is grammatical, it should be obvious that this in no way entails that there is a distinction between the mental

¹⁶² More accurately, the grammar specifies mappings between phonetic and semantic representations.

system that is used to assign grammaticality and the system that parses sentences. In the PIG model grammaticality is just a special case of parsability which makes the idealization of unbounded resources available to the grammar.¹⁶³

I should emphasize here that I am using the term *parsing* to refer to the construction of structural descriptions for sentences, and not to the many different cognitive processes that contribute to *comprehension*. Therefore, the well-known fact that there are many sentences that are easily comprehensible but not accepted as grammatical (e.g. *that-trace* violations) is fully consistent with my proposal. See §5.4 below for further discussion of this point.

The architecture of the PIG model is rather similar to the model of language that proposed by Miller and Chomsky in 1963, one of the earliest formulations of a model of language use in the generative grammar tradition.¹⁶⁴ In the terms of Berwick & Weinberg (1983) it adopts the strongest possible form of token-for-token transparency between the parser and the grammar. This means that it also provides the most straightforward account of how the parser and the grammar are related. However, despite the appeal of its simplicity this model has been considered impossible by most linguists and psycholinguists since soon after it was proposed. It will therefore be useful to review the reasons why the model has been thought to be unfeasible.

5.3 Objections to the PIG Model

A number of objections have been raised against identifying the parser with the grammar. Good summaries of the main objections can be found in Fodor, Bever & Garrett 1974, Levelt 1974 (vol. III) and Fillenbaum 1971. The main arguments presented in the 1970s against identifying the grammar with the parser were the following, and they have not changed much in the intervening 20 years:

- The grammars under consideration simply could not be directly implemented as sentence recognition devices (§5.3.1), given standard conceptions of the form of the grammar.
- Something very close to a direct implementation of a transformational grammar as a parser had been explored experimentally (a.k.a. the

¹⁶³ One aspect of parsing which I have had almost nothing to say about here is the parsing of the acoustic waveform into phonological segments and morphosyntactic units. I am therefore not in a good position to comment on whether the grammar–parser distinction is necessary at this level or not. However, I analysis-by-synthesis models of speech perception (a.k.a. *Motor Theory of Speech Perception*) have been investigated in considerable detail (cf. Liberman et al. 1967, Liberman & Mattingly 1985 for formulations, Remez 1994 for review).

¹⁶⁴ The identity of the parser and the grammar that I am advocating here is maybe a stronger position than what Miller & Chomsky assumed, because they assign a role for a preprocessor which “take[s] a sentence as input and give[s] us as output a relatively superficial analysis of it (perhaps the derived phrase marker ...)” (p. 480). However, the focus in their model is on just the role of the grammar and the finite workspace in language use, and not on any specialized preprocessing device.

The Parser and the Grammar

Derivational Theory of Complexity), and was widely considered to have been discredited (§5.3.2).

- Many phenomena in language processing were discovered which it was thought could not be explained if the parser and the grammar were identical (§5.3.3).

In this section I reconsider these three main arguments against the PIG model and their current status, taking into account the modified assumptions about the form of the grammar and of the parser presented in Chapters 2–4 and elsewhere.

It is unfortunate that the second argument, involving the DTC, has received the lion's share of the attention in the past, because this was never very crucial to the PIG model, and in any case the strength of the evidence against the DTC was exaggerated, as we shall see. I hope to show in this section that the first and third of these arguments are the most serious arguments against the identity of the parser and the grammar, and that these arguments do not apply to the theory suggested here.

5.3.1 The Grammar couldn't be a Parser

The most important objection to the claim that the grammar *is* the parser, although unfortunately not the one that has attracted most attention, is extremely simple and does not depend on any special assumptions about parsing, except for the uncontroversial fact that many sentences of natural language are recognizable in real time. The argument is that standard models of grammar could not be used directly as sentence recognition devices that could capture even this simple fact.

Let us assume for the sake of discussion the form of the grammar proposed in Chomsky's *Aspects of the Theory of Syntax* (Chomsky 1965), in which the grammar consists of a set of phrase structure rules which generate deep structures, and a set of transformational rules which map deep structures onto surface structures. I consider the status of other grammatical theories at the end of this section. Working on the basis of an *Aspects* style theory, there are two ways in which we might try to use a transformational grammar directly as a sentence recognition device.

5.3.1.1 Analysis-by-synthesis

The first possible implementation of the grammar is what is known as an *analysis-by-synthesis* recognition device. This kind of device recognizes input sentences by trying to generate a sentence to match the input, using the rules of the grammar. If the grammar manages to generate a matching form then it accepts the input, otherwise it rejects the input. In other words, we parse sentences by attempting to generate them for ourselves.

There is, however, one enormous problem that such models face. The analysis-by-synthesis model as stated specifies nothing about which sentences the grammar should generate in its search for a match to the input. One possibility is random search through the infinite space of sentences that the grammar can generate, in the hope that one will happen to match the input. But

random search is clearly not a particularly plausible model of how people recognize and understand sentences.

There are, of course, better ways than random search of choosing what the grammar should generate in its search for a match. But even these are not particularly constraining. Taking into consideration which words the input sentence contains ought to help, but it is not obvious that they do help, given that the words are the output of rather than the input to the grammatical derivation. Knowing the order in which the words occur is not particularly useful if the sentence is a transformed sentence, and the grammar therefore has to first generate an underlying phrase marker in which the words are ordered differently.

Whether or not these strategies succeed in restricting the search space of an analysis-by-synthesis sentence recognizer, the fact that they are necessary raises an important point. If the strategies that are used to narrow the search space are not a part of the grammar, then this amounts to adding a special preprocessing device to the grammar in order to make analysis-by-synthesis possible. If such moves prove to be necessary—which they clearly are if the grammar is an *Aspects* style transformational grammar—then we are forced to give up the claim that the grammar and the parser are one and the same thing. The grammar is used fairly directly in parsing, but only in conjunction with additional parsing strategies.

Furthermore, not only does an analysis-by-synthesis model of sentence recognition force the addition of some kind of preprocessor to the transformational grammar; once the preprocessor's search mechanism is spelled-out more explicitly the preprocessor quickly takes over a good deal of the work of parsing the surface structure of the input sentence. In other words, a preprocessor which would be at all effective in saving an analysis-by-synthesis model from the random search problem is anything but a trivial addition to such a model of sentence recognition.

The second problem for the analysis-by-synthesis grammar as parser view is that a plausible model of parsing must be able to parse sentences incrementally from left-to-right. If the grammar does not provide any characterization of incremental sentence recognition, but instead just recognizes completed sentences, then it fails this basic criterion of adequacy.

For reasons such as these, direct implementations of the grammar as an analysis-by-synthesis parser, which were briefly explored in the early 1960s (e.g. Matthews 1962) have not in general been considered a serious possibility since around the mid-1960s.¹⁶⁵

5.3.1.2 Analysis-by-analysis

The second possible implementation of the grammar as a sentence recognizer is what are known as *analysis-by-analysis* models. In this kind of model sentences are recognized by taking the input sentence—presumably the output of a well-formed grammatical derivation—and attempting to run the

¹⁶⁵ This is not to say that the combination of a grammar and a parser could not make a highly effective analysis-by-synthesis parsing system. All that this argument is intended to show is that the grammar on its own is insufficient.

The Parser and the Grammar

grammatical derivation 'in reverse' to derive an appropriate underlying structure. For example, if the grammar is conceived as starting with the symbol *S*, and then applying a sequence of rewrite rules and transformations to arrive at the output string of words, an analysis-by-analysis recognizer based on this grammar would start with the string of words and attempt to run the transformations and rewrite rules in reverse order to arrive back at the symbol *S*.¹⁶⁶

While this model appears feasible enough in the abstract, it runs into essentially the same kinds of difficulties that the analysis-by-synthesis model faces when put into practice. As Fodor et al. (1974) show, in order for the analysis-by-analysis model to be a plausible model of sentence recognition there must be some relatively efficient way of determining which transformations and which rewrite rules to 'un-apply' in order to arrive back at the start state. There are two main obstacles to this.

First, consider an untransformed and completely unambiguous sentence like (4a), which has a complex NP as the object. Let us assume that it is generated by the abbreviated set of phrase structure rules in (4b).¹⁶⁷

- (4) a. John read the news that his company was about to lay off 5000 workers.
- b.
- | | | |
|----|-----|----------|
| A. | S | NP VP |
| B. | VP | V NP |
| C. | NP | Det N' |
| D. | N' | N that S |
| E. | N | John |
| F. | V | read |
| G. | Det | the |
| H. | N | news |
- etc.

Rules (A) through (D) in (4b) correspond to the order in which rules would be applied in a derivation from the starting symbol *S* to the output string in (1a). Rule B cannot apply until rule A has applied, rule C cannot apply until after rule B has applied, and so on. An analysis-by-analysis recognition device would therefore have to run through this derivation in reverse, recognizing first the output of rule D, then the output of rule C, then rule B and finally rule A. Notice that this means that no structure can be built until the very end of the sentence has been reached, because rule D cannot be recognized until the end of the sentential complement of the noun *news* has been completed, and no other rules can be recognized until after rule D has been recognized. Clearly, though, speakers of English are able to understand sentences like (4a) incrementally, a

¹⁶⁶ If the grammar is viewed as not involving any kind of derivation, then the possibility of 'running the grammatical derivation backwards' is obviously not available.

¹⁶⁷ In (4b) I omit the rules that are required in order to generate the sentential complement of *news*, because these are not important to the discussion.

The Parser and the Grammar

basic fact that is surprising under the analysis-by-analysis implementation of a traditional transformational grammar.

Of course, this problem can be avoided by allowing rules to be hypothesized in advance of complete confirmation by input words (for example, the string NP V may be taken as an indication that the rule S → NP VP has applied), but this is no longer a pure analysis-by-analysis implementation of the grammar: it is a parser that is based upon that grammar.¹⁶⁸

The second obstacle facing an analysis-by-analysis approach to sentence recognition is that it is not even clear what is supposed to be recognized in the parsing of sentences that have undergone transformations. Sentence (4a) could be recognized easily by picking out the expansions of the various rewrite rules that produced it, but in a transformed sentence it is not so straightforward to recognize what transformations have applied so that they may be 'undone'. The grammar specifies exhaustively the list of symbols that are the output of any rewrite rule, and therefore it is relatively feasible for a parsing device to search for the output of rewrite rules. In the case of transformational rules, on the other hand, the outputs of these rules are not necessarily exhaustively specified lists of symbols that can be recovered from a surface string. If a transformational rule contains as part of its context a variable symbol which ranges over a variety of different environments, then it will be hard to match this variable in the output. For example, how is the string 'wh-NP X Y' to be identified in a sentence to which the *wh*-movement rule in (5a) has applied? Worse still, if a rule deletes an element in its input, then it will be impossible to find it in the output, as will be the case if an ellipsis rule like (5b) applies.

(5) a. *Wh-movement*

	X	wh-NP	Y
	1	2	3
2	1	0	3

b. *VP-ellipsis*

X	VP ₁	Y	VP ₂	Z
1	2	3	4	5
1	2	3	0	5

Condition: VP₁ = VP₂

Again, just as with the analysis-by-synthesis model, it is not difficult to find ways of getting around these problems. For example, in the case of the *wh*-movement rule (5a) the presence of a *wh*-phrase could initiate a search for the gap position included in the output of the *wh*-movement rule, and the literature contains a number of proposals for how such a search might proceed. In the case of the VP-ellipsis rule (5b) the content of the deleted VP can easily be

¹⁶⁸ This does not mean that a grammar could not in principle incorporate procedures for generating nodes for as yet unconfirmed input. The point is only that standard phrase structure grammars do not include such procedures.

The Parser and the Grammar

determined from the form of its antecedent, and the ellipsis site may be identified by the presence of a string such as *did so too*. But here as before the crucial point is that these heuristics are not part of the grammar proper, and therefore amount to the introduction of a separate system for sentence recognition and hence the giving up of the PIG model.

This is not to say that a parser which is constructed in such a way that it ‘undoes’ transformations and rewrite rules in an efficient way is not a viable approach to sentence parsing. The point is just that the grammar alone cannot do this, given standard views of what the grammar looks like.

Given the failure of both the analysis-by-synthesis and the analysis-by-analysis implementations of the grammar as a parser, the first objection to the PIG model of language is simply that the grammars under consideration in the late 1960s and early 1970s could not be used as sentence recognizers without either failing to capture the most basic facts about human language comprehension or adding a significant amount of machinery specifically for parsing.

Moreover, the force of these arguments is so general that they also apply to most other theories of grammar that have been proposed over the last 20 years. The arguments against direct use of the grammar for analysis-by-synthesis apply to any grammatical model which requires something other than incremental left-to-right structure building. And as for the arguments against the analysis-by-analysis implementation of grammar that were based on the difficulty of identifying the output of movement operations: these apply equally well to those non-transformational models of grammar in which long-distance dependencies are represented in a manner similar to the output of (5a).

Current Status. The objection that the PIG model could not be used as a parser, given what is known about the form of the grammar, seems to be the most compelling argument against the model, and it is rather more compelling than certain other criticisms of the model which have received a good deal more attention. Nevertheless, I think that this criticism does not apply to a left-to-right grammar of the kind advocated in Chapters 2 and 4.

Recall that the main problem with an analysis-by-synthesis sentence recognition device is that if it consists of only the grammar then it has difficulty in using the input to incrementally generate a matching sentence. In a grammar in which sentences are built strictly from left-to-right, however, it is quite straightforward to constrain the range of possibilities that the grammar searches for a match to the input. The words of the input sentence are added to the grammatical derivation one at a time and in the order in which they occur. This immediately provides an extremely tight constraint on the derivations the grammar has to consider, and makes an analysis-by-synthesis derivation almost deterministic. Any derivation which uses words that are different from those in the input or which occur in a different order is automatically excluded. The only exceptions to this deterministic assembly of words should therefore occur in exactly the situations where parsing is not deterministic for humans, i.e. structural ambiguities.

Therefore, what had been a fatal problem for the analysis-by-synthesis parser based on traditional transformational grammars is straightforwardly handled under the left-right approach.¹⁶⁹

5.3.2 The Derivational Theory of Complexity

The first argument against the PIG model showed that the grammar could not be a parser, under a very general set of assumptions about the form of the grammar. This ought to be enough to dismiss the purest form of the PIG model, but it leaves open the question of how closely related the grammar and the parser are. This section revisits a classic set of arguments that the operations of the grammar and parser are distantly related to one another at best.

The passage from Miller & Chomsky's 1963 paper that appears to have attracted most attention is the following one, in which they point out one way in which their model might receive direct confirmation:

'The psychological plausibility of a transformational model of the language user would be strengthened, of course, if it could be shown that our performance on tasks requiring an appreciation of the structure of transformed sentences is some function of the nature, number and complexity of the grammatical transformations involved.' (Miller & Chomsky 1963: p. 481)

The idea was that if the transformational operations of the grammar correspond to specific mental operations in the processing of sentences, then it might be possible to see this in reaction time measures. Therefore, it should be possible to construct sentence minimal pairs, such that the only difference between the two sentences is that one is generated using a given transformation and the other is not. Processing the transformed sentence should therefore be more complex than processing the untransformed sentence, and this difference should be measurable in terms of processing time (or processing accuracy, or memory space, etc.). One specific interpretation of this later became known as the *Derivational Theory of Complexity* (DTC).

Issues relevant to the DTC were extensively tested for a brief period in the 1960s, and then seem to have been largely ignored in the experimental psycholinguistics literature from around 1970 onwards, although the issues have occasionally resurfaced in discussions of linguistic theory and of computational issues (Bresnan 1978, Berwick & Weinberg 1983, 1984, Pritchett & Whitman 1993, 1995). It will be worthwhile to discuss this topic in more detail, for the reason that the 'failure of the DTC' is often regarded as an early chapter in the history of psycholinguistics, which showed that the parser and the grammar

¹⁶⁹ Given that the derivations of the left-right grammar that I am proposing here closely resemble the derivations of an analysis-by-analysis parser for a traditional transformational grammar, the terms 'analysis-by-synthesis' and 'analysis-by-analysis' run the risk of becoming confusing. By analysis-by-synthesis I mean 'deriving sentences in exactly the manner that the grammar specifies', and I take analysis-by-analysis to refer to the grammar run 'backwards'.

Therefore, if the grammar specifies a left-right mapping from surface strings to underlying representations, it makes little sense to pursue the possibility of an analysis-by-analysis implementation of this grammar.

could not be identical. In this section and the appendix I aim to show that the failure of the DTC is somewhat of a fiction; mostly, the DTC just stopped being an issue that psycholinguists were interested in pursuing.¹⁷⁰

5.3.2.1 The Received View of the DTC

The ‘received’ view of the DTC corresponds closely to the view presented in Fodor, Bever & Garrett 1974 (see also Levelt 1974, vol. III). The stylized version of the main conclusions corresponds roughly to the following claims:

- A large body of experimental work led to the disconfirmation of the Derivational Theory of Complexity.
- These experiments found evidence for the ‘psychological reality’ of the representations but not for the transformational operations proposed by generative grammars.

This picture oversimplifies the conclusions of Fodor et al, whose original conclusion was rather more guarded.

‘Perhaps the fairest summary of the evidence for and against the DTC is that it fails to yield a decisive resolution. This is not solely because of shortcomings in the experiments. Any test of DTC must be predicated on *some* assumptions about the grammar, and it is always possible that such assumptions might turn out to be false.’ (Fodor, Bever & Garrett 1974, p. 328)

Fodor et al.’s conclusions were based on careful analysis of a number of experiments in the context of a very specific set of assumptions about the form of a transformational grammar. The conclusions do not necessarily apply to a different set of assumptions about the grammar, as has since been pointed out by a number of people.

I should also emphasize that I think the importance of the DTC to the issue of parser–grammar identity has been overemphasized. The really damaging argument against the PIG model is the argument already presented in §5.3.1, which did not depend on any experimental predictions, and held across a wide range of different assumptions about the grammar. The DTC, on the other hand, was never more than one potential correlate of Miller & Chomsky’s proposal. It pointed out a hypothetical set of experimental results which could be explained if their model was correct, not a set of predictions upon which their model should stand or fall. Moreover, any test of the DTC was necessarily predicated on a very specific set of assumptions about the form of the grammar and the parser.

¹⁷⁰ Useful reviews of DTC-related experiments can be found in Fodor, Bever & Garrett 1974, pp. 226–268, 320–328, Levelt 1974 (vol. III), pp. 92–103 and Fillenbaum 1971. Some DTC-relevant findings are discussed in detail in Berwick & Weinberg 1983, 1984. Other views of the status and relevance of the DTC experiments can be found in Garnham 1983, Bever 1988 and Wanner 1988. A recent reevaluation of the DTC experiments can be found in Pritchett & Whitman 1993. Part of the appendix draws heavily on the discussion in that paper, which I repeat here for the reason that the paper is not readily available.

The Parser and the Grammar

A number of experiments produced results which could be explained by the DTC, utilizing a variety of experimental measures, and then a great many studies produced results into which the DTC did not seem to provide any insight. The counterarguments consisted of (i) demonstrations of possibly confounding factors in some (but not all) of the supporting results, and (ii) a series of sentence pairs whose relative 'perceptual complexity' did not seem to correspond to the predictions of transformational analyses of the time.

The experimental arguments against the DTC were never particularly overwhelming, and I think it is fair to say that the demise of studies relevant to this question was not because the DTC was taken to have been 'disconfirmed', but rather because the most fruitful research questions in psycholinguistics at the time lay elsewhere, in issues of relatively grammar-independent processing strategies.

5.3.2.2 Results Supporting DTC

A large number of different studies investigated the relative perceptual complexity of transformations like passivization, negation or *yes/no* question formation. In the theory of Chomsky 1957 these transformations were assumed to be operations that applied optionally to *kernel* sentences, which were active affirmative sentences. A typical study used as its target stimuli various combinations of these transformations to the same kernel sentences. An example is the following set of eight sentences from Mehler 1963.

- | | | | |
|-----|----|--|--------------|
| (6) | a. | the secretary has typed the paper | K (kernel) |
| | b. | the paper has been typed by the secretary | P (passive) |
| | c. | the secretary has not typed the paper | N (negative) |
| | d. | has the secretary typed the paper | Q (question) |
| | e. | hasn't the secretary typed the paper? | NQ |
| | f. | the paper hasn't been typed by the secretary | PN |
| | g. | has the paper been typed by the secretary | PQ |
| | h. | hasn't the paper been typed by the secretary | PNQ |

Miller, McKean and Slobin (described in Miller 1962 and in Miller & McKean 1964) presented subjects with a page containing two lists of sentences. The subjects were asked to match sentences from the first list with a corresponding sentence in the second list that was related by a prespecified transformation (P, N, or PN). Miller & McKean (1964) performed a variant of the same experiment in which sentences were tachistoscopically presented one at a time, and after each sentence was presented a search list was presented. In the earlier paper and pencil study the time to search the list was recorded; in the Miller & McKean study, subjects were instructed to perform the transformation mentally before triggering the display of the search list; only the time to read the sentence and perform the required transformation was recorded, not the search time. Baseline times in variants of the task in which no transformations were required were subtracted out from the results. The residual time was assumed to correspond to the time required to perform the transformation.

Both experiments showed results which conform to the predictions of the DTC. The two conditions which required two transformations (e.g. PN to K, P

The Parser and the Grammar

to N) required longer than the four conditions which required just one transformation (e.g. PN to P, N to K), and in fact the time required to perform the double transformations was predicted rather well by the time required to perform the two transformations individually. See below for more detailed discussion of this study and some objections that were raised about whether the results really supported the DTC.

McMahon (1963) asked subjects to confirm or reject as quickly as possible the truth of sentences of varying degrees of transformational complexity. The sentences used in this experiment were arithmetical statements like (7).

- | | | | | |
|-----|----|-----|-----------------------------------|------------|
| (7) | a. | i. | seven precedes thirteen | K (true) |
| | | ii. | thirteen precedes seven | K (false) |
| | b. | i. | thirteen is preceded by seven | P (true) |
| | | ii. | seven is preceded by thirteen | P (false) |
| | c. | i. | thirteen does not precede seven | N (true) |
| | | ii. | seven does not precede thirteen | N (false) |
| | d. | i. | seven is not preceded by thirteen | PN (true) |
| | | ii. | thirteen is not preceded by seven | PN (false) |

As in the Miller & McKean experiments, the length of reaction times were predicted by transformational complexity, including the fact that times on conditions requiring both P and N transformations could be fairly accurately predicted by times on conditions in which either P or N was required on its own.

A variant on this verification task was used in another study by McMahon (1963) and in two studies by Gough (1965, 1966). In these studies subjects were shown a picture and then presented with a sentence which they had to judge whether it was true of the picture. Here again, the supposed derivational complexity of the sentences correlated rather well with the time required to perform the task.¹⁷¹

Finally, Savin & Perchonock (1965) used a task in which subjects were asked to memorize both a sentence and a set of unrelated words.¹⁷² The transformational complexity of the sentence was varied in different conditions. The idea of this study was to test whether increasing the transformational complexity of a sentence leads to corresponding decreases in the ability to recall word lists, as might be expected if derivational operations place a burden on short term storage. Savin & Perchonock's results seemed to indicate that more complex derivations do indeed use up more memory space, since there was a decrease in performance on the memory task as the sentences became more

¹⁷¹ There was a residual effect of truth vs. falsity of the target sentence on reaction times in these studies, meaning that something more needs to be said about the processes involved in truth value judgement tasks, in order to explain why denial takes longer than affirmation. However, this does not detract from the point that the transformational complexity of sentence types accounts for a good deal of the remaining variation in response times.

¹⁷² This study used the same transformations (P, N, Q) used in the other studies described, plus *wh*-questions (e.g., *who has typed the paper?*) and emphatic forms (e.g., *the secretary DID type the paper*).

complex. (See below for further discussion of this study and some criticisms that were raised against it.)

In sum, then, these various experiments provided promising initial support for the DTC.

5.3.2.3 Some Concerns about the DTC Experiments

The results that claimed to support the DTC were criticized for a number of reasons, some of which remain damaging for the DTC, and others of which do not. In general, the force of the various criticisms has been exaggerated. I give here a brief summary of the main objections to the early DTC experiments, but defer detailed discussion to the appendix to this chapter.

Some of the experimental findings, such as McMahon's, were never seriously challenged.

Fodor et al. (1974) give a detailed criticism of Miller & McKean's 1964 study, arguing that its predictions were incorrect interpretations of the DTC, and that a more accurate set of predictions was not supported by their results, thereby making their findings irrelevant to the DTC. In the appendix I show that Fodor et al.'s criticism of the Miller & McKean study is quite unjustified. When we revise Miller & McKean's predictions in order to answer Fodor et al.'s interpretation of the DTC, we actually obtain a better model of the results than Miller & McKean's own predictions!

Slobin's (1966) criticism of earlier studies of the passive transformation has received a good deal of attention in later reviews of the DTC experiments (cf. Bresnan 1978, Berwick & Weinberg 1983, 1984). Slobin argued that once effects of reversibility in passives are controlled for, the slowdown due to passivization reported in other studies disappears. However, as I show in the appendix, Slobin's results are an artifact of the task used.

Savin & Perchonock's (1965) claim that transformational complexity has a measurable effect on memory capacity was both replicated and challenged in a number of studies. In the appendix I show that the most serious challenge to their finding was to show that transformational complexity has a measurable time cost but no independent memory cost.

Most of the early investigations of transformational complexity focused on a small number of transformations—passive, negation, and yes/no question formation. Later in the 1960s a number of studies investigated a broader range of assumed transformations, and generally found a poor correlation between transformational and perceptual complexity. The appendix surveys these studies (building on earlier reviews by Berwick & Weinberg 1983 and Pritchett & Whitman 1993), and shows that these results are quite consistent with more recent and well-motivated syntactic analyses of the constructions in question.

It is granted that the collection of experiments which were most consistent with the DTC never amounted to overwhelming evidence for that theory, as opposed to some other interpretation of the data. However, the received view that initially corroborating experimental evidence for the DTC was subsequently discredited is far from an accurate summary of what happened. It is true that some of the experiments required reinterpretation, but this never amounted to a serious challenge to the DTC, and sometimes even lent stronger support to the DTC than the original authors had claimed.

5.3.2.4 Transformations that Reduce Perceptual Complexity.

There is a further set of cases in which transformational complexity is inversely correlated with perceptual complexity, even under contemporary analyses. These examples serve to illustrate an important point about the scope of the DTC or theories like it.

Consider the two sentence pairs in (8-9). (8a) is a multiply center embedded sentence, and (8b) replaces two of the active clauses of (8a) with passive clauses (cf. Fodor et al. 1974, p. 327). (9a-b) show a dative construction with a heavy direct object both with and without Heavy NP Shift (cf. Pritchett & Whitman 1993).

- (8) a. The first shot the tired soldier the mosquito bit fired missed.
b. The first shot fired by the tired soldier bitten by the mosquito missed.
- (9) a. I gave a complete set of the annotated works of H.H. Munro to Felix.
b. I gave to Felix a complete set of the annotated works of H.H. Munro.

Both of the (b) examples are widely considered to be transformationally more complex than the corresponding (a) examples, now as in the 1960s.¹⁷³ But it is also fairly clear that the (b) examples are easier to understand than the (a) examples. These sentences clearly do not show perceptual complexity differences in the direction predicted by the DTC.

It is fairly easy to see why the sentences in (8b) and (9b) are easier to process than the (a) versions. Multiply center embedded sentences are well-known to be very difficult to understand, and the sentences in (8b) and (9b) avoid the center embedding that is found in the (a) sentences and replace it with a more right-branching structure. Therefore, if the use of more right-branching configurations and the avoidance of center-embedding is an aid to comprehension independent of transformational complexity, then the facts in (8-9) are quite easily reconciled with the DTC. Passivization or HNPS may increase the transformational complexity of a sentence, but this may be more than cancelled out by the benefits of avoiding center-embedded structures.

Of course, the authors who have used examples like (8-9) as evidence against the DTC were well aware of this fact. They were quite justified in using examples like these to show that transformational complexity is unlikely to be the *sole* determinant of complexity in sentence processing, but I can see no justification for using them as an argument against the DTC, or at least against that version of the DTC which is a reasonable inference from the assumption that sentence recognition recapitulates the grammatical derivation of a sentence.

¹⁷³ There are analyses of Heavy NP Shift in which the heavy NP is not moved rightwards but instead the light NP is moved leftwards (Larson 1989). But even in such accounts, the HNPS sentence in (9b) involves moving something that does not undergo movement in (9a).

The Parser and the Grammar

Consider the following three possible interpretations of the DTC, which should not be confused with one another.

- (10) a. *Derivational operations add to perceptual complexity*
Parsing involves recapitulating the grammatical derivation of a sentence; additional derivational operations are a source of increased perceptual complexity.
- b. *More derivational steps means greater perceptual complexity*
Sentences with more derivational steps are perceptually more complex than sentences with fewer derivational steps.
- c. *Perceptual complexity is determined by derivational complexity*
The perceptual complexity of a sentence is entirely a function of its derivational history.

Alternative (10a) seems to be a reasonable prediction to derive from the claim that recognizing a sentence involves running through its grammatical derivation. None of the studies that we have reviewed in this section have cast doubt upon this claim. Examples like (8–9) are, however, inconsistent with stronger claims like (10b) and (10c). But neither of these claims follows from the assumption that the parser and the grammar are identical, nor do they follow from the passage from Miller & Chomsky quoted above. If the parser is identical to the grammar this does not mean that the only possible source of difficulty in sentence processing is the number of transformational operations in the sentence's grammatical derivation.

5.3.2.5 Other reevaluations of the DTC experiments

Although there was initially little interest in defending the DTC against the criticisms levelled at it in the 1960s and early 1970s, there have been periodic attempts to reopen discussion of the DTC, and the implications of the experimental results. A variety of different responses to the DTC results have been taken, involving modifications of the linguistic grammar, of the parser, or of both. In all of these cases, though, it has been taken for granted that there was *something* wrong with the DTC as it was advanced in the 1960s. I think that even this conclusion is too strong.

Bresnan (1978) argues that the 'rather pessimistic conclusions that have [...] been drawn in the psycholinguistic literature' (p.1) support the idea that the transformational component of a transformational grammar should be restricted, as it is in her Extended Lexical Theory, a precursor to Lexical Functional Grammar.¹⁷⁴ In other words, something like the DTC should be resurrected, but with respect to a modified grammar.

¹⁷⁴ Note, however, that Bresnan does not use Slobin's (1966) claims about the perceptual complexity of passives to support her non-transformational analysis of the passive, although Bresnan's paper has sometimes been represented as making this argument (cf. Berwick & Weinberg 1983, 1984, Newmeyer 1986). Bresnan was aware of results of Forster and colleagues which called into question Slobin's

The Parser and the Grammar

Berwick & Weinberg (1983, 1984) respond to the DTC results in a rather different way. They suggest that the results did not call into question the validity of a transformational grammar, rather they undermined certain assumptions that critics of the DTC had made about how ‘perceptual complexity’ should be realized. They show that in a non-serial model of parsing increased complexity can easily be realized without increase in time demands.

Pritchett & Whitman (1993, 1994) argue that the DTC can be resurrected by making different assumptions about both the parser and the grammar. They are not in fact interested in proposing a parsing model which recapitulates the operations of the grammar (they assume the model of parsing proposed in Pritchett 1992). They are just interested in the question of whether the complexity of a sentence as defined by the grammar contributes to the sentence’s perceptual complexity. They propose replacing the DTC with a *Representational Theory of Complexity* (RTC), which predicts greater perceptual complexity for LF representations that contain greater numbers of chain links. Given that chain links of LF representations effectively encode the derivational history of a sentence from D-structure via S-structure to LF in the grammatical model that Pritchett & Whitman assume, they are able to predict a correlation between grammatical and perceptual complexity, without being committed to the view that the parser recapitulates operations of the grammatical derivation.

All of these authors assume that the DTC experiments showed that something was wrong with the conception of the parser/grammar relation suggested in Miller 1962 and Miller & Chomsky 1963, involving either the form of the grammar or the form of the parser, or the way in which they interact. However, I do not think that even these relatively modest responses to the DTC experiments are forced by the experimental results from the 1960s. The main problem with the early proposals about the parser–grammar relation is not that they were shown to be wrong, only that they were never made particularly precise.

5.3.2.6 Conclusion

When we look back at the various DTC-related studies we find something rather suspicious. For a theory that was supposedly so attractive initially, and a demise that was supposedly so conclusive, it is rather surprising that we find no attempts to defend the DTC against the criticisms levelled against it. This is even more surprising given how easy it is to defend the DTC against many of these criticisms, as we have seen here.

What *really* seems to have happened is that psycholinguists found other topics which they found more interesting to work on, and so interest in DTC-related issues simply died out. Problems with some of the earlier experiments were correctly pointed out, but in general there does not seem to have been much interest in correcting for these problems. Syntacticians, meanwhile, do not enter the picture.

In sum, I have no misgivings about espousing an account of the parser–grammar relation which entails assumptions closely reminiscent of the the DTC.

conclusions about passives, using more sensitive techniques (Forster & Olbrei 1973, Forster 1976).

In any case, as I have emphasized a number of times already, the DTC was never the best argument against the PIG model of language. Better and simpler arguments were widely available. These involved the demonstration that something more than the grammar was contributing to parsing performance. The next section considers this kind of objection to the PIG model.

5.3.3 There are Non-Grammatical Parsing Strategies

The third argument against the PIG model is more telling than the previous one. It is also rather simple. The argument is that if we find evidence for the operations of parsing strategies¹⁷⁵ that cannot be attributed to either the grammar the finite resources, then we have straightforward evidence for the existence of a parser which is distinct from both the grammar and the resources. An advantage of this argument is that it does not depend in any way on how directly or indirectly the grammar is used in parsing. This means, for example, that the argument that a left-to-right grammar is highly *compatible* with implementation as a parser is quite irrelevant to this objection to the PIG model. Even if the form of the grammar does not *necessitate* an independent parser, it could still be the case that an independent parser exists.

Therefore, any evidence about the operations of parsing which cannot be reduced to a component of the grammar (and cannot be reduced to properties of the workspace) could potentially seal the argument against the identity of the grammar and the parser.

Some of the best candidates for non-grammatical parsing principles are the various ambiguity resolution strategies that have been proposed over the years. On the assumption that ambiguity resolution is not something that the grammar ever has to be concerned with, special principles for ambiguity resolution are unlikely to be reducible to principles of grammar.

For example, a strategy like *Minimal Attachment* (Frazier & Fodor 1978) which attempts to keep the number of phrase structure nodes to a minimum, does not have any correlate in most theories of grammar. Similarly, grammars generally do not make reference to strategies which select as the optimal parse of a string the structure in which the fewest number of predicate-argument relations are left open (Pritchett 1988, 1992; Gibson 1991), nor to strategies which analyze phrases as arguments wherever possible (Ford, Bresnan & Kaplan 1982, Abney 1987, 1989, Schütze & Gibson 1996). Likewise, the strategy favoring local over non-local attachment which most theories of parsing incorporate in some form or another does not seem like a grammatical principle, under standard assumptions about grammar. If none of these are derivable from grammar, then they are presumably properties of an independent parsing system.

Current Status. The argument just presented for the existence of an independent parser depends on the premise that ambiguity resolution involves rejecting one alternative because another alternative is *better*, and that the grammaticality of a sentence never depends on such considerations. This premise was generally valid until recent years, in which a large number of

¹⁷⁵ The use of the term 'strategy' is not intended to imply that the strategy is consciously accessed.

The Parser and the Grammar

grammatical proposals have appeared which make precisely the assumption that the grammar may rule out some structures just because there is a better alternative available. Comparison mechanisms do some of the work in determining grammaticality in the framework of Chomsky's *Minimalist Program* (e.g. Chomsky 1991, 1993, 1995), and this is taken further in *Optimality Theory* (Prince & Smolensky 1993, Grimshaw 1995), in which any given form is only ever ungrammatical because there is a better alternative.

This change in the way that many people look at grammatical phenomena means that it is now relatively easy to find grammatical proposals which are similar to if not identical to the kinds of parsing strategies listed above.

For example, Frazier's Minimal Attachment principle for ambiguity resolution finds a potential grammatical parallel in principles which minimize the number of syntactic projections in a sentence (Grimshaw 1995, Speas 1990).¹⁷⁶

There is still no direct grammatical correlate that I am aware of for parsing principles which minimize the number of open predicate-argument relations,¹⁷⁷ but more distant parallels can be found. This kind of parsing principle can be viewed as a special instance of an imperative to satisfy grammatical constraints (such as the theta-criterion) as soon as possible. A grammatical analog to this is found in the *Earliness Principle* proposed by Pesetsky (1989), which requires that grammatical requirements be satisfied as early as possible in a derivation.

What is especially relevant for my purposes here is the fact that the parsing principle Branch Right, which I argued in Chapter 3 to have wide reaching effects in ambiguity resolution, is exactly the same principle that was appealed to in Chapter 2 to account for the predominance of right-branching structures. In both parsing and grammaticality assignment Branch Right selects the most right-branching structure that is consistent with what the speaker knows about the intended meaning, and in both parsing and grammaticality assignment deviations from the most right-branching structure are licensed if another interpretation is thereby created, which is more consistent with what is known about the intended meaning.

Therefore, ambiguity resolution strategies no longer provide the straightforward argument against the identity of the parser and the grammar that they once did. Of course, the fact that both the grammar and the parser include economy principles does not guarantee that they include the *same* economy principles. But the mere fact that theories of parsing make reference to economy principles is no longer a sufficient argument for the existence of an independent parser.

Quite independent of the status of structural economy conditions in parsing and grammar, non-structural factors in parsing, such as the use of contextual

¹⁷⁶ See Gorrell 1995 for discussion of this possibility.

¹⁷⁷ Such principles are closely related to grammatical principles like the *theta criterion* (Chomsky 1981) or the *principle of completeness* in LFG (Bresnan 1982), but the connection is indirect, since the grammatical principles are stated over complete phrase markers, and do not tolerate the existence of any incomplete predicate-argument relations, whereas the parsing principles choose between different partial phrase markers containing more or less incomplete predicate-argument relations.

knowledge and probabilistic information, might be seen as properties of a parser that really cannot be reduced to grammar. However, I see no need to attribute these aspects of comprehension to a special-purpose parsing system, so they do not bear on the issue of whether the parser and the grammar are identical or not.

Summarizing this section. I have reviewed the main arguments that have been raised against the PIG model. The most important objections that were raised in the 1960s and 1970s were that (i) prevalent models of grammar could not be used alone as parsers, (ii) there was evidence that non-grammatical strategies were employed in parsing. As we have seen, though, the modifications to the grammar and parser proposed in Chapters 2–4 defuse these objections to a large extent.

There are further potential objections to the PIG model, which I address in the next section.

5.4 More Objections to Parser–Grammar Identity

In this section I address a series of other objections to the claim that the grammar and the parser are the same system, many of which I have encountered in discussions of this topic. As we will see, some of these objections are easily answered, others are more serious concerns which will require further investigation.

PARSING PREFERENCES ARE ‘SOFT’, GRAMMATICALITY IS ‘HARD.’ This is commonly cited as a reason why there must be a parser separate from the grammar. However, I think that this argument arises from overlooking why it is that preferences in parsing appear to be soft and violable.

A now sizeable number of studies of ambiguity resolution have shown that claimed structural biases in parsing can be weakened or even reversed by manipulating such factors as plausibility, discourse context or lexical frequency (e.g. Crain & Steedman 1985, Taraban & McLelland 1988, MacDonald et al. 1994). I take studies like these to have shown that many situations which are in principle structurally ambiguous are in fact only partially ambiguous, because non-structural factors are independently biasing the listener/reader towards one of the possible structural continuations. In this case, what has been shown is that the notion of ambiguity is a soft one, and not that the structural component of ambiguity resolution is itself soft.¹⁷⁸

There is another way in which the parser has been argued to show ‘soft’ behavior. This involves the observation that speakers’ performance on parsing tasks is not always consistent: sentences that are easy to understand on one testing may be hard to understand on the next. In this respect the behavioral sides of parsing and grammaticality judgements are identical. Grammaticality judgements are sometimes clear one way or another, but there is a broad middle ground of uncertainty and inconsistency. However, it should be obvious that the fuzziness of grammar-related or parser-related behavior does not tell us whether

¹⁷⁸ This does not, of course, say anything about how the various sources of structural and non-structural information contribute to ambiguity resolution. See Chapter 3 for discussion of this issue.

The Parser and the Grammar

the fuzziness is due to the grammar/parser or to other systems involved in these behaviors. This is not a retreat into abstraction—it simply reflects the fact that it is impossible to look at mental systems in isolation.

A GENERATIVE GRAMMAR IS A RECURSIVE CHARACTERIZATION OF A SET OF SENTENCES, NOT AN ALGORITHM. It is often claimed that collapsing the grammar and the parser is a conceptual error akin to confusing the axioms of logic with a system for proving theorems of logic. A generative grammar is nothing more than a recursive characterization of a set of sentences, and there is no reason to expect it to be algorithmic.

I think that this point is correct insofar as it is not a *necessary* property of a generative grammar that it be algorithmic. But this does not mean that a generative grammar *cannot* be algorithmic. As we have seen in the main body of this thesis, there are a number of advantages to a grammar which has some properties of a parsing algorithm. The attractiveness of this possibility of course does not guarantee its truth, but it does make it an alternative worth pursuing.

IF THE GRAMMAR AND PARSER ARE IDENTICAL, THEN WE CANNOT EXPLAIN WHY SPEAKERS WITH SIMILAR GRAMMATICALITY JUDGEMENTS CAN SHOW DIFFERENT PARSING PERFORMANCE. I do not accept this argument. It is quite crucial to remember that the PIG model does not just consist of a grammar. It consists of a grammar plus resources. Variations among speakers in the resources can easily account for the fact that speakers' parsing performance may vary, even when their grammaticality judgements are uniform. In fact, there have been a number of experiments which have shown that parsing performance varies in interesting ways according to various properties of speakers' memory span (cf. MacDonald, Just & Carpenter 1992, Pearlmutter & MacDonald 1995).

GRAMMAR CANNOT BE REDUCED TO PARSING, BECAUSE OF THE ARBITRARINESS OF VARIATION AMONG LANGUAGES. Remarks like this can be found in a number of places in the literature (e.g. Pritchett 1992, p.3). Such remarks point to a rather different enterprise from the one proposed here.

An implicit assumption behind this kind of argument is that the human parser is uniform across languages, in contrast to the grammars of particular languages, which show a good deal of variation. A further assumption behind such objections is that while grammars may encode all manner of linguistic arbitrariness, the parser is a relatively well-adapted and hence non-arbitrary system. Given both of these assumptions, the argument that grammar cannot be reduced to parsing goes through rather straightforwardly. However, in identifying the grammar and parser here I am quite willing to accept that this single syntactic system incorporates a good deal of arbitrariness and cross-linguistic variation.

THERE ARE MANY SENTENCES WHICH ARE EASY TO PARSE AND UNDERSTAND, THOUGH UNGRAMMATICAL. The claim that grammaticality is parsability in the limit makes the simple prediction that if a sentence is parsable then it should also be grammatical. There are a number of cases which look at first sight like

The Parser and the Grammar

clear counterexamples to this prediction. These examples involve sentences which are judged ungrammatical but which are quite easy to understand.

A well-known example of this is the *that*-trace effect illustrated in (21). (11a) and (11b) are equally easily understood, but the presence of the complementizer *that* in (11b) makes the sentence ungrammatical for speakers of most dialects of English.

- (11) a. Who do you think left the room?
b. * Who do you think that left the room?

A similar point is made by sentences which violate restrictions on double object constructions, but which are easily understood (cf. Green 1974, Baker 1979, Gropen et al. 1989, Pinker 1989).

- (12) a. * John donated the museum a painting.
b. * John reported the police an accident.

Under the characterization of ‘grammatical’ and ‘parsable’ given above, the set of sentences that are parsable should be a subset of the set of sentences that are grammatical, because parsing just involves running through derivations of the grammar under tight memory and time limitations. Given this, we should not expect to find sentences which are parsable but ungrammatical. However, there are a number of reasons why I do not take examples like (11–12) to show the independence of the grammar and the parser.¹⁷⁹

First, as I have mentioned on a number of occasions above, I draw a sharp distinction between parsing, which is a technical notion referring to the assembly of syntactic structure, and comprehension, which is a complex cognitive process with parsing as one of its parts. The sentences in (11–12) are clearly comprehended, but this does not entail that they are successfully parsed. Notice that when sentences like (12a–b) are altered such that the intended meaning is no longer so obvious, it becomes less clear whether the sentences are read as ill-formed double object constructions or *to*-less dative constructions (13).

- (13) * John reported the policeman the policewoman.

Second, the grammar may draw a distinction between representations that can be constructed and representations that are acceptable. This notion is familiar from grammatical formalisms which assume a very powerful phrase-structure generating component, which vastly ‘overgenerates’ structures, supplemented by a series of constraints which filter the representations built by the phrase structure module. If this is the case, then it is possible that people are able to build ungrammatical representations for overgenerated sentences.

¹⁷⁹ This view differs from Phillips 1995, which began with the claim that examples like (11–12) make it certain that the parser and the grammar are different systems.

5.6 On the Competence–Performance Distinction

Discussion of the relation between the parser and the grammar generally leads to some discussion of the so-called *competence-performance distinction*. Unfortunately, this rarely clarifies matters. I therefore bring up this topic here in order to first clarify where I think confusion has crept in, to state my position, and then to set the matter aside.

The competence–performance distinction is regularly used to refer to a series quite different distinctions. One of these distinctions I think we can accept on virtually *a priori* grounds. The others are entirely empirical hypotheses, with varying likelihoods. This ambiguity of usage between necessary and contingent distinctions is probably in large part to blame for the fact that the competence–performance distinction has seemed so mysterious to many people. Each of the following interpretations of the terms *competence* and *performance* appears in print in a number of places.

The first distinction is the distinction between an organism’s behavior and whatever mechanisms underlie that behavior. A speaker’s language-related behavior is the result of the complex interaction of many different linguistic and non-linguistic properties of the speaker’s brain. Therefore the language-related properties of a speaker’s brain and the speaker’s linguistic behavior are obviously quite different. This is one sense in which the distinction between *competence* and *performance* has been made.

On the other hand, the terms competence and performance are often used to distinguish between different kinds of mechanism internal to the speaker, on the one hand specialized systems for comprehending sentences and producing sentences (‘performance systems’), and on the other hand a central repository of linguistic knowledge which the input-output systems both interact with (a ‘competence system’). This kind of distinction is an entirely empirical hypothesis about the organization of linguistic knowledge.

I should clarify what I mean by this, for it is sometimes claimed that since comprehension and production of sentences are related but different processes, there must be both mechanisms that the two processes share (i.e. grammatical knowledge) and mechanisms that they do not share (i.e. comprehension and production systems). This argument is sound, except that it does not entail that there are specialized *linguistic* systems for comprehension and production. For example, motor cortex and the vocal tract are mechanisms implicated in speech production and probably not speech perception, but it is generally assumed that they are not specialized for language; similarly, the cochlea and the peripheral auditory pathways are systems that are involved in auditory sentence comprehension and not in production (ignoring for the moment issues of self-monitoring of speech), but they are also not specialized for language. Meanwhile, it remains an entirely empirical question whether there are separate *linguistic* systems for comprehension and production. My aim in this thesis has been to argue that this particular distinction is unnecessary.

Both of these distinctions are quite straightforward, and I think it is uncontroversial to claim that the first is an *a priori* distinction and the second is a distinction to be empirically established or denied. Unfortunately, because the term competence–performance distinction has been used in both cases, this has led the terms to be far more mysterious than they need to be.

The Parser and the Grammar

There is a third kind of competence–performance distinction, which is one to which I subscribe. Earlier in this chapter I said that a sentence is parsable if it can be generated by the grammar and a certain limited set of resources, and grammatical if it could in principle be generated by the grammar, if unlimited resources were available. We might equate what is parsable with performance, and what is grammatical with competence, in which case the competence–performance distinction refers to the differing capacities of a system when supplied with differing amounts of memory resources etc.

Notice that neither of these notions refers to any kind of behavior: both competence and performance refer to sentences that the grammar has the *potential* to generate under different circumstances. Also, notice that this version of the competence–performance distinction is also very much an empirical hypothesis. It depends on the possibility of distinguishing the grammar and the resources, so that the grammar can be held constant while varying the resources available to it. If the grammar and the resources cannot be individuated in this way, then this particular competence–performance distinction cannot exist.

Finally, there is a fourth use of the term *competence–performance distinction*, which corresponds to a distinction drawn in formal theories of computation, and then taken to apply automatically to the formal study of natural language. For any formal (as opposed to natural) grammar it is possible to distinguish between *what* the grammar computes and *how* the grammar does the computing, i.e. what algorithm it uses. This distinction is sometimes equated with the competence–performance distinction.¹⁸⁰ Now while this distinction needs no justification in the case of formal languages which (by *fiat*) exist independently of the systems that realize them, the distinction is not automatically relevant to an understanding of natural language. In fact, if the form of natural language is quite heavily shaped by properties of the system that realizes it—the human brain—then there is probably good reason to avoid drawing the what/how distinction made in theories of computation. In any case, I consider it an entirely empirical matter whether such a distinction will turn out to be relevant to natural language.¹⁸¹

¹⁸⁰ As, for example, in the following passage from Crocker 1996:
'...it is worth taking some time to discuss the implications of the *a priori* distinction between linguistic competence—our knowledge of the language, and performance—how we use this knowledge. Indeed, there is much confusion in the literature about whether or not such a distinction is necessary, desirable, or even meaningful. In fact, the distinction is a formal property of any processing system: inherent to any process is both a declarative semantics, a characterization of *what* the process computes, and an operational semantics, a specification of how the declarative specification is used, thereby realising a particular algorithm. That is, the process is characterised by the conjunction of a declarative specification and some operational semantics.' (pp. 8–9).

¹⁸¹ One way in which this distinction might turn out to be relevant to natural language would be in the investigation of plasticity in the neural representation of language. For example, if it turned out that certain aspects of linguistic knowledge can be supported by a variety of different neural structures, then this would be evidence for implementation-independence of certain aspects of language, and therefore the value of the what/how distinction. Although such questions fall way beyond our current understanding of the neuroscience of language, they are real issues in the neuroscience of sensory systems in animals, as for example in the work of Mriganka Sur and colleagues on the 'rewiring' of ferret auditory cortex to process visual information (e.g. Sur, Pallas & Roe 1990).

The Parser and the Grammar

Summarizing, there are at least four ways in which the term competence–performance distinction has been used, given in (14).

- (14) a. *Competence–Performance I*
The distinction between the mechanisms implicated in linguistic behavior and linguistic behavior itself.
- b. *Competence–Performance II*
The distinction between specialized mechanisms for comprehension and production of language and a central repository of linguistic knowledge.
- c. *Competence–Performance III*
The distinction between the capacities of a grammar with bounded and unbounded resources respectively.
- d. *Competence–Performance IV*
The distinction between what is computed and how it is computed, borrowed from formal theories of computation.

The arguments presented in this thesis take for granted the distinction in (14a), but call into question the distinction in (14b). I assume that the distinction in (14c) is real, which is likely but by no means necessary. The distinction in (14d) plays no role here.

Having stated my position on the competence–performance distinction, I prefer to avoid using this terminology, since it seems to lead to more confusion than clarity, given that it is used in so many different ways.

5.7 Conclusion

We end where we began. What I have been attempting to motivate throughout this thesis is the following two claims, which are not unrelated:

- Grammatical derivations proceed in a strictly left-to-right fashion.
- The grammar and the parser are the same system.

Over the course of the thesis I hope to have shown that these are more than just programmatic claims, and that they receive considerable empirical motivation, quite apart from any intuitive appeal that they might have.

I should emphasize that I have not been trying to show that sentences *can* be assembled starting at the beginning and ending at the end. Instead I have argued that they *must* be assembled in this order, in order to account for a range of generalizations about the involvement of ‘partial’ phrase markers in grammatical phenomena.

Similarly, I did not try to show that the parser and the grammar *could* be made to look identical if we are so inclined. Rather I tried to show that independently motivated accounts of grammar and of parsing turn out to have more in common than is standardly assumed.

The Parser and the Grammar

If the model of language knowledge and use that I have proposed here is correct, then this implies that parsing is more of an active process of construction than a passive process of analysis. We comprehend sentences by trying to formulate them for ourselves.

Of course, I have mostly confined my attention to just the aspect of the parser–grammar which deals with syntactic representation, and only certain aspects of that system. A good deal of further work is required in order to test the feasibility of this approach in other domains, including the phonological and interpretive aspects of language. Nevertheless, the initial steps have been encouraging.

Appendix: Reevaluating Criticisms of the DTC

Miller & McKean 1964

I discuss the criticism of Miller & McKean’s experiment in some detail, in order to show how carefully objections to the DTC experiments need to be evaluated.

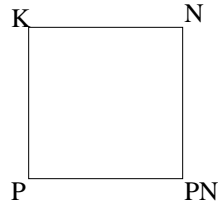
Fodor et al. 1974 (pp. 227–234) present a detailed analysis of the results of Miller & McKean 1964. They argue that the results of this experiment actually do not support the DTC, despite its apparent success, because the authors ignored one relevant transformation (*do*-support), and failed to pay attention to the effect of grammatical ordering of those transformations that they did consider, which makes a big difference to the relative transformational complexity of the different conditions in their experiment. We shall see that although Fodor et al.’s comments were mostly correct, they had no reason to conclude from this that Miller & McKean’s conclusions were flawed. Quite the opposite: their observations make Miller & McKean’s results *more consistent* with the DTC than Miller & McKean themselves had thought, as the following demonstration shows.

Recall that Miller & McKean were studying the effects of transformations on the time required to convert one member of a sentence family (e.g. P) into another sentence in the same family (e.g. PN). Probe sentences and target sentences differed from one another in varying degrees of transformational complexity. The experiment used 4 types of sentence: a kernel sentence (K), a passive (P), a negative (N) and a passive negative (PN), and the pairs of probe and target sentences were any two from this set. Miller & McKean reasoned that the transformational distance between these should be predicted by the path between vertices of the square shown in (15).¹⁸²

¹⁸² Had there been 3 transformations involved in the task, then the results would have been predicted by paths along the edges of a cube. Further transformations would require more dimensions. This analogy is due to Miller (1962).

The Parser and the Grammar

(15)



Based on (15) there were two pairs of sentences which were separated by two transformations and 4 pairs of sentences which were separated by one transformation.

(16) Transformational distance between sentences, based on (15)

- | | |
|------------|------------|
| a. K N: 1 | e. K PN: 2 |
| b. K P: 1 | f. P N: 2 |
| c. N PN: 1 | |
| d. P PN: 1 | |

Miller & McKean's results fit rather well with these predictions, as the table in (17) shows. The rows of the table represent the probe sentences, and the columns the target sentences. The times shown are the number of seconds required to perform the transformation task, minus a baseline time from control conditions in which no transformations were required. Notice that the longest times are for converting between K and PN and between N and P, as predicted by (16). Moreover, the times for converting between K and PN is predicted quite accurately by adding the times required to convert between K and N and between K and P independently.

(17) Conversion times in Miller & McKean 1964

	K	N	P	PN
K	—	0.39	0.74	1.14
N	0.40	—	1.80	1.03
P	0.88	1.83	—	0.40
PN	1.34	0.99	0.44	—

Fodor et al. raise the following objections to the predictions in (15–16). They point out that Miller & McKean failed to take into account (i) the fact that some transformations require ordering, and (ii) the contribution of the *do*-support transformation to the transformational complexity of the task.

Fodor et al. argue that the negation transformation must be ordered after the passive transformation, otherwise ungrammatical passive negatives like (18b) will be derived.

- (18) a. John didn't see Mary N
 b. * Mary didn't be seen by John PN

The Parser and the Grammar

Therefore, N cannot be transformed into PN by just one transformation. N must first be *detransformed* into K, then P can apply, and then N. If we add the fact that the *do*-support transformation must also be undone, converting N to PN requires a total of 4 transformations. Meanwhile, converting P to PN requires only one transformation (e.g. *Mary was seen by John* → *Mary wasn't seen by John*). This means that what for Miller & McKean were two equally related sentence pairs (based on distances along the edges of the square in (15)) are actually very different in their degree of relatedness according to the DTC.

The implication of these observations was supposed to be that if Miller & McKean's predictions were actually very different from what the DTC *really* predicts, then their results no longer supported the DTC. A critical step that was missing in this argument, however, was that of rechecking the new predictions against Miller & McKean's results, to see whether the new predictions received less support. In fact, a revised set of predictions, which answers Fodor et al.'s objections, conforms better to the data than Miller & McKean's original set of predictions did.

First, let us take into consideration the role of *do*-support in the transformation task, as Fodor et al. suggest. We can give this transformation the label D. Once we do this, we find that there is no longer a need to order the passive and negative transformations; all that is needed is to order both of these before the *do*-support transformation. This in turn means that the difference in transformational complexity between N and PN is closer to the difference between P and PN than Fodor et al. claim, as (19a–b) show. Converting N to PN requires two operations (–D,+P), and converting P to PN requires one (+N).

- (19) a. John didn't see Mary. N
 John TNS_{past} not see Mary (–D)
 Mary wasn't seen by John (+P) PN
- b. Mary was seen by John. P
 Mary wasn't seen by John (+N) PN

Taking into account the role of *do*-support, we can now derive the following predictions about the operations required to convert between each pair of sentence types. Now the differences between sentence pairs range between one and three transformations.

- (20) Modified predictions for steps involved in Miller & McKean task

	K	N	P	PN
K	—	+n,+d	+p	+p,+n
N	–d,–n	—	–d,–n,+p	–d,+p
P	–p	–p,+n,+d	—	+n
PN	–n,–p	–p,–d	–n	—

The next step is to check whether these predictions fit Miller & McKean's results better or worse than Miller & McKean's own predictions. We can do this by first generating estimates of the time required for each individual transformation—both for Miller & McKean's predictions (15) and for the

The Parser and the Grammar

modified predictions in (20)—and by then using these estimates to generate time predictions for all of the sentence pairs. These estimates can be compared with the observed data to find which set of predictions best models the data.

For the Miller & McKean predictions we derive an estimate of 0.41s required for N and 0.91s required for P, based on the results for those sentence pairs separated by *only* N or *only* P. For the predictions in (20) we derive estimates of 0.42s for N and 0.81s for P in the same manner. D never applies alone, so I use as the estimate for D the mean slowdown due to D in sentence pairs in which neither sentence was a kernel (i.e. PN N). N PN requires 0.22s longer than the estimate for P alone, and PN N requires 0.18s longer. We therefore take 0.20s as an estimate for the time required for D when no kernel sentence is involved, and assume that just these instances of D (which we can call D_{-k}) carry a time cost, rather than all instances of D.¹⁸³ Using these estimates for individual operations, predictions for all sentence pairs can be generated, and compared to the observed data.

(21) a. Observed data in Miller & McKean 1964

	K	N	P	PN
K	—	0.39	0.74	1.14
N	0.40	—	1.80	1.03
P	0.88	1.83	—	0.40
PN	1.34	0.99	0.44	—

b. Predictions and goodness of fit of Miller & McKean model

pred.	K	N	P	PN	diff.	K	N	P	PN
K	—	0.41	0.91	1.32	K	—	0.02	0.17	0.18
N	0.41	—	1.32	0.91	N	0.01	—	0.48	0.12
P	0.91	1.32	—	0.41	P	0.03	0.51	—	0.01
PN	1.32	0.91	0.41	—	PN	0.02	0.08	0.03	—

sum of squares of differences = 0.5754s, root mean square = 0.219s

c. Predictions and goodness of fit of revised model,
designed to answer Fodor et al.'s objections

pred.	K	N	P	PN	diff.	K	N	P	PN
K	—	0.42	0.81	1.23	K	—	0.03	0.07	0.09
N	0.42	—	1.43	1.01	N	0.02	—	0.37	0.02
P	0.81	1.43	—	0.42	P	0.07	0.40	—	0.02
PN	1.23	1.01	0.42	—	PN	0.11	0.02	0.02	—

sum of squares of differences = 0.3298s, root mean square = 0.166s

¹⁸³ Taking D_{-k} rather than D into account is not an innocent assumption, and it makes a sizeable difference to the success of the model. However, it is not entirely *ad hoc*, because Miller & McKean note that there is an overall time cost when neither of the pair of sentences is a kernel. Also, see below for a suggestion for why D_{-k} rather than D is relevant here.

The Parser and the Grammar

As the tables demonstrate, altering the predictions so as to answer Fodor et al.'s objections actually yields a better model of Miller & McKean's results than Miller & McKean's own predictions did. The revised model still significantly underestimates the time required to transform between P and N, but since this is the most complex conversion in the experiment, this problem could be answered by using something more sophisticated than an additive model of the time required for multiple operations. But this issue is not important to us here.

One concern with this demonstration is that Miller & McKean's transformation task does not require transformations to be used in the manner that any transformational theory of parsing would claim, and therefore the success of the experiment is irrelevant. Parsing a sentence may involve deriving its underlying structure, but it does not involve deriving a kernel sentence plus a list of transformations. This view of 'sentence families' is particular to the theory of Chomsky 1957. The authors themselves conclude that their transformation task leads to time delays which are considerably greater than the times required in comprehension experiments performed by other investigators which implicitly require the same transformations.¹⁸⁴

However, based on the model of the grammar–parser adopted here we can give an account of the Miller & McKean task which matches the revised predictions given in (13), but gives a better model of the results, in that it suggests an explanation for why (de)passivization should take roughly twice as long as (de)negativizing, and for why the *do*-support transformation only exacts a time cost when neither the probe nor the target is a kernel sentence.

Consistent with the left-to-right theory of the parser and the grammar, I assume that wherever possible subjects perform transformation operations in a single left-right parse of the sentence. There are some transformations which the subjects could have performed on a single left-right parse. For example, in converting from N to K (*John didn't see Mary* → *John saw Mary*) the two required operations –N and –D can be performed in sequence in one pass through the sentence. First *didn't* is translated into its syntactic features INFL + NEG, then NEG is suppressed, and then INFL can immediately be combined with the main verb to yield *saw*.

On the other hand, transformations involving passivization or depassivization cannot be performed in a single pass through the sentence. On a first pass through the sentence the thematic roles of the participants in the sentence can be recovered by generating an underlying structure for the probe sentence, and then a separate pass through the sentence is required in order to generate an active or passive sentence from the underlying structure.¹⁸⁵ Therefore, it is not surprising that the passive transformation should take roughly twice as long as the negative transformation in this paradigm. Experimental paradigms which required only *comprehension* of passives do not

¹⁸⁴ For example, studies using similar transformations in the picture verification paradigm (e.g. Gough 1965, 1966) shows time costs on the order of 0.1s per transformation.

¹⁸⁵ How the active or passive sentence is generated from an underlying structure depends on an account of sentence planning and production, which I have not provided here.

show a similar time increase relative to negation (for example, see Slobin's results discussed below).

We can also suggest a reason for the general slowdown that Miller & McKean observed in conditions in which neither the probe nor the target was a kernel sentence. In the revised model of Miller & McKean's predictions above this slowdown was attributed to the transformation D_{-k} (i.e. a *do*-support transformation when no kernel sentence is involved), but without any account of why D_{+k} imposes no time cost. What is special about the conditions involving D_{-k} is that they require conversion of an inflected auxiliary into its underlying syntactic features (e.g. was $\text{Infl}_{3\text{sg.past}}$), and then retention of this form in memory from the first pass through the sentence to the second pass through the sentence. On the other hand, in the D_{+k} transformation the inflectional features can be immediately moved onto the verb in the same pass through the sentence.

In sum, it is appropriate to level against Miller & McKean's study the criticism that its transformation task was rather artificial. This is something that Miller & McKean themselves comment upon. But it is not appropriate to claim that the results of the experiment fail to support an account of the task which uses grammatical transformations. As we have seen, such an account in fact fares rather well.

Passivization and Reversibility (Slobin 1966)

Studies by McMahon (1963) using a truth value judgement task (see above) and by Gough (1965, 1966) showed that passive and negative sentences are associated with a slowdown in a picture verification task. However, an important study by Slobin (1966: based on Slobin 1964) argued that only *reversible* passives exact a time cost in the picture verification task. Passives in which the thematic roles of the participants could be inferred without recourse to grammatical information did not lead to a similar slowdown.¹⁸⁶

Findings like these have been taken to undermine either the role of 'detransformation' operations as a component of the normal understanding of passives (Fodor et al. 1974) or the status of passives as transformationally derived constructions in the grammar (Bresnan 1978). This is because we would expect to find slowdown across-the-board in comprehending passives, if it were the case that detransformation is always required in understanding a passive.

It is true that the DTC predicts that even with non-reversible sentences, comprehension of the passive form should involve more operations than comprehension of the active form. However, the procedure used in Slobin's study made it highly unlikely that such a difference would be found. Sample stimuli for reversible and non-reversible sentence types are shown in (22). Reversed versions of all of the stimuli were also used. Note that the reversed versions of (22e-h) are highly anomalous (e.g., *the girl is being watered by the flowers*).

¹⁸⁶ Walker et al. (1968) found similar effects of reversibility in a probe latency task, in which subjects were asked to decide whether a probe word occurred in a presented sentence.

The Parser and the Grammar

- | | | | |
|------|----|--|---------------|
| (22) | a. | The dog is chasing the cat. | K (rev.) |
| | b. | The cat is being chased by the dog. | P (rev.) |
| | c. | The dog is not chasing the cat. | N (rev.) |
| | d. | The cat is not being chased by the dog. | PN (rev.) |
| | e. | The girl is watering the flowers. | K (non-rev.) |
| | | The flowers are watering the girl. | |
| | f. | The flowers are being watered by the girl. | P (non-rev.) |
| | | The girl is being watered by the flowers. | |
| | g. | The girl is not watering the flowers. | N (non-rev.) |
| | | The flowers are not watering the girl. | |
| | h. | The flowers are not being watered by the girl. | PN (non-rev.) |
| | | The girl is not being watered by the flowers. | |

Sentences were read aloud by the experimenter, after which the test picture was immediately presented. The time from presentation of the picture to the subject's truth-value decision (indicated by pushing a button) was recorded. (23) shows Slobin's results for both reversible and non-reversible sentences.

(23)		Reversible	Passive			Non-reversible	Passive	
			-P	+P			-P	+P
	Negative	-N	1.02	1.16	Negative	-N	0.75	0.75
		+N	1.40	1.49		+N	1.05	1.03

Notice that the effect of non-reversibility is more than just the facilitation of passives, it also leads to a sizeable speed up in verification times for active sentences. The times that were measured in this experiment are somewhat distant from the processes involved in comprehending the sentences, because timing did not begin until after the entire sentence had been presented acoustically, which presumably gave the subjects a few seconds to process the sentence before timing began. In the non-reversible conditions, subjects only needed to judge the plausibility of the sentence, independent of seeing the picture, in order to perform the task correctly. The task is therefore likely to have been mostly complete before any timing took place.¹⁸⁷

In sum, Slobin's study raises an issue relevant to the testing of the DTC, but given that most comprehension processes occurred well before measurement began in this study, the DTC was not actually tested in this experiment. What is perhaps surprising, then, is that *any* slowdowns were found relative to the kernel sentence. However, the across-the-board slowdown due to negation is a well-known phenomenon, which reflects a general difficulty in evaluating the truth of negative statements (cf. Johnson-Laird 1983 and references cited therein). This leaves a residual timing cost for passive in the reversible conditions, which may be attributed to either the comprehension or the storage of the passive sentences, both of which are quite consistent with the DTC.

¹⁸⁷ Forster & Olbrei (1973) and Forster (1976) used a rapid serial visual presentation task (RSVP), and found little or no effect of reversibility on the speed of processing passives.

Transformations and Memory Capacity

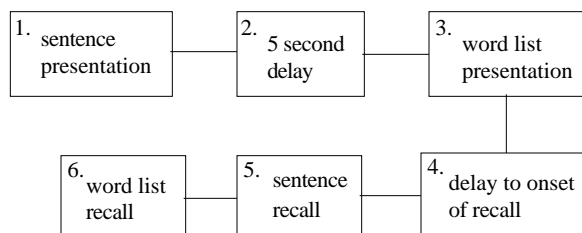
In a paper that was quite influential at the time, Savin & Perchonock (1965) argued that as the transformational complexity of a sentence increases, so its demands on short term memory also increase. Their argument was supported by an experiment in which they presented subjects with both a sentence and a list of 8 unrelated words, and asked the subjects to recall from memory first the sentence and then as many words as possible from the random list. The sentences used involved various by now familiar manipulations of transformational complexity (e.g., P, Q, N, PN), and Savin & Perchonock showed that subjects' accuracy in the recall task was inversely correlated with the transformational complexity of the sentence that the words were paired with, as would be expected if transformations add to the memory storage demands of a sentence.

A number of other experimenters subsequently replicated Savin & Perchonock's findings (Matthews 1968, Wright 1968, Epstein 1969, Glucksberg & Danks 1969), but generally with weaker results than Savin & Perchonock and with suggestions of alternative explanations for the original findings. The original success and later criticism of Savin & Perchonock's experiment was taken as representative of the demise of the initial optimism that parsing was a fairly direct implementation of the operations of the grammar (cf. Fillenbaum 1971, Levelt 1974, Wanner 1988). Therefore, it is useful to look back at the studies criticizing Savin & Perchonock's study to see which specific aspects of the interpretation of their results was called into question, and how this bears on the notion that parsing a sentence involves recapitulating its transformational history (among other things).

First, it is important to note that transformationally more complex sentences may be more complex to parse without necessarily reducing the amount of memory space available for other processes. Increased complexity could result in a time cost or a memory cost, or both, but it is quite possible for there to be a time penalty without a memory penalty, and vice versa.

The stages in Savin & Perchonock's task are shown schematically in (24).

(24)



Stages in Savin & Perchonock's (1965) sentence & word list recall task.

Savin & Perchonock showed that there was a correlation between the derivational complexity of the sentence presented (step 1) and subjects' performance on the recall task (step 6), but as later authors pointed out, the

intervening stages in the experimental task introduce a number of additional variables which may have contributed to Savin & Perchonock's results.

For example, Matthews (1968) shows that the length of the time delay between the presentation of the word list and when the subject is instructed to begin recall of the word list (i.e. step 4) has an effect on recall in this task. This is not unexpected, given that memory traces decay over time. Glucksberg & Danks (1969) therefore hypothesized that this decay-over-time factor is responsible for Savin & Perchonock's results, because transformationally more complex sentences take longer to recall (steps 4 & 5), thereby increasing the delay to the start of the word list recall (step 6). They replicated Savin & Perchonock's basic finding, but showed that the time delay from the end of word list presentation to the onset of word-list recall is a rather better predictor of performance on the memory task than the transformational complexity of the sentence used.¹⁸⁸ In other words, the time spent on steps 4 and 5 is a better predictor of performance on step 6 than the transformational complexity of the sentence in step 1 is. Importantly, though, the transformational complexity of the sentences in step 1 correlated well with the time required for steps 4 and 5, and therefore was an indirect predictor of performance on the memory task.

Therefore, it should be clear that Savin & Perchonock's findings were called into question in only a rather limited sense. They had argued that additional transformations use up additional space in short-term memory, which is just one way in which syntactic complexity may influence 'perceptual' complexity. What Glucksberg & Danks showed was that there is no clear evidence that the memory task is *directly* influenced by the complexity of the sentence. But at the same time they confirmed what many other experimenters had found, that transformational complexity does cause a latency difference in tasks requiring the recall of the sentence. Glucksberg & Danks acknowledge this in the discussion of their experiments.¹⁸⁹ So while one particular extension of the DTC was not

¹⁸⁸ The experiment contained 11 sentence types: a kernel sentence, 6 sentence types which were analyzed as containing one transformation (passive, *yes/no* question, negative, negative question (cf. Katz & Postal 1964), emphatic and *wh*-question), and 4 sentence types analyzed as containing two transformations (passive with each of question, negative, negative question and emphatic). Sentence types were grouped together according to their transformational complexity (i.e. 0, 1 or 2 transformations) in the statistical analyses, i.e. there was a much coarser-grained comparison of transformation types than in the other studies reviewed here.

Given the fact that sentences of many different types and lengths were combined in the analyses, it is remarkable that any effects of transformational complexity were found (for example: a sentence classified as a '1' that were derived from the same kernel sentence varied in length from 5 to 8 words, and thereby introduced extra variation into the data).

¹⁸⁹ Strangely, though, Glucksberg & Danks are less circumspect in the abstract to their paper, in which they claim that: 'these findings, as well as others in the literature, fail to support a transformational interpretation of sentence processing.' (p. 113). This claim seems rather distant from the conclusions reached in the text.

Glucksberg & Danks tested for a correlation between transformational complexity and two different time measures, the delay to onset of sentence recall (step 4) and the time to onset of word list recall (steps 4 and 5 combined). Times for step 4 alone showed a non-significant trend in the predicted direction, but combined

supported, the overall set of predictions that had been outlined in Miller 1962 and Miller & Chomsky 1963 was not called into question.

Studies of other transformations

Following the studies of passive, negation and question formation, a number of studies tested the relative complexity of other transformations, and in many cases found that the perceptually more complex sentence was the less transformationally complex sentence, according to analyses of the middle and late 1960s. The examples in (25) summarize some of these findings, repeating and extending presentations in Berwick & Weinberg 1983 and especially Pritchett & Whitman 1993. As both of these earlier reviews have shown, an interesting property of these findings is that in almost all instances the transformational relation between the sentences has been either abandoned or reversed by subsequent work in syntax. The symbol '<' here means that the sentence on the left was found to be 'perceptually less complex' on some measure or other.

- (25) a. the red house is on fire < the house which is red is on fire
(Fodor & Garrett 1967)
- b. Pablo looked the number up < Pablo looked up the number
(Bever, Fodor, Garrett & Mehler 1967)
- c. Slowly, Gustav opened the door < Gustav opened the door slowly
(Bever & Mehler 1967: cf. Bever 1968)
- d. John swims faster than Bob < John swims faster than Bob swims
(Jenkins, Fodor & Saporta 1965)
- e. Fido was kissed < Fido was kissed by Tom
(Fodor & Garrett 1967b)
- f. Sleeping dogs are unlikely to bark < That sleeping dogs bark is unlikely.
(Fodor, Bever & Garrett 1974)
- g. It is obvious that dogs are man's best friend < That dogs are man's best friend is obvious.
(Fodor, Bever & Garrett 1974)
- h. De studenten zijn te jong om te ontgroenen
the students are too young for to initiate/be initiated
Reading 1: 'The students are too young to be initiated.'
=
Reading 2: 'The students are too young to initiate (somebody).'
(Dutch: Levelt & Bonarius 1968)

times for steps 4 and 5 showed a highly significant correlation with the transformational complexity of the stimulus sentence, a point which is overlooked in some reviews of this study.

It should be clear to most readers today that these findings about relative ‘perceptual complexity’ are not particularly troubling to the DTC, given progress in the study of these constructions over the intervening 20–30 years. This is not the place to give detailed discussion of the various constructions listed in (25), so I just give pointers to relevant literature.

Very few people assume that adjectives are derived from relative clause constructions (see Williams 1975 for arguments), so (25a) is not troubling. As for (25b), the question of how *V NP particle* and *V particle NP* orders are derived is a matter of some controversy. Some authors have given arguments that *V NP Particle* order is underlying (Kayne 1984), others have argued that *V particle NP* is underlying (Ross 1967, Guéron 1990), and others still have argued that *both* surface orders are derived (Johnson 1991, Bowers 1993).

Most theories of adverb placement since Jackendoff 1972 have assumed that adverbs may be generated in various different positions, and do not undergo movement, so (25c) is not troubling. Regarding the comparative constructions in (25d), Hankamer (1975) has argued that *than NP* expressions are derived without deletion, while Ross (1967) provides arguments that comparative clauses like *than Bob swims* involve movement of a null operator, based on the fact that they exhibit *wh*-island effects.

Regarding the long/short passive contrast in (25e), most transformational analyses since the early 70s have not assumed that the derivation of short passives includes an agent deletion rule, and therefore short passives are derivationally simpler than long passives (Chomsky 1970, Emonds 1970, Fiengo 1974).

The pair of sentences in (25f) is unlikely to be transformationally related, given that one contains two finite clauses and the other one finite and one non-finite clause. The sentences in (25f) are equally complex transformationally, if Koster (1978) is correct in claiming that sentential subjects undergo topicalization. Similarly, since the arguments of Emonds 1970 it has been commonly assumed that the expletive subject construction in (25g) is not derived from a sentence with a clausal subject.

The Dutch sentences in (25h) represent an interesting case: the verb is ambiguous between an active (agentive) and a middle reading, and therefore the same surface string supports two different readings. Under the assumption that the subject of a middle verb is derived by raising its underlying object, then the middle has a more complex derivation than the agentive.¹⁹⁰ However, the derivational status of middles in Dutch and English is currently the status of some controversy, with arguments both for movement analyses (Stroik 1992, Hoekstra & Roberts 1993) and against movement analyses (Fagan 1988, Ackema & Schoorlemmer 1994, 1995). Therefore, the relevance of the Levelt & Bonarius experiment to the DTC is uncertain, given the current level of understanding of middle constructions.

¹⁹⁰ Note that this reasoning hides the required additional assumption that there is no derivational cost involved in suppression of the object argument in the agentive reading of *ontgroenen*. If there is some cost to the suppression of the object, then this could easily cancel out the differences in complexity between the active and middle reading of (25h).

The Parser and the Grammar

These examples illustrate clearly how difficult it is to test something like the DTC, which is heavily dependent on specific syntactic analyses, which are liable to undergo rapid changes, given the ease with which relevant data is collected and new arguments are constructed. Of the long list of claimed counterexamples to the DTC from the late 1960s, none are problematic for the DTC today, given the more detailed study of the various constructions that has taken place over the intervening 30 years.

The Parser and the Grammar

References

- Abbott, Barbara. 1976. Right Node Raising as a test for constituenthood. *Linguistic Inquiry* 7, 639–42.
- Abney, Steven. 1987. Licensing and parsing. *Proceedings of NELS 17*, Amherst, Mass.: GLSA.
- Abney, Steven. 1989. A computational model of human parsing. *Journal of Psycholinguistic Research* 18, 129–144.
- Ackema, Peter & Maaïke Schoorlemmer. 1994. The middle construction and the syntax–semantics interface. *Lingua* 93, 59–90.
- Ackema, Peter & Maaïke Schoorlemmer. 1995. Middles and nonmovement. *Linguistic Inquiry* 26, 173–197.
- Ades, Anthony & Mark Steedman. 1982. On the order of words. *Linguistics & Philosophy* 4, 517–558.
- Alphonse, Carl & Henry Davis. 1996 (to appear). Motivating non-directional movement. In H. van Riemsdijk, D. LeBlanc & D. Beermann (eds.), *Rightward movement*. Amsterdam: John Benjamins.
- Altmann, Gerry. 1988. Ambiguity, parsing strategies, and computational models. *Language and Cognitive Processes* 3, 73–98.
- Altmann, Gerry & Mark Steedman. 1988. Interaction with context during human sentence processing. *Cognition* 30, 191–238.
- Andrews, Avery. 1983. A Note on the constituent structure of modifiers. *Linguistic Inquiry* 14, 695–697.
- Aoun, Joseph & Yen-hui Audrey Li. 1989. Scope and constituency. *Linguistic Inquiry* 20, 141–172.
- Babyonyshev, Maria. 1996. *Structural connections in syntax and processing: studies in Japanese and Russian*. PhD thesis, MIT.
- Babyonyshev, Maria & Edward Gibson. 1995. Processing overload in Japanese. In C. Schütze, J. Ganger & K. Broihier (eds.), *Papers on Language Processing and Acquisition*. MITWPL 26, 1–35.
- Baker, C. Lee. 1979. Syntactic theory and the projection problem. *Linguistic Inquiry* 10, 533–581.
- Baker, Mark. 1985a. The Mirror Principle and Morphosyntactic Explanation. *Linguistic Inquiry* 16, 373–415.
- Baker, Mark. 1985b. *Incorporation: A theory of grammatical function changing*. PhD dissertation, MIT.
- Baltin, Mark & Paul Postal. 1996. More on reanalysis hypotheses. *Linguistic Inquiry* 27, 127–145.
- Barss, Andrew & Howard Lasnik. 1986. A Note on Anaphora and Double Objects. *Linguistic Inquiry* 17, 347–354.

References

- Berwick, Robert & Samuel D. Epstein. 1995. On the convergence of 'minimalist' syntax and categorial grammar. ms, MIT and Harvard University.
- Berwick, Robert & Sandiway Fong. 1993. Madama Butterfly Redux: parsing English and Japanese with a principles and parameters approach. In R. Mazuka & N. Nagai (eds.), *Japanese Sentence Processing*. Hillsdale, NJ: Erlbaum, 177–208.
- Berwick, Robert & Amy Weinberg. 1983. The role of grammars in models of language use. *Cognition* 13, 1–61.
- Berwick, Robert & Amy Weinberg. 1984. *The Grammatical Basis of Linguistic Performance*. Cambridge, MA: MIT Press.
- Bever, Thomas. 1968. A survey of some recent research in psycholinguistics. In W.J. Plath (ed.), *Specification and utilization of a transformational grammar: Scientific report number three*. Yorktown Heights, NJ: IBM Thomas Watson Research Center.
- Bever, Thomas. 1970. The cognitive basis for linguistic structures. In: J.R. Hayes (ed.), *Cognition and the Development of Language*, New York: Wiley, 279–352.
- Bever, Thomas. 1988. The psychological reality of grammar: a student's-eye view of cognitive science. In W. Hirst (ed.), *The making of cognitive science: essays in honor of George A. Miller*. Cambridge, UK: Cambridge University Press, 112–142.
- Bever, Thomas, Jerry Fodor, Merrill Garrett & Jacques Mehler. 1966. Transformational operations and stimulus complexity. ms, MIT.
- Bever, Thomas & Brian McElree. 1988. Empty categories access their antecedents during comprehension. *Linguistic Inquiry* 19, 35–43.
- Bever, Thomas & Jacques Mehler. 1967. The coding hypothesis and short-term memory. *AF Technical Report*. Cambridge, MA: Harvard Center for Cognitive Studies.
- Boas, Franz. 1947. Kwakiutl grammar, with a glossary of the suffixes. *Transactions of the American Philosophical Society*, 37, 201–377.
- Bobaljik, Jonathan. 1994. What does adjacency do? In C. Phillips & H. Harley (eds.), *The Morphology–Syntax Connection*. MITWPL 22, 1–32.
- Bobaljik, Jonathan. 1995. *Morphosyntax: The Syntax of Verbal Inflection*. PhD dissertation, MIT. Cambridge, MA: MITWPL.
- Boland, Julie & Michael Tanenhaus. 1991. The role of lexical representations in sentence processing. In: G.B. Simpson (ed), *Understanding Word and Sentence*, North-Holland: Elsevier, 331–366.
- Bonet, Eulàlia. 1991. *Morphology after syntax*. PhD dissertation, MIT. Cambridge, MA: MITWPL.
- Booij, Geert & Jerzy Rubach. 1987. Postcyclic versus postlexical rules in Lexical Phonology. *Linguistic Inquiry* 18, 1–44.
- Borsley, Robert & María-Luisa Rivero. 1994. Clitic auxiliaries and incorporation in Polish. *Natural Language and Linguistic Theory* 12, 373–422.
- Borsley, Robert, María-Luisa Rivero & Janig Stephens. 1996. Long Head Movement in Breton. In R. Borsley & I. Roberts (eds.), *Celtic and Beyond*, 53–74. Cambridge, UK: Cambridge University Press.
- Bošković, Željko. 1995a. Participle movement and second position cliticization in Serbo-Croatian. *Lingua* 96, 245–266.
- Bošković, Željko. 1995b. *Principles of economy in nonfinite complementation*. PhD dissertation, University of Connecticut. [Distributed by MITWPL.]
- Bowers, John. 1993. The syntax of predication. *Linguistic Inquiry* 24, 591–656.
- Bresnan, Joan. 1974. The Position of Certain Clause-Particles in Phrase Structure. *Linguistic Inquiry* 5, 614–619.

References

- Bresnan, Joan. 1978. A realistic transformational grammar. In G. Miller, M. Halle & J. Bresnan (eds.) *Linguistic Theory and Psychological Reality*. Cambridge, MA: MIT Press, 282–390.
- Bresnan, Joan. 1982. Control and complementation. In J. Bresnan (ed.) *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press.
- Brody, Michael. 1994. Phrase Structure and Dependency. ms., University College, London and Hungarian Academy of Sciences.
- Browne, Wayles. 1974. On the problem of enclitic placement in Serbo-Croatian. In R.D. Brecht & C.V. Chvany (eds.), *Slavic Transformational Syntax: Michigan Slavic Materials 10*, 36–52. Ann Arbor, MI: University of Michigan.
- Brysbaert, Marc & Don Mitchell. 1995. Modifier attachment in sentence parsing: evidence from Dutch. *Quarterly Journal of Experimental Psychology*.
- Carrier, Jill & Janet Randall. 1992. The argument structure and syntactic structure of resultatives. *Linguistic Inquiry* 23, 173–234.
- Carroll, J., Thomas Bever & C. Pollack. 1981. The non-uniqueness of linguistic intuitions. *Language* 57, 368–383.
- *avar, Damir & Chris Wilder. 1994. Clitic Third in Croatian. *Linguistics in Potsdam* 1, 25–63.
- Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1970. Remarks on nominalization. In R. Jacobs & P. Rosenbaum (eds.), *English Transformational Grammar*. Waltham, MA: Ginn.
- Chomsky, Noam. 1973. Conditions on Transformations. In S. Anderson & P. Kiparsky (eds.), *A Festschrift for Morris Halle*, 232–286. New York: Holt, Reinhart & Winston.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1991. Some Notes on economy of derivation and representation. In R. Freidin (ed.), *Principles and Parameters in Comparative Grammar*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1993. A minimalist program for linguistic theory. In K. Hale & S.J. Keyser (eds.), *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1995a. Bare Phrase Structure. In G. Webelhuth (ed.), *Government-Binding Theory and the Minimalist Program*. Oxford: Blackwell, pp 383–439.
- Chomsky, Noam. 1995b. Categories and Transformations. In *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, Noam & Howard Lasnik. 1977. Filters and control. *Linguistic Inquiry* 8, 425–504.
- Cinque, Guglielmo. 1986. Bare quantifiers, quantified NPs and the notion of operator at S-structure. *Rivista di Grammatica Generativa* 11.
- Clifton, Charles Jr., Shari Speer & Steven Abney. 1991. Parsing arguments: Phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language* 30, 251–271.
- Cormack, Annabel. 1984. VP Anaphora: Variables and Scope. In F. Landman & F. Veltman (eds.), *Varieties of Formal Semantics*. Dordrecht: Foris.
- Crain, Stephen & Mark Steedman. 1985. On not being led up the garden path: the use of context by the psychological parser. In D. Dowty, L. Karttunen & A. Zwicky (eds.), *Natural Language Processing: Psychological, Computational and Theoretical Perspectives*. Cambridge, UK: Cambridge University Press, 320–358.
- Crocker, Matthew. 1996. *Computational Psycholinguistics*. Dordrecht: Kluwer.

References

- Cuetos, Fernando & Don Mitchell. 1988. Cross-linguistic differences in parsing: restrictions on the use of the Late Closure strategy in Spanish. *Cognition* 30, 73–105.
- Davis, Henry & Carl Alphonse. 1992. Parsing, *wh*-movement and linear asymmetry. In *Proceedings of NELS 22*. Amherst, MA: GLSA, 87-100.
- Demiraj, Shaban. 1986. *Gramatikë Historike e Gjuhës Shqipe*. University of Tirana, Faculty of History and Philology, Tirana.
- Diesing, Molly. 1992. *Indefinites*. Cambridge, MA: MIT Press.
- Dimitrova-Vulchanova, Mila & Lars Hellan. Clitics and Bulgarian clause structure. In M. Dimitrova-Vulchanova & L. Hellan (eds.), *Papers from First Conference on Formal Approaches to South Slavic Languages*. University of Trondheim Working Papers in Linguistics 28, 363–409.
- Dowty, David. 1988. Type-Raising, Functional Composition, and Non-Constituent Coordination. In R. Oehrle, E. Bach & D. Wheeler (eds.), *Categorial Grammars and Natural Language*. Dordrecht: Reidel, pp 153–198.
- Dryer, M.S. 1980. The positional tendencies of sentential noun phrases in Universal Grammar. *The Canadian Journal of Linguistics* 25, 123–195.
- Embick, David & Roumyana Izvorski. 1995. Participle-Auxiliary Word Orders in Slavic. In *Proceedings of the Fourth Annual Conference on Formal Approaches to Slavic Linguistics*.
- Emonds, Joseph. 1970. *Root and structure-preserving transformations*. PhD dissertation, MIT.
- Emonds, Joseph. 1976. *A transformational approach to English syntax*. New York: Academic Press.
- Emonds, Joseph. 1977. The verbal complex V/V' in French. *Linguistic Inquiry* 8.
- Engdahl, Elisabet. 1983. Parasitic gaps. *Linguistics and Philosophy* 6, 5–34.
- Epstein, Samuel D. 1995. The nature of syntactic relations. ms, Harvard University.
- Epstein, W. 1969. Recall of word lists following learning of sentences and of anomalous and random strings. *Journal of Verbal Learning and Verbal Behavior* 8, 20–25.
- Ernst, Thomas. 1994. M-Command and precedence. *Linguistic Inquiry* 25, 327–335.
- Erteshik-Shir, Nomi. 1987. Right Node Raising. *MITWPL* 9, 105–117.
- Fagan, Sarah. 1988. The English middle. *Linguistic Inquiry* 19, 181–203.
- Ferreira, Fernanda & Charles Clifton Jr. 1986. The independence of syntactic processing. *Journal of Memory and Language*, 25, 348-368.
- Ferreira, Fernanda & J.M. Henderson. 1990. The use of verb information in syntactic parsing: A comparison of evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory and Cognition* 16, 555–568.
- Fiengo, Robert. 1974. *Semantic conditions on surface structure*. PhD dissertation, MIT.
- Fiengo, Robert. 1977. On trace theory. *Linguistic Inquiry*, 8, 35–61.
- Fillenbaum, Samuel. 1971. Psycholinguistics. *Annual Review of Psychology* 22, 251–308.
- Fodor, Janet. 1978. Parsing strategies and constraints on transformations. *Linguistic Inquiry* 9, 427–473.
- Fodor, Janet. 1993. Processing empty categories: a question of visibility. In G. Altmann & R. Shillcock (eds.), *Cognitive models of speech processing: The second Sperlonga meeting*. Hillsdale, NJ: Erlbaum.
- Fodor, Janet & Lyn Frazier. 1983. Local attachment in a one-stage parser. ms. University of Connecticut and UMass, Amherst.
- Fodor, Jerry. 1983. *The modularity of mind*. Cambridge, MA: MIT Press.

References

- Fodor, Jerry, Thomas Bever & Merrill Garrett. 1974. *The Psychology of Language*. New York: McGraw Hill.
- Fodor, Jerry & Merrill Garrett. 1967. Some syntactic determinants of sentential complexity. *Perception and Psychophysics* 2, 289–296.
- Fodor, Jerry, Merrill Garrett & Thomas Bever. 1968. Some syntactic determinants of sentence complexity. II Verb structure. *Perception and Psychophysics* 3, 453–461.
- Ford, Marilyn, Joan Bresnan & Ronald Kaplan. 1982. A competence-based theory of syntactic closure. In J. Bresnan (ed.), *The Mental Representation of Grammatical Relations*. Cambridge, MA: MIT Press, pp. 727–851.
- Forster, K. & I. Olbrei. 1973. Semantic heuristics and syntactic analysis. *Cognition* 2, 319–348.
- Forster, Kenneth. 1976. The autonomy of syntactic processing. Paper presented at the Convocation on Communication, MIT.
- Fowler, Carol & L. Rosenblum. 1991. The perception of phonetic gestures. In I. Mattingly & M. Studdert-Kennedy (eds.), *Modularity and the motory theory of speech perception: Proceedings of a conference to honor Alvin M. Liberman*. Hillsdale, NJ: Erlbaum, 33–59.
- Fox, Danny. 1995a. Economy and scope. *Natural Language Semantics* 3, 283–341.
- Fox, Danny. 1995b. Locality in variable binding. to appear in P. Barbosa et al. (eds.), *Is the Best Good Enough*. Cambridge, MA: MIT Press & MITWPL.
- Frazier, Lyn. 1978. *On comprehending sentences: Syntactic parsing strategies*. PhD thesis, University of Connecticut.
- Frazier, Lyn. 1985. Syntactic complexity. In D. Dowty, L. Karttunen & A. Zwicky (eds.), *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*. Cambridge, UK: Cambridge University Press.
- Frazier, Lyn. 1987. Sentence Processing: A Tutorial Review. In: M. Coltheart (ed) *Attention and Performance XII*. Hillsdale, N.J.: Erlbaum.
- Frazier, Lyn. 1987. Syntactic processing: evidence from Dutch. *Natural Language and Linguistic Theory* 5, 519–560.
- Frazier, Lyn. 1994. Sentence (re)analysis. ms, UMass Amherst.
- Frazier, Lyn. 1995. Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research*, 24, 437–468.
- Frazier, Lyn & Charles Clifton Jr. 1996. *Construal*. Cambridge, MA: MIT Press.
- Frazier, Lyn & Janet Fodor. 1978. The sausage machine: a new two-stage parsing model. *Cognition* 6, 291–325.
- Frazier, Lyn & Keith Rayner. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178–210.
- Frazier, Lyn & Keith Rayner. 1987. Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language* 26, 505–526.
- Ganger, Jennifer. 1996. ms, MIT.
- Garnham, Alan. 1983. Why psycholinguists don't care about the DTC: reply to Berwick & Weinberg. *Cognition* 15, 263–269.
- Gazdar, Gerald. 1981. Unbounded dependencies and coordinate structure. *Linguistic Inquiry* 12, 155–184.
- Gibson, Edward. 1987. paper presented at ESCOL conference.
- Gibson, Edward. 1991. *A Computational Theory of Human Language Processing: Memory Limitations and Processing Breakdown*. PhD thesis, Carnegie Mellon University.

References

- Gibson, Edward & Kevin Broihier. 1996 (in press). Optimality theory and human sentence processing. In P. Barbosa et al. (eds), *Is the Best Good Enough?* Cambridge, Mass.: MIT Press.
- Gibson, Edward & Gregory Hickok. 1993. Sentence processing with empty categories. *Language and Cognitive Processes* 8, 147-161.
- Gibson, Edward, Gregory Hickok & Carson Schütze. 1994. Processing empty categories in a parallel parsing framework. *Journal of Psycholinguistic Research* 23, 381-405.
- Gibson, Edward, Neal Pearlmutter, Gregory Hickok & Enriqueta Canseco-Gonzalez. 1996. Recency Preferences in the Human Sentence Processing Mechanism. *Cognition* 59, 23-59.
- Gibson, Edward, Neal Pearlmutter & Vicenc Torrens. 1996. ms, MIT.
- Gilboy, Elizabeth, Josep-Maria Sopena, Charles Clifton Jr. & Lyn Frazier. 1995. Argument structure and association preferences in Spanish and English complex NPs. *Cognition* 54, 131-167.
- Glucksberg, S. & J. Danks. 1969. Grammatical structure and recall: a function of the space in immediate memory or of recall. *Perception and Psychophysics* 6, 113-117.
- Goodall, Grant. 1987. *Parallel Structures in Syntax*. Cambridge, UK: Cambridge University Press.
- Gorrell, Paul. 1987. *Studies of human sentence processing: ranked parallel versus serial models*. PhD dissertation, University of Connecticut.
- Gorrell, Paul. 1993. Evaluating the direct association hypothesis: a reply to Pickering and Barry. *Language and Cognitive Processes* 8, 129-146.
- Gorrell, Paul. 1995. *Syntax and Parsing*. Cambridge, UK: Cambridge University Press.
- Gorrell, Paul. 1996. Syntactic analysis and reanalysis in sentence processing. ms, Max Planck Institute for Cognitive Neuroscience, Potsdam. to appear in volume on reanalysis edited by J.D. Fodor & F. Ferreira.
- Gough, Philip. 1965. Grammatical transformations and speed of understanding. *Journal of Verbal Learning and Verbal Behavior* 4, 107-111.
- Gough, Philip. 1966. The verification of sentences. The effects of delay of evidence and sentence length. *Journal of Verbal Learning and Verbal Behavior* 5, 492-496.
- Green, Georgia. 1974. *Semantics and syntactic regularity*. Bloomington, IN: University of Indiana Press.
- Grimshaw, Jane. 1996. Projection, heads and optimality. *Linguistic Inquiry*, to appear.
- Gropen, Jess, Steven Pinker, Michelle Hollander, Richard Goldberg & Ronald Wilson. 1989. The learnability and acquisition of the dative alternation in English. *Language* 65, 203-257.
- Grosu, Alexander. 1976. A Note on Subject Raising to Object and Right Node Raising. *Linguistic Inquiry* 7, 642-645.
- Guéron, Jacqueline. 1990. Particle, prepositions and verbs. In J. Mascaró & M. Nespó (eds.), *Grammar in Progress*. Dordrecht: Foris, pp. 153-166.
- Guéron, Jacqueline & Robert May. 1984. Extraposition and logical form. *Linguistic Inquiry* 15, 1-31.
- Hale, Kenneth. 1975. Papago intonation and word order. ms, MIT.
- Hale, Kenneth, Laverne Jeanne & Paul Platero. 1977. Three cases of overgeneration. In P. Culicover et al. (eds.), *Formal Syntax*. New York: Academic Press.
- Hale, Kenneth & Lisa Selkirk. 1987. Government and Tonal Phrasing in Papago. *Phonology Yearbook* 4, 151-183.

References

- Halle, Morris & Alec Marantz. 1993. Distributed Morphology and the Pieces of Inflection. In K. Hale & S.J. Keyser (eds.) *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*. Cambridge, MA: MIT Press.
- Halle, Morris & Kenneth Stevens. 1964. Speech recognition: a model and a program for research. In J.A. Fodor & J.J. Katz (eds.), *The structure of language: Readings in the philosophy of language*. Englewood Cliffs, NJ: Prentice Hall.
- Halpern, Aaron. 1992. *Topics in the Placement and Morphology of Clitics*. PhD dissertation, Stanford University.
- Hankamer, Jorge. 1973. Unacceptable ambiguity. *Linguistic Inquiry* 4, 17–68.
- Hankamer, Jorge. 1975. Why there are two *than*'s in English. *Proceedings of the ninth regional meeting of the Chicago Linguistics Society*.
- Hawkins, John. 1995. *A performance theory of word order and constituency*. Cambridge, UK: Cambridge University Press.
- Heim, Irene. 1985. Notes on comparatives and related matters. ms, University of Texas at Austin.
- Hewitt, S. The progressive in Breton in the light of the English progressive. In M.J. Ball, J. Fife, E. Poppe & J. Rowland (eds.), *Celtic Linguistics: Readings in the Brythonic Languages*. Amsterdam: John Benjamins, 167–88.
- Hickok, Gregory. 1993. Parallel parsing: evidence from reactivation in garden-path sentences. *Journal of Psycholinguistic Research* 22, 239–250.
- Hirschbühler, Paul. 1982. VP-Deletion and Across-the-Board Quantifier Scope. In J. Pustejovsky & P. Sells (eds.), *Proceedings of NELS 12*, pp. 132–139. Amherst, MA: GLSA.
- Hoekstra, Teun. 1988. Small clause results. *Lingua* 74, 101–139.
- Hoekstra, Teun. 1992. Aspect and theta-theory. In I. Roca (ed.), *Thematic structure: its role in grammar*. Berlin: Mouton de Gruyter.
- Hoekstra, Teun & Ian Roberts. 1993. Middle constructions in Dutch and English. In E. Reuland & W. Abraham (eds.), *Knowledge & Language, Volume 2: Lexical and Conceptual Structure*. Dordrecht: Kluwer, 183–220.
- Hoji, Hajime. 1985. *Logical Form Constraints and Configurational Structures in Japanese*. PhD dissertation, University of Washington, Seattle.
- Holmes, V. 1987. Syntactic parsing: In search of the garden path. In: M. Coltheart (ed.), *Attention and Performance XII*, Hillsdale, NJ: Erlbaum, 587–599.
- Hudson, Richard. 1976. Conjunction Reduction, Gapping, and Right Node Raising. *Language* 52, 535–562.
- Inkelas, Sharon & Draga Zec (eds.). 1990. *The phonology-syntax connection*. Stanford, CA: CSLI Publications.
- Inoue, Atsu. 1991. *A comparative study of parsing in English and Japanese*. PhD dissertation, University of Connecticut.
- Inoue, Atsu & Janet Fodor. 1995. Information-paced parsing of Japanese. In R. Mazuka & N. Nagai (eds.), *Japanese Sentence Processing*. Hillsdale, NJ: Erlbaum, 9–63.
- Izvorski, Roumyana, Tracy Holloway King & Catherine Rudin. 1996. Against *-li* lowering in Bulgarian. ms., UPenn, Stanford U. & Wayne State College.
- Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
- Jackendoff, Ray. 1990. On Larson's treatment of the double object construction. *Linguistic Inquiry* 21, 427–455.
- Jakobovits, L. & M. Miron (eds.). 1967. *Readings in the psychology of language*. Englewood Cliffs, NJ: Prentice Hall.
- Jenkins, James, Jerry Fodor and Sol Saporta. 1965. An introduction to psycholinguistic theory. ms, Center for Advanced Studies, Palo Alto, CA.

References

- Johnson, Kyle. 1991. Object positions. *Natural Language and Linguistic Theory* 9, 577–636.
- Johnson, Mark. 1989. Use of knowledge of language. *Journal of Psycholinguistic Research* 18 (1).
- Johnson-Laird, Philip. 1983. *Mental models*. Cambridge, MA: Harvard University Press.
- Juliano, Cornell & Michael Tanenhaus. 1993. Contingent frequency effects in syntactic ambiguity resolution. *Proceedings of the 15th annual conference of the Cognitive Science Society*, Hillsdale, NJ: Erlbaum, 593–598.
- Just, M.A., Carpenter, P.A. & Wooley, J.D. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 3, 228–238.
- Katz, Jerrold & Paul Postal. 1964. *An integrated theory of linguistic descriptions*. Cambridge, MA: MIT Press.
- Kayne, Richard. 1984. *Connectedness and binary branching*. Dordrecht: Foris.
- Kayne, Richard. 1985. Principles of particle constructions. In J. Guéron, H.-G. Obenauer & J.-Y. Pollock (eds.), *Grammatical Representation*. Dordrecht: Foris.
- Kennedy, A., W. Murray, F. Jennings & C. Reid. 1989. Parsing complements: Comments on the generality of the principle of minimal attachment. *Language and Cognitive Processes* 4, 51–76.
- Kennison, Shelia & Charles Clifton Jr. 1994. Resolving the ambiguity of the word *her*. Paper presented at the Psychonomics Society, St. Louis, MO.
- Kimball, John. 1973. Seven Principles of Surface Structure Parsing in Natural Language. *Cognition* 2, 15–47.
- Kitagawa, Yoshihisa. 1986. *Subjects in Japanese and English*. PhD dissertation, UMass, Amherst. Amherst, MA: GLSA.
- Klavans, Judith. 1985. The Independence of Syntax and Phonology in Cliticization. *Language* 61, 95–120.
- Klima, Edward. 1964. Negation in English. In J. Fodor & J. Katz (eds.), *The Structure of Language*. Englewood Cliffs, NJ: Prentice Hall.
- Koizumi, Masatoshi. 1993. Object agreement phrases and the Split VP Hypothesis. In J. Bobaljik & C. Phillips (eds.), *Papers on Case & Agreement I*. MIT Working Papers in Linguistics 18, 99–148.
- Koizumi, Masatoshi. 1994. Secondary Predicates. *Journal of East Asian Linguistics* 3:1.
- Koizumi, Masatoshi. 1995. *Phrase structure in minimalist syntax*. PhD dissertation, MIT. Distributed by MITWPL.
- Konieczny, Lars, Barbara Hemforth & Christoph Scheepers. 1995. PP- and NP-attachment preferences differ according to verb-placement in German sentences. Talk presented at the 8th annual CUNY Sentence Processing Conference, Tucson, AZ.
- Konieczny, Lars, Barbara Hemforth, Christoph Scheepers & Gerhard Strube. 1995. PP-attachment in German: results from eye-movement studies. In J.M. Findlay, R. Walker & R.W. Kentridge (eds.), *Eye-movement research. Mechanisms, processes and applications*. Amsterdam: North Holland, 405–420.
- Koopman, Hilda. 1983. *The syntax of verbs*. Dordrecht: Foris.
- Koopman, Hilda. 1992. On the absence of case-chains in Bambara. *Natural Language and Linguistic Theory* 10, 555–594.
- Koster, Jan. 1978. Why subject sentences don't exist. In S.J. Keyser (ed.), *Recent transformational studies in European Languages*. Cambridge, MA: MIT Press, 53–64.
- Koster, Jan. 1987. *Domains and dynasties*. Dordrecht: Foris.

References

- Kuno, Susumu. 1973. *The structure of the Japanese language*. Cambridge, MA: MIT Press.
- Kuno, Susumu. 1973. Constraints on internal clauses and sentential subjects. *Linguistic Inquiry* 4, 363–385.
- Kurtzman, Howard. 1985. *Studies in syntactic ambiguity resolution*. PhD dissertation, MIT.
- Larson, Richard. 1988. On the double object construction. *Linguistic Inquiry* 19, 335–392.
- Larson, Richard. 1989. *Light predicate raising*. Lexicon Project Working Paper #27, Department of Linguistics & Philosophy, MIT.
- Larson, Richard. 1990. Double objects revisited: reply to Jackendoff. *Linguistic Inquiry* 21, 589–632.
- Lasnik, Howard. 1995a. The forms of sentences. In L. Gleitman & M. Liberman (eds.), *Language: An Invitation to Cognitive Science, Volume 1 (second edition)*. Cambridge, MA: MIT Press, pp. 283–310.
- Lasnik, Howard. 1995b. Verb movement and inflection: Syntactic Structures meets the Minimalist Program. ms, University of Connecticut.
- Lema, José & María-Luisa Rivero. 1989. Long Head Movement: ECP vs. HMC. In *Proceedings of NELS 20*. Amherst, MA: GLSA, pp. 333–347.
- Lema, José & María-Luisa Rivero. 1991. Types of verbal movement in Old Spanish: modals, futures and perfects. *Probus* 3, 237–278.
- Levelt, Willem & M. Bonarius. 1968. *Suffixes as deep structure cues*. Heymans Bulletin HB-68-22EX. University of Groningen.
- Levelt, Willem. 1974. *Formal Grammars in linguistics and psycholinguistics* (3 volumes). The Hague: Mouton.
- Levin, Beth & Malka Rappaport Hovav. 1995. *Unaccusativity: at the syntax–lexical semantics interface*. Cambridge, MA: MIT Press.
- Levine, Robert. 1980. On the lexical origin of the Kwakwaka'wakw passive. *International Journal of American Linguistics* 46, 240–259.
- Levine, Robert. 1985. Right Node (Non-)Raising. *Linguistic Inquiry* 16, 492–497.
- Liberman, Alvin. 1957. Some results of research on the perception of speech. *Journal of the Acoustical Society of America* 29, 117–123.
- Liberman, Alvin, F. Cooper, Donald Shankweiler & Michael Studdert-Kennedy. 1967. Perception of the speech code. *Psychological Review* 74, 431–461.
- Liberman, Alvin & Ignatius Mattingly. 1985. The motor theory of speech perception revised. *Cognition* 21, 1–36.
- MacDonald, Maryellen. 1989. Priming effects from gaps to antecedents. *Language and Cognitive Processes* 4, 1–72.
- MacDonald, Maryellen. 1993. The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language* 32, 692–715.
- MacDonald, Maryellen, Marcel Just & Patricia Carpenter. 1992. Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology* 24, 56–98.
- MacDonald, Maryellen, Neal Pearlmutter & Mark Seidenberg. 1994. The lexical basis of syntactic ambiguity resolution. *Psychological Review* 101, 676–703.
- Mahajan, Anoop. 1990. *Movement Theory and the A/A' Distinction*. PhD thesis, MIT. Distributed by MITWPL.
- Maling, Joan. 1972. On “Gapping and the Order of Constituents”. *Linguistic Inquiry* 3, 101–108.
- Maling, Joan. 1980. Inversion in embedded clauses in Modern Icelandic. *Islenskt mal og almenn malfræði* 2, 176–193.

References

- Marantz, Alec. 1988a. Clitics, morphological merger and the mapping to phonological structure. In M. Hammond and M. Noonan (eds.) *Theoretical Morphology*. San Diego, CA: Academic Press, 253–270.
- Marantz, Alec. 1988b. Apparent exceptions to the projection principle. In *Morphology and Modularity: in honor of Henk Schultink*. Dordrecht: Foris.
- Marantz, Alec. 1989. Clitics and Phrase Structure. In M. Baltin & A. Kroch (eds.), *Alternative Conceptions of Phrase Structure*. Chicago: University of Chicago Press.
- Marcus, M., D. Hindle & M. Fleck. 1983. D-Theory: Talking about talking about trees. *Association for Computational Linguistics* 21, 129–136.
- Matthews, G.H. 1962. Analysis by synthesis of sentences of natural languages. In *Proceedings International Congress on Machine Translation and Applied Language Analysis*. London: HMSO.
- Matthews, W.A. 1968. Transformational complexity and short term recall. *Language and Speech* 11, 120–128.
- May, Robert. 1985. *Logical form*. Cambridge, MA: MIT Press.
- Mazuka, Reiko & Kenji Itoh. 1995. Can Japanese speakers be led down the garden path? In R. Mazuka & N. Nagai (eds.), *Japanese sentence processing*. Hillsdale, NJ: Erlbaum.
- Mazuka, Reiko & Noriko Nagai (eds.) 1995. *Japanese sentence processing*. Hillsdale, NJ: Erlbaum.
- McCawley, James. 1982. Parentheticals and Discontinuous Constituents. *Linguistic Inquiry* 13, 91–106.
- McCawley, James. 1988. *The syntactic phenomena of English*. Chicago, IL: University of Chicago Press.
- McCloskey, James. 1986. Right Node Raising and Preposition Stranding. *Linguistic Inquiry* 17, 183–186.
- McCloskey, James. 1996. On the scope of verb movement in Irish. *Natural Language and Linguistic Theory* 14.
- McMahon, Lee. 1963. *Grammatical analysis as part of understanding a sentence*. PhD dissertation, Harvard University.
- Mehler, Jacques. 1963. Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior* 2, 346–351.
- Miller, George. 1962. Some psychological studies of grammar. *American Psychologist* 17, 748–762. Also in Jakobovits & Miron 1967.
- Miller, George & Noam Chomsky. 1963. Finitary models of language users. In R.D. Luce, R.R. Bush & E. Galanter (eds.), *Handbook of Mathematical Psychology, volume 2*. New York: Wiley.
- Miller, George. & Katherine McKean. 1964. A chronometric study of some relations between sentences. *Quarterly Journal of Experimental Psychology* 16, 297–308.
- Milsark, Gary. 1983. On length and structure in sentence parsing. *Cognition* 13, 129–134.
- Mitchell, Don. 1987. Lexical guidance in human parsing: locus and processing characteristics. In M. Coltheart (ed.), *Attention and Performance XII*. Hillsdale, NJ: Erlbaum, 601–681.
- Mitchell, Don, Martin Corley & Alan Garnham. 1992. Effects of context in human sentence parsing: Evidence against a discourse-based proposal mechanism. *Journal of Experimental Psychology: Language, Memory and Cognition*, 18, 69–88.
- Mitchell, Don & Fernando Cuetos. 1991. The origins of parsing strategies. In C. Smith (ed.), *Current Issues in Natural Language Processing*. Austin, TX: Center for Cognitive Science, University of Texas.

References

- Miyagawa, Shigeru. 1995. Against optional scrambling. to appear in *Linguistic Inquiry*.
- Moltmann, Friederike. 1992. *Coordination and Comparatives*. PhD dissertation, MIT. Distributed by MITWPL.
- Muadz, Husni. 1991. *Coordinate Structures: A Planar Representation*. PhD dissertation, University of Arizona, Tucson.
- Nicol, Janet & David Swinney. 1989. The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* 18, 5–19.
- Noyer, Rolf. 1992. *Features, Positions and Affixes in Autonomous Morphological Structure*. PhD dissertation, MIT. Cambridge, MA: MITWPL. [to be published by Garland Press]
- Oehrle, Richard. 1976. *The Grammatical Status of the English Dative Alternation*. PhD dissertation, MIT. Cambridge, MA: MITWPL.
- Oehrle, Richard. 1991. Categorical frameworks, coordination and extraction. In *Proceedings of WCCFL 9*, 411–425. Stanford, CA: CSLI Publications.
- Payne, Doris & Thomas Payne. 1990. Yagua. In D.C. Derbyshire & G.K. Pullum (eds.), *Handbook of Amazonian Languages, Volume 2*, 249–474. Berlin: Mouton de Gruyter.
- Pearlmutter, Neal & Maryellen MacDonald. 1995. Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language* 34, 521–542.
- Penčev, J. 1993. *Bu^olgarski sintaksis. Upravljenje i svârzvane*. Plovdiv: Plovdivsko Universitetsko Izdatelstvo.
- Percus, Orin. 1993. The captious clitic: problems in Serbo-Croatian clitic placement. ms., MIT.
- Percus, Orin. 1995. A horse raced to Logical Form. In C. Schütze et al. (eds), *Papers on Language Processing and Acquisition*. MITWPL 26, 473–519.
- Perlmutter, David & John Robert Ross. 1970. Relative Clauses with Split Antecedents. *Linguistic Inquiry* 1, 350.
- Pesetsky, David. 1985. Morphology and logical form. *Linguistic Inquiry*, 16, 193–246.
- Pesetsky, David. 1987. *Wh-in-situ: Movement and Unselective Binding*. In E. Reuland & A. ter Meulen (eds.), *The Representation of (In)definiteness*. Cambridge, MA: MIT Press.
- Pesetsky, David. 1989. Language specific processes and the Earliness Principle. ms, MIT.
- Pesetsky, David. 1995. *Zero Syntax: Experiencers & Cascades*. Cambridge, MA: MIT Press.
- Phillips, Colin. 1995. Right Association in Parsing and Grammar. In C. Schütze, J. Ganger & K. Broihier (eds.), *Papers on Language Processing and Acquisition*, MITWPL 26, 37–93.
- Phillips, Colin & Edward Gibson. 1996 (in press). On the strength of the local attachment preference. *Journal of Psycholinguistic Research*.
- Pickering, Martin & Guy Barry. 1991. Sentence processing without empty categories. *Language and Cognitive Processes* 6, 229–259.
- Pickering, Martin & Guy Barry. 1993. Dependency Categorical Grammar and Coordination. *Linguistics* 31, 855–902.
- Pickering, Martin & Matthew Traxler. 1996. Plausibility and the processing of unbounded dependencies: an eye-tracking study. *Journal of Memory and Language* 35, 454–475.
- Pinker, Steven. 1989. *Learnability and cognition: the acquisition of argument structure*. Cambridge, MA: MIT Press.

References

- Pinker, Steven. 1994. *The language instinct*. New York: Morrow.
- Platzack, Christer. 1987. The Scandinavian languages and the Null-Subject Parameter. *Natural Language and Linguistic Theory* 5, 377–401.
- Pollock, Jean-Yves. 1989. Verb Movement, UG and the Structure of IP. *Linguistic Inquiry* 20, 365–424.
- Postal, Paul. 1971. *Cross-over Phenomena*. New York: Holt, Reinhart & Winston.
- Postal, Paul. 1974. *On Raising*. Cambridge, Mass.: MIT Press.
- Postal, Paul. 1986. *Studies of Passive Clauses*. Albany, NY: SUNY Press.
- Postal, Paul. 1993. Parasitic gaps and the across-the-board phenomenon. *Linguistic Inquiry* 24, 735–754.
- Postal, Paul. 1994. Parasitic and Pseudoparasitic Gaps. *Linguistic Inquiry* 25, 63–117.
- Pranka, Paula. 1983. *Syntax and word formation*. PhD dissertation, MIT.
- Prince, Alan & Paul Smolensky. 1993. *Optimality theory: constraint interaction in generative grammar*. Rutgers University Center for Cognitive Science report #1.
- Pritchett, Bradley. 1988. Garden path phenomena and the grammatical basis of language processing. *Language* 64, 539–576.
- Pritchett, Bradley. 1991. Subjacency in a principle-based parser. In R. Berwick, S. Abney & C. Tenny (eds.), *Principle-based parsing: computation and psycholinguistics*. Dordrecht: Kluwer, 301–345.
- Pritchett, Bradley. 1992. *Grammatical Competence and Parsing Performance*. Chicago, IL: University of Chicago Press.
- Pritchett, Bradley & John Whitman. 1993. Grammatical and perceptual complexity: The DTC in current syntax. ms., Carnegie Mellon University & Cornell University.
- Pritchett, Bradley & John Whitman. 1995. Syntactic representation and interpretive preference. In R. Mazuka & N. Nagai (eds.), *Japanese Sentence Processing*. Hillsdale, NJ: Erlbaum, pp. 65–76.
- Progovac, Ljiljana. 1993. Clitics in Serbian/Croatian: Comp as the second position. ms., Wayne State University.
- Pullum, Geoff & Arnold Zwicky. 1988. The phonology-syntax interface. In F. Newmeyer (ed.), *The Cambridge Survey of Linguistics, volume 1*. Cambridge, UK: Cambridge University Press.
- Rayner, Keith, Marcia Carlson & Lyn Frazier 1983. The interaction of syntax and semantics during sentence processing: eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior* 22, 358–374.
- Reinhart, Tanya. 1981. Definite NP anaphora and c-command domains. *Linguistic Inquiry* 12, 605–636.
- Reinhart, Tanya. 1983. *Anaphora and Semantic Interpretation*. London: Croom Helm.
- Remez, Robert. 1994. A guide to research on the perception of speech. In M. Gernsbacher (ed.), *Handbook of Psycholinguistics*. New York: Academic Press, 145–172.
- Riemsdijk, Henk van & Edwin Williams. 1986. *An Introduction to the Theory of Grammar*. Cambridge, MA: MIT Press.
- Rivero, María-Luisa. 1991. Long Head Movement and Negation: Serbo-Croatian vs. Slovak and Czech. *The Linguistic Review* 8, 319–351.
- Rivero, María-Luisa. 1993. Bulgarian and Serbo-Croatian *Yes/No* Questions. V^0 -raising to -LI versus -LI Hopping. *Linguistic Inquiry* 24, 567–575.
- Rivero, María-Luisa. 1994. Clause Structure and V-Movement in the Languages of the Balkans. *Natural Language and Linguistic Theory* 12, 63–120.

References

- Rivero, María-Luisa. 1996. Verb Movement and Economy: Last Resort. In M. Dimitrova-Vulchanova & L. Hellan (eds.), *Papers from First Conference on Formal Approaches to South Slavic Languages*. University of Trondheim Working Papers in Linguistics 28, 211–227.
- Rizzi, Luigi. 1986. On Chain Formation. In H. Borer (ed.), *Syntax and Semantics 19: The Syntax of Pronominal Clitics*. New York: Academic Press, pp 65–96.
- Roberts, Ian. 1991. Excorporation and Minimality. *Linguistic Inquiry* 22, 209–218.
- Roberts, Ian. 1992. *Verbs and diachronic syntax*. Dordrecht: Kluwer.
- Roberts, Ian. 1994a. Two types of head movement in Romance. In D. Lightfoot & N. Hornstein (eds.), *Verb Movement*. Cambridge, UK: Cambridge University Press.
- Roberts, Ian. 1994b. Second position effects and agreement in Comp. ms., University of Wales and University of Maryland.
- Ross, John Robert. 1967. *Constraints on variables in syntax*. PhD dissertation, MIT. [Published 1986 by Ablex as *Infinite Syntax!*]
- Rudin, Catherine. 1996. On pronominal clitics. In M. Dimitrova-Vulchanova & L. Hellan (eds.), *Papers from First Conference on Formal Approaches to South Slavic Languages*. University of Trondheim Working Papers in Linguistics 28, 229–245.
- Sag, Ivan. 1976. *Deletion and Logical Form*. PhD dissertation, MIT. Distributed by MITWPL.
- Sag, Ivan & Janet Fodor. 1994. Extraction without traces. In *Proceedings of WCCFL XIII*. Stanford, CA: CSLI Publications.
- Saito, Mamoru. 1985. *Some asymmetries in Japanese and their theoretical implications*. PhD thesis, MIT. Distributed by MITWPL.
- Saito, Mamoru. 1992. Long distance scrambling in Japanese. *Journal of East Asian Linguistics* 1, 69–118.
- Sauerland, Uli. 1996. Case matching effects in relative clauses. Presentation to sentence processing lab, MIT.
- Savin, H. & E. Perchonock. 1965. Grammatical structure and the immediate recall of English sentences. *Journal of Verbal Learning and Verbal Behavior* 4, 348–353.
- Schütze, Carson. 1994. *Serbo-Croatian second position clitic placement and the phonology-syntax interface*. In A. Carnie, H. Harley & T. Bures (eds.), *Papers on Phonology and Morphology*. MITWPL 21, 373–473.
- Schütze, Carson. 1995. PP-attachment and argumenthood. In C. Schütze, J. Ganger & K. Broihier (eds.), *Papers on Language Processing and Acquisition*. *MIT Working Papers in Linguistics* 26, 95–151.
- Schütze, Carson & Edward Gibson. 1996. The role of argument preference in English PP-attachment. Poster presented at the 9th annual CUNY conference on sentence processing.
- Seiter, W. 1980. *Studies in Niuean syntax*. New York: Garland.
- Sigurðsson, Halldor. 1989. *Verbal syntax and case in Icelandic*. PhD dissertation, University of Lund.
- Slobin, Dan. 1964. *Grammatical transformations in childhood and adulthood*. PhD dissertation, Harvard University.
- Slobin, Dan. 1966. Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior* 5, 219–227.
- Smith, Ian & Steve Johnson. 1979. The cliticization of case-markers in Nganhcara. Paper presented at the Australian Linguistic Society, Newcastle.
- Speas, Margaret. 1990. *Phrase structure in natural language*. Dordrecht: Kluwer.

References

- Spivey-Knowlton, Michael & Julie Sedivy. 1995. Resolving attachment ambiguities with multiple constraints. *Cognition* 55, 227–267.
- Stabler, Edward. 1992. *The logical approach to syntax: foundations, specifications and implementations of theories of government and binding*. Cambridge, MA: MIT Press.
- Stabler, Edward. 1994. Syntactic preferences in parsing for incremental interpretation. ms, UCLA.
- Steedman, Mark. 1985. Dependency and Coordination in the Grammar of Dutch and English. *Language* 61, 523–568.
- Steedman, Mark. 1988. Combinators and grammars. In R. Oehrle, E. Bach, and D. Wheeler (eds.) *Categorial Grammars and Natural Language Structures*. Dordrecht: Reidel.
- Steedman, Mark. 1996 (to appear). *Surface structure and interpretation*. Cambridge, MA: MIT Press.
- Stephens, Janig. 1982. *Word order in Breton*. PhD dissertation, University of London.
- Stowe, Laurie. 1986. Evidence for on-line gap location. *Language and Cognitive Processes* 1, 227–245.
- Stowell, Tim. 1981. *Origins of phrase structure*. PhD thesis, MIT.
- Stroik, Thomas. 1990. Adverbs as V-Sisters. *Linguistic Inquiry* 21, 654–661.
- Stroik, Thomas. 1992. Middles and movement. *Linguistic Inquiry* 23, 127–137.
- Stroik, Thomas. 1996. *Minimalism, Scope and VP-Structure*. New York: Sage Publications.
- Sturt, Patrick & Matthew Crocker. 1996. Monotonic syntactic processing: a cross-linguistic study of attachment and reanalysis. To appear in *Language and Cognitive Processes*.
- Tancredi, Christopher. 1992. *Deletion, Deaccenting and Ellipsis*. PhD dissertation, MIT. Distributed by MITWPL.
- Taraban, Roman & James McLelland. 1988. Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language* 27, 597–632.
- Thompson, Sandra & A. Mulac. 1991. The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics* 15, 237–251.
- Tomić, Olga Mišeska. 1996 (in press). The Balkan Slavic clausal clitics. *Natural Language and Linguistic Theory*.
- Travis, Lisa de Mena. 1984. *Parameters and effects of word order variation*. PhD dissertation, MIT. Cambridge, MA: MITWPL.
- Trueswell, John, & Michael Tanenhaus. 1991. Tense, temporal context and syntactic ambiguity resolution. *Language and Cognitive Processes*, 6, 303-338.
- Trueswell, John & Michael Tanenhaus. 1994. Toward a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier & K. Rayner (eds.), *Perspectives on Sentence Processing*. Hillsdale, NJ: Erlbaum.
- Trueswell, John, Michael Tanenhaus & C. Kello. 1993. Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition* 19, 528–553.
- Voorst, J. van. 1986. *Event structure*. Doctoral dissertation. University of Ottawa, Ontario.
- Walker, E., P. Gough & R. Wall. 1968. Grammatical relations and the search of sentences in immediate memory. *Proceedings of the Midwestern Psychological Association*.

References

- Wanner, Eric. 1988. Psychology and linguistics in the sixties. In W. Hirst (ed.), *The making of cognitive science: essays in honor of George A. Miller*. Cambridge, UK: Cambridge University Press, 143–152.
- Wasow, Thomas. 1972. *Anaphoric Relations in English*. PhD dissertation, MIT. Distributed by MITWPL.
- Webelhuth, Gert & Hans den Besten. 1987. Remnant topicalization and the constituent structure of VP in the Germanic SOV languages. *GLOW Newsletter* 18, 15–16.
- Wehrli, Eric. 1988. Parsing with GB-Grammar. In U. Reyle & C. Rohrer (eds.), *Natural Language Parsing and Linguistic Theories*. Dordrecht: Reidel, 177–201.
- Wexler, Kenneth & Peter Culicover. 1980. *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.
- Wilder, Chris. 1994. Some properties of ellipsis in coordination. *Geneva Generative Papers* 3, 23–61.
- Wilder, Chris & Damir Ćavar. 1994. Long Head Movement? Verb movement and cliticization in Croatian. *Lingua* 93, 1–58.
- Williams, Edwin. 1975. Small clauses in English. In J. Kimball (ed.), *Syntax and semantics*. New York: Seminar Press.
- Williams, Edwin. 1977. Discourse and Logical Form. *Linguistic Inquiry* 8, 101–139.
- Williams, Edwin. 1978. Across-the-Board Rule Application. *Linguistic Inquiry* 9, 31–43.
- Williams, Edwin. 1990. The ATB theory of parasitic gaps. In *Linguistic Review*, 6, 265–279.
- Wold, Dag. 1995. Antecedent Contained Deletion in Comparative Constructions. ms, MIT.
- Woods, W.A. 1970. Transition network grammars for natural language analysis. *Communications of the Association for Computing Machinery* 3, 591–608.
- Wright, P. 1968. Sentence retention and transformational theory. *Quarterly Journal of Experimental Psychology* 20, 265–272.
- Yngve, Victor. 1960. A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society* 104, 444–466.
- Yngve, Victor. 1961. The depth hypothesis. In R. Jakobson (ed.), *Structure of language and its mathematical aspects*. Providence, RI: American Mathematical Society.
- Yoshimura, Noriko. 1989. Parasitic pronouns. Paper presented at the Southern California Conference on Japanese/Korean Linguistics, UCLA.
- Zec, Draga & Sharon Inkelas. 1990. Prosodically constrained syntax. In S. Inkelas & D. Zec (eds.), *The phonology-syntax connection*, 365–378. Chicago: University of Chicago Press.
- Zwicky, Arnold & Geoffrey Pullum. 1986. The principle of phonology-free syntax: introductory remarks. *Ohio State University Working Papers in Linguistics* 32, 63–91.

References