# Coherence, Reference, *and the* Theory of Grammar

*Andrew Kehler*

# Coherence, Reference, and the Theory of Grammar

# Coherence, Reference, and the Theory of Grammar

Andrew Kehler

CSLI PUBLICATIONS
Center for the Study of Language and Information
Stanford, California

CSLI Publications
Center for the Study of Language and Information
Leland Stanford Junior University
Printed in the United States
06 05 04 03 02 5 4 3 2 1

*Library of Congress Cataloging-in-Publication Data*

Kehler, Andrew, 1966–
Coherence, reference, and the theory of grammar /
Andrew Kehler.
p. cm. — (CSLI lecture notes ; no. 104)
Includes bibliographical references and index.

ISBN 1-57586-215-8 (alk. paper)
ISBN 1-57586-216-6 (pbk. : alk. paper)

1. Cohesion (Linguistics)
2. Discourse analysis.
3. Grammar, Comparative and general.
I. Title. II. Series.
P302.2.K44 1999
401′.41—dc21 99-36170
CIP

∞ The acid-free paper used in this book meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

CSLI was founded early in 1983 by researchers from Stanford University, SRI International, and Xerox PARC to further research and development of integrated theories of language, information, and computation. CSLI headquarters and CSLI Publications are located on the campus of Stanford University.

CSLI Publications reports new developments in the study of language, information, and computation. In addition to lecture notes, our publications include monographs, working papers, revised dissertations, and conference proceedings. Our aim is to make new results, ideas, and approaches available as quickly as possible. Please visit our web site at
http://cslipublications.stanford.edu/
for comments on this and other titles, as well as for changes and corrections by the author and publisher.

# Contents

1 Introduction 1
2 A Theory of Discourse Coherence 11
3 Coherence and VP-Ellipsis 35
4 Coherence and Gapping 81
5 Coherence and Extraction 101
6 Coherence and Pronoun Interpretation 143
7 Coherence and Tense Interpretation 181
8 Conclusion 207
References 211
Index 223

# 1

# Introduction

In *The Psychology of Science*, the eminent psychologist Abraham Maslow remarked, as a comment on the mechanistic tradition of research in behaviorism, "I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail" (Maslow 1966, pp. 15-16). This maxim (and many variations of it) has since become widely cited, for reasons that are not difficult to see: It embodies a universal tendency that professionals working in various fields – including research, engineering, business, law, teaching, and yes, carpentry – find themselves succumbing to in the normal course of pursuing their trade. After all, when you lack access to the proper tools for a task, you are basically left with three choices. First, you can go out and acquire a suitable tool. This is usually the most difficult of the three choices, since you may have to search far and wide, or even invent the necessary tool yourself. Second, you can simply ignore the problem or task at hand and pursue something else. Lastly, you can try to use a tool that you *do* have. This is often the convenient middle road chosen, even if the fit to the task, and the result of applying the tool to it, leaves something to be desired.

I would perhaps not be going too far out on a limb if I suggested that linguistic theory is not immune to Maslow's maxim. We are all familiar with studies in which a researcher has seemingly addressed an interesting and recalcitrant linguistic fact by attempting to retrofit it to an existing set of tools, with a result that is not unlike how a pistachio might look were one to try to split it with a hammer instead of a nutcracker. This type of approach not only typically fails to shed much light on the phenomenon at hand, it often leads to the contribution of unwarranted assumptions about the object of study itself: If the tool does not fit, and one continues to maintain the assumption that it should, then the only conclusion one can draw is that the object of study must not be as it appears. It is perhaps natural to view a potentially deep and complex

phenomenon through the lens provided by a familiar (even if comparatively narrow) set of linguistic tools, but it can have dire ramifications for the state of our understanding. Indeed, a continual cycle of analyses based on a fundamentally flawed premise can keep entire bodies of literature in a state of stagnancy. The stagnancy results not from a lack of cleverness on the part of the linguists who wield their tools, but because the tools are inherently not, in and of themselves, capable of explaining the data.

My goal in this book is to introduce – or rather, reintroduce – a set of tools which I claim has a relevance to linguistic theory that has largely gone unacknowledged. The tools pertain to the methods by which hearers establish the *coherence* of multi-clause sentences and discourses. Discourse coherence is an area that to this point has typically been studied in relative isolation, as a postcursor to the production and processing of the syntactic and semantic structures of individual sentences. I intend to show, however, by way of developing a theory of coherence and then using it as a crucial component in analyses of five diverse linguistic phenomena, that coherence is not only a useful tool for analyzing mainstream linguistic problems, but also a necessary one.

## What is Coherence?

When we comprehend a discourse, we do not merely interpret each utterance within it. We also attempt to recover ways in which these utterances are related to one another. To see this, consider the rather unremarkable passage given in (1).

(1) John took a train from Paris to Istanbul. He has family there.

In most discourse situations, we will likely infer that John's having family in Istanbul is the reason for his taking a train there. While this inference is not explicitly stated, it is a natural one to draw under the assumption that the utterances bear some relationship to each other, that is, that the discourse is coherent.

We can compare passage (1) in this respect with passage (2), from Hobbs (1979).

(2) ? John took a train from Paris to Istanbul. He likes spinach.

Most people find this version to be notably odd, yet, like (1), the sentences that comprise it are both well formed and readily interpretable. Interpretation instead goes wrong during one's attempt to infer a connection between them. After all, what does going to Istanbul have to do with liking spinach? In asking this, we are questioning the coherence of the passage.

As Hobbs (1979) points out, with a little thought one might come up with a scenario in which passage (2) would become coherent. For instance, one could conjecture that perhaps the spinach crop failed in France, and Turkey is the closest country in which spinach is available. Under this assumption, one can now infer a cause-effect relationship analogous to the one we identified for passage (1), and as a result the passage is more natural.

The fact that hearers infer such relations when interpreting passages like (1), and even go so far as to contemplate additional assumptions that would license such inferences for passages like (2), illustrates that the need to establish coherence is basic to our natural language understanding capacity. Just as we attempt to identify syntactic and semantic relationships when presented with a sequence of words in an utterance, we attempt to identify coherence relationships when presented with a sequence of utterances in a discourse. The establishment of coherence is hence a powerful mechanism that allows us to communicate, and conversely understand, considerably more meaning than that conveyed by individual sentences alone. In this sense, the meaning of a discourse is greater than the sum of the meanings of its parts.

Having argued for the centrality of coherence establishment to language interpretation, we would naturally like to have a theory that characterizes the possible ways in which successive utterances can be connected to form a coherent discourse. Several researchers have in fact attempted such a characterization, in which a set of connections is enumerated as a list of *coherence relations* (Halliday and Hasan 1976, Hobbs 1979, Longacre 1983, Mann and Thompson 1987, Polanyi 1988, Hobbs 1990, inter alia; see Hovy (1990) for a compendium of over 350 relations that have been proposed in the literature). I will likewise present a list of relations here, but one in which these relations are seen to arise from a fundamental cognitive distinction. This distinction was first articulated by the philosopher David Hume in his *Inquiry Regarding Human Understanding* (1748), who makes the following general statement concerning the types of connections that can hold between ideas.

> "Though it be too obvious to escape observation that different ideas are connected together, I do not find that any philosopher has attempted to enumerate or class all the principles of association—a subject, however, that seems worthy of curiosity. To me there appear to be only three principles of connection among ideas, namely *Resemblance*, *Contiguity* in time or place, and *Cause* or *Effect*."

Hobbs (1990, pp. 101-102) was the first to point out that Hume's principles could be used as a basis for categorizing coherence relations, but he did not pursue such a categorization in depth. I will in fact argue for such a categorization. Indeed, the position I take is a strong one: That Hume's categories comprise a small set of basic types of cognitive principles that, when applied to the domain of discourse interpretation, give rise to such relations.

Upon having offered my theory of coherence, I will then utilize it as a fundamental component within analyses of five diverse and well-studied linguistic phenomena. I will describe how Maslow's maxim has raised its head in the literatures of these areas, showing that each can be characterized by two contradictory properties: (i) an implicit assumption that the data can be explained with a uniform set of tools (e.g., solely by syntactic rules, semantic mechanisms, or uniform discourse-level strategies), and (ii) a set of data that would seem to defy this assumption. In each case, I argue that the data can be explained with a cross-modular theory that interfaces a relatively straightforward account of the properties of the linguistic phenomenon in question with the effect of discourse-level interpretation processes used to establish coherence. I synopsize the book in greater detail in the following section.

## Overview of the Book

I will begin by presenting my neoHumian categorization of coherence relations in detail in Chapter 2. Therein a core set of relations is proposed along with a specification of the constraints that each imposes. For instance, the *Explanation* relation will be seen to capture the type of connection that we established for passage (1) and sought to establish for passage (2).

**Explanation:** Infer $P$ from the assertion of $S_0$ and $Q$ from the assertion of $S_1$, where normally $Q \rightarrow P$.

A clear pattern emerges from these definitions that accords with Hume's principles in terms of two criteria: (i) the type of arguments over which the constraints of each relation apply, and (ii) the type of inference processes that are used to establish these constraints.

In Chapter 3 I address the linguistic phenomenon of VP-ellipsis, illustrated in example (3).

(3) George likes his mother, and Al does too.

I focus particularly on the fundamental question that serves as the starting point for any analysis of VP-ellipsis: the level of linguistic representation at which it is resolved. We will see that analyses that operate at

the level of syntax, which generally require parallel syntactic structure between the antecedent and elided verb phrases, predict the unacceptability of examples like (4), in which the antecedent clause has been passivized.

(4) # This problem was looked into by John, and Bob did too. [look into the problem]

Analyses that operate at a purely semantic level of representation do not predict this unacceptability. Examples like (5) are in fact acceptable, however, as predicted by semantic, but not syntactic, analyses.

(5) This problem was to have been looked into, but obviously nobody did. [look into the problem]

I show that the seemingly contradictory data exhibits a pattern that correlates with the type of coherence relation holding between the antecedent and elided clauses. The account specifies the interaction between two independently-motivated sets of properties: the syntactic and referential properties of VP-ellipsis, and the properties of the processes for establishing each type of relation. This interaction will be seen to predict the pattern found in the data.

Chapter 4 addresses the gapping construction, illustrated along with its ungapped counterpart in examples (6a-b), taken from Levin and Prince (1986).

(6) a. Sue became upset and Nan $\emptyset$ downright angry.
b. Sue became upset and Nan became downright angry.

Gapping is similar to VP-ellipsis in that material has been elided from within a clause. In this case, however, only bare constituents remain in the clause, none of which is a stranded auxiliary. The behavior of gapping strongly suggests that it is primarily a syntactically-governed phenomenon, and hence most previous approaches have addressed it at that level. However, these approaches fail to predict an interesting fact about examples (6a-b). Example (6b) can be understood two ways, depending on the type of connection that is inferred between the two clauses: Either Sue and Nan became emotional independently, albeit perhaps in response to the same external stimulus (the 'symmetric' reading), or Nan's becoming angry was *caused* by Sue's becoming upset (the 'asymmetric' reading). Example (6a), on the other hand, has only the first of these readings. I will show that the ability to gap in such cases depends not only on the syntactic properties of the clauses, but also on the type of coherence relation that holds between them. As in the case of VP-ellipsis, I will show how this dependency is predicted by the interaction between two independently-motivated sets of properties:

the syntactic and referential properties of gapping (which differ in important respects from those of VP-ellipsis), and those of the inference processes underlying the establishment of coherence relations.

In Chapter 5 I address extraction from coordinate structures, with particular reference to the Coordinate Structure Constraint (CSC) originally proposed by Ross (1967).

> In a coordinate structure, no conjunct may be moved, nor may any element contained in a conjunct be moved out of that conjunct.

Ross proposed the CSC to account for the ungrammaticality of sentences such as (7).

(7) * What book did John buy and read the magazine?

Counterexamples to the CSC are well-attested, however. For instance, Ross himself notes that examples such as (8) are acceptable, in which extraction has occurred "across-the-board", that is, out of all conjuncts.

(8) What book did John buy and read?

Furthermore, Ross also points out that example (9a) is acceptable despite the fact that extraction occurs out of only the second conjunct, and Goldsmith (1985) and Lakoff (1986) note that examples (9b-c) respectively are acceptable despite the fact that extraction occurs out of only the first conjunct.

(9) a. Here's the whiskey which I went to the store and bought.
b. How much can you drink and still stay sober?
c. That's the stuff that the guys in the Caucasus drink and live to be a hundred.

I will show that this data patterns with my neoHumian categorization of coherence relations. I will then demonstrate how this pattern results from the interaction between independently motivated conditions on extraction and the constraints that need to be met in establishing each type of coherence relation.

Chapter 6 addresses the problem of pronoun interpretation. In sentences (10a-b), adapted from an example from Winograd (1972), the pronoun *they* is typically understood to refer to *the city council* and *the demonstrators* respectively. This difference is presumably due to the fact that these reference assignments make the scenario described in each passage most plausible, especially considering the fact that the syntactic conditions are the same in each case.

(10) The city council denied the demonstrators a permit because...

a. ...they *feared* violence.

b. ...they *advocated* violence.

However, in example (11), informants universally interpret the (unaccented) pronoun *her* to refer to Hillary Clinton, even though the more semantically plausible referent is Margaret Thatcher. Similarly, example (12) tends to generate a garden-path effect, in which hearers initially identify John as the referent of the pronoun instead of Bill, despite the fact that subsequent information suggests the latter as the referent.

(11) Margaret Thatcher admires Hillary Clinton, and George W. Bush absolutely worships her.

(12) John can open Bill's safe. He made a promise to get the combination changed soon.

Once again, I will show that these examples and others discussed in the pronoun interpretation literature display a pattern with respect to my neoHumian categorization of coherence relations. I then demonstrate how the inference processes underlying the establishment of these relations interact with the linguistic properties of pronouns – particularly their tendency to signal immediate interpretability with respect to a currently salient referent – to predict this behavior.

Chapter 7 addresses the problem of tense interpretation. Previous approaches have attempted to attribute the forward movement of time normally inferred between successively-described events in a narrative to the meaning of tense itself, typically the simple past. Such approaches fail to account for the fact that the simple past is compatible with any temporal ordering between events; for instance, the two events in examples (13a-d) are understood as displaying forward movement of time, backward movement of time, identical times, and no implied ordering, respectively.

(13) a. Max slipped. He spilt a bucket of water.

b. Max spilt a bucket of water. He tripped on his shoelace.

c. Max spilt a bucket of water. He spilt it all over the rug.

d. Max spilt a bucket of water. John dropped a jär of cookies.

I present this and additional data that is problematic for two types of approach to tense interpretation. I then provide an account that combines a theory of tense with the constraints imposed by coherence relations that correctly predicts this data, as well as having additional advantages over other analyses.

Finally, Chapter 8 concludes with some final thoughts and suggestions for future research directions.

As the foregoing phenomena are rather varied, the reader may be approaching this book with an interest in only a subset of them. I have tried my best to organize the book so that only Chapter 2 is a prerequisite to reading any of the individual analyses in Chapters 3, 5, 6, and 7. Chapter 4 is also dependent on the central parts of the analysis of VP-ellipsis described in Chapter 3. However, I cannot resist the temptation to encourage the reader to read the book in its entirety, since I would argue that the fact that the same theory of coherence can be used to address outstanding issues in a diverse range of areas adds to the strength of the underlying argumentation in each particular case.

It has also been my intention to present the analyses in as theory-neutral a manner as I possibly could. Whereas many underlying theoretical assumptions unavoidably remain, my main goal is to convince the reader that coherence establishment processes must be accounted for in analyses of the linguistic phenomena addressed herein, and by extension, of other interclausal phenomena that have yet to be analyzed in these terms. I therefore sought to impose as few theoretical obstacles between the reader and this message as possible, by including only those concepts that I found necessary to make the analyses concrete. I hope that the insights expressed herein will find their proper influence in a wide range of contemporary linguistic frameworks and perspectives.

## Who Might Find this Book of Interest?

This book draws significantly on work in a broad range of theoretical traditions, bringing together insights from formal, functional, cognitive, and computational linguistics. I believe that this book will be equally of interest to practitioners in each of these areas, and furthermore that it demonstrates ways in which the work of theorists in each area can be seen to tie together. With respect to areas of language processing, this book bears heavily on work in syntax, semantics, pragmatics, and discourse processing. The influence of Hume's work on my theories also make the book of interest to philosophers of language, as it illustrates a set of concrete linguistic applications of a broad and influential piece of philosophical thought. The analyses are also rooted heavily in cognitive science and artificial intelligence – particularly with respect to how we perceive our world as coherent and how we focus our attention while doing so – making the book of interest to practitioners of these fields. Finally, many of the theories expressed herein contradict prevailing assumptions in psycholinguistics, and at the same time are empirically testable themselves. As such, the work will be of interest to psycholinguists working in all areas of language processing.

## Acknowledgments

This book coalesces and significantly expands a body of research that I have been developing over the last eight years. Previous analyses of VP-ellipsis I have presented (Kehler 1993b, Kehler 1994a, Kehler 1995) are greatly revised and extended in Chapter 3. Chapter 4 similarly expands on the analyses of gapping which received a briefer treatment in early work (Kehler 1994b, Kehler 1994a, Kehler 1995). Chapters 3 and 4 also draw on material published in Kehler (2000a), which was done with kind permission from Kluwer Academic Publishers. A considerably more brief and preliminary version of the account of extraction from coordinate structures offered in Chapter 5 appeared in Kehler (1996). The analysis of pronominal reference in Chapter 6 received a preliminary treatment in Kehler (1993b) and Kehler (1995). Finally, the analysis of tense in Chapter 7 expands on the work presented in Kehler (1994c), Kehler (1995), and Kehler (2000c).

The work described herein has profited from discussions with and comments from many people whom I hereby thank, including Jennifer Balogh, Stanley Chen, Mary Dalrymple, Mark Gawron, Joshua Goodman, Barbara Grosz, Julia Hirschberg, Jerry Hobbs, David Israel, Megumi Kameyama, Chris Kennedy, Shalom Lappin, Robert Levine, Christine Nakatani, Fernando Pereira, Maria Polinsky, Ivan Sag, Liz Shriberg, Candy Sidner, Stuart Shieber, Gregory Ward, and Bonnie Webber. Mary Dalrymple deserves special thanks for providing comments on a draft. I humbly ask for forgiveness from anyone whom I have missed. Talks based on this book were presented at Stanford University, Northwestern University, the University of California at San Diego, and San Diego State University, and the book has benefited from questions and comments from those audiences. Of course, all opinions and errors remain my own.

I also owe a debt of gratitude to my publisher, Dikran Karagueuzian, for his assistance and patience. As he will be the first to tell you, this book was a long time in coming.

This book would not have been possible without the financial support I have received. This work was supported in part by grants to Stuart Shieber and Barbara Grosz at Harvard University, including National Science Foundation Grants IRI-90-09018, IRI-94-04756, CDA-94-01024, IRI-91-57996, IRI-93-50192, and a grant from the Xerox Corporation; grants to Jerry Hobbs at SRI International, including National Science Foundation and Advanced Research Projects Agency Grant IRI-9314961; and a grant to Jerry and myself, National Science Foundation Grant IIS-9619126.

Finally, I would also like to thank those who have had the frustrating experience of dealing with me as I went through the trials and tribulations of preparing this book. First, my family: my parents Frank and Ruby, and my brother Rick. Most of all, I thank my wife Jill, for whom the preparation of this book was almost as difficult as it was for me.

This book is dedicated to the memory of my friend and colleague, Megumi Kameyama, who passed away on January 23, 1999. She is sorely missed.

# 2

# A Theory of Discourse Coherence

We are so used to identifying the coherence of our surroundings that it generally escapes our notice. Constantly confronted with novel scenarios involving new and varied sets of events and objects, we typically remain oblivious to the many inferences we have to draw to see each as a coherent, interconnected situation.

For instance, consider yourself witnessing a complex scenario unfold, such as a team of firefighters preparing to fight a fire. You see people in uniforms running around frantically, yelling toward each other. Several are carrying one end of a hose, running to the back of a house. At the same time, another team is hooking the other end of the hose to a fire hydrant. This team opens the hydrant and also runs to the back of the house. These actions all make sense to you because of what you know about fire fighting, and you readily make assumptions to allow for the situation to be interpreted as a fire-fighting event. For instance, while you do not actually see the fire, you nonetheless assume that it is the house to which everyone is running that is burning. If someone mentioned to you that it was not that house but instead one down the street, you would scratch your head in confusion. Similarly, you assume that the firefighters opening the hydrant are trying to get water to flow into the hose. If no water came out, and none of the firefighters seemed to mind that fact, you would be puzzled. You would likewise be baffled if they opened the hydrant without attaching the hose first, allowing water to gush into the street. While these are all possible sequences of events, each would be strange because they contradict the assumptions you need to make to view the situation as coherent.

Understanding a discourse requires the same type of reasoning. If someone described each event in the fire fighting situation to you as I just did, you would have to make the same sorts of inferences to understand the situation. Thus, if I followed the initial description of events by

telling you that the fire was down the street, that no water came out of the hose, or that the hose was attached after the hydrant was opened, you would have been puzzled and presumably have awaited or even requested an explanation. In this case, this information would contradict the inferences you made to view the *discourse* as coherent.

Let us take a closer look at some of the constraints at play by considering examples of coherent discourse as well as cases in which coherence is not so readily identifiable. For starters, consider passage (14), adapted from a passage within a CNN news story.

(14) The domestic pharmaceutical industry fears the institution of a Medicare drug benefit. They do not want to reveal the true costs of their proprietary medicines.

This discourse demonstrates several properties of coherence that I discussed briefly in Chapter 1. It contains two statements about the pharmaceutical industry, one that describes a fear that they have, and another that describes an event that they do not want to occur. Although it is possible to interpret these statements as independent and unrelated, under normal communicative circumstances a hearer will not do so. Instead, this hearer will infer a relationship between the two; in this case, the second sentence is most naturally understood as a cause and the first sentence as its effect. Example (14) is thus a case in which the Explanation relation (see Chapter 1) is operative, since the content of the second sentence explains the content of the first.

The inference of Explanation in this example generates a presupposition, specifically that the true costs of healthcare products which qualify for Medicare coverage must be made public. In the case that the hearer knew this already, this presupposition is likely to go unnoticed, as many do in normal modes of discourse. On the other hand, if the hearer did not already know this (as was the case when I first read this passage), the new information is likely to be assumed and accommodated into the hearer's knowledge store as long as it remains consistent with his or her existing beliefs about the world. In fact, given that this passage came from a reliable source, one might feel relatively comfortable communicating this presupposed information as fact to someone else, even though it was not actually stated anywhere in the passage.

Of course, if this information did not accord with the hearer's beliefs (for instance, she had information that contradicted it), she would be well within her rights to question it, for instance, with a response of the sort given in (15).

(15) Actually, the current bill for expanding Medicare to cover prescription drugs drops the cost disclosure requirement.

Again, nowhere in (14) does it directly say otherwise. The response is licensed only by the inference that the hearer must make to establish passage (14) as coherent.

On the other hand, sometimes the inferences needed to establish the intended relationships between utterances are not so easy to draw. Consider (16), a variant of example (14):

(16) The domestic pharmaceutical industry fears the institution of a Medicare drug benefit. Newt Gingrich has been campaigning for George W. Bush.

Example (16) comes across as less coherent, precisely because the two statements seem to be unrelated. In particular, it is hard to determine exactly what connection the hearer should establish or accommodate. This notwithstanding, a hearer might be lead to *attempt* to identify a scenario that would support such a connection. For instance, one might believe, or find it plausible to believe, that since Newt Gingrich is unpopular, his campaigning for Bush is likely to help Bush's opponent, Al Gore, become elected. This assumption, combined with the knowledge that Gore supports a Medicare drug benefit, could explain how Newt Gingrich's support of Bush might cause someone to fear the institution of a Medicare drug benefit. In this case, as with example (14), the hearer is assuming information that will allow an Explanation relation to be established between the propositions denoted by the utterances. The fact that the causal chain is less readily recoverable in this case adversely impacts the coherence of the passage, assuming that the current context does not already make this information available.

Of course, there are other types of connection that can serve as the basis for establishing coherence. Consider example (17).

(17) Al supports a medicare drug benefit. George favors a tax cut.

Passage (17) is coherent by virtue of what has been called a *Parallel* relation, which is indicated by the fact that similar relations (in this case, positions on political issues of the day) are attributed to similar entities (two politicians running for president). The coherence of passage (17) can be compared to the less coherent example (18).

(18) Al tried to distance himself from Bill. George smirked a lot.

Without sufficient context, it is more difficult to identify a reasonably specific set of commonalities between the two sentences in passage (18) than it is for passage (17). However, if passage (18) is stated in response to the question *What did the presidential candidates do during their campaigns?*, the passage becomes much more coherent under the common topic provided by the question. Indeed, this context makes

the basis for establishing a Parallel relation between the propositions explicit.

Finally, consider passage (19), adapted from an example discussed by Hobbs (1990).

(19) A flashy-looking campaign bus arrived in Iowa. Soon afterward, George W. Bush gave his first speech of the primary season.

Passage (19) is coherent by virtue of what Hobbs calls an *Occasion* relation. As with the other relations, establishing Occasion requires that certain inferences be made; for instance, a hearer is likely to infer that George W. Bush was on the bus and that his speech was delivered in Iowa. Passage (19) can be compared to passage (20).

(20) A flashy-looking campaign bus arrived in Iowa. About two hours later, Al was giving a response in Tennessee.

Again, without additional assumptions, passage (20) is of questionable coherence. To understand the passage as coherent, a hearer could nonetheless make a number of inferences: Bush was on the bus, he arrived for the purpose of making a speech, he in fact made a speech, and Al Gore was responding to that speech. Again, the greater extent to which information must be assumed renders the passage marginal as compared to (19).

In each of these examples, a hearer is presented with two sentences that independently are readily understood. Interpretation does not stop there, however, as the hearer is further inclined to assume unstated information necessary to explain the co-occurrence of those sentences. As Hobbs (1979) says:

> .. the very fact that one is driven to such explanations indicates that some desire for coherence is operating, which is deeper than the notion of a discourse just being "about" some set of entities. (p. 67)

Indeed, examples (14) and (16–20) show that the need to resolve coherence is a central facet of language understanding. Just as naturally as humans attempt to recover the implicit syntactic structure of a string of words communicated to them, they appear to also attempt to recover the implicit "coherence structure" of a sequence of utterances communicated to them.

I will call the process of determining the coherence of a discourse *coherence establishment.* As we will see, the evidence for coherence establishment has led a number of researchers to posit a set of possible coherence relations that can hold between adjacent segments of a discourse, including Explanation, Parallel, and Occasion. Some of these re-

searchers have also described computational interpretation mechanisms for establishing such relations. Despite the centrality of coherence establishment processes to language interpretation, however, their potential influence is rarely considered by researchers who develop theories of any number of linguistic phenomena that operate across clauses.

In this chapter, I present an analysis of coherence relations and the inference processes that underlie their establishment in terms of a small and basic set of core principles. In the remainder of this book, I then show how this account can be utilized in analyses of the linguistic forms described briefly in Chapter 1. In each of these analyses, properties of the constraints imposed by the different relations will prove crucial in accounting for data that is beyond the scope of other theories.

## 2.1 A Theory of Coherence Relations

Recall that examples (14), (17), and (19) illustrate three types of connection between utterances that a hearer might try to establish as a basis for determining coherence. Given suitable assumptions, a Parallel relation can be established for example (17), an Explanation relation for example (14), and an Occasion relation for example (19).

In this section, I claim that these three relations are exemplars of the three broader types of "connection among ideas" posited by Hume (1748) and discussed in Chapter 1, specifically *Resemblance*, *Cause-Effect*, and *Contiguity*, and provide an analysis that captures this fact. I analyze a larger set of relations, many taken or adapted from a set of relations proposed by Hobbs (1990), as belonging to one of these classes. The three classes are shown to differ systematically in two respects: (i) in the type of arguments over which the coherence constraints are applied, and (ii) in the type of inference process underlying this application.

### 2.1.1 Resemblance Relations

The first class I consider is Resemblance. The recognition of Resemblance requires that commonalities and contrasts among corresponding sets of entities and relations be recognized. For each relation, the hearer identifies a relation $p_1$ that applies over a set of entities $a_1, ..., a_n$ from the first sentence $S_1$, and a corresponding relation $p_2$ that applies over a corresponding set of entities $b_1, ..., b_n$ from the second sentence $S_2$. Coherence results from inferring a common (or contrasting) relation $p$ that subsumes $p_1$ and $p_2$, along with a suitable set of common (or contrasting) properties $q_i$ of the arguments $a_i$ and $b_i$. I refer to the set of $q_i$, the identity of which is determined as part of the inference process, as a *property vector*, written $\vec{q}$. I will also refer to corresponding arguments $a_i$ and $b_i$ as *parallel elements*, or alternatively *parallel arguments*.

The canonical instance of a Resemblance relation is *Parallel.*

**Parallel:** Infer $p(a_1, a_2, ...)$ from the assertion of $S_1$ and $p(b_1, b_2, ...)$ from the assertion of $S_2$, where for some property vector $\vec{q}$, $q_i(a_i)$ and $q_i(b_i)$ for all $i$.

An example of the Parallel relation is given in sentence (21).

(21) Dick Gephardt organized rallies for Gore, and Tom Daschle distributed pamphlets for him.

Here, the parallel arguments $p_1$ and $p_2$ correspond to the relations denoted by *organized rallies for* and *distributed pamphlets for* respectively; the common relation $p$ that subsumes these might thus be roughly the relation denoted by *do something to support.* Likewise, the parallel elements $a_1$ and $b_1$ correspond to Dick Gephardt and Tom Daschle, who share the common property $q_1$ of being people who are presumably known by the discourse participants to be high-ranking democratic politicians. The parallel elements $a_2$ and $b_2$ correspond to the meanings of *Gore* and *him*, which share a trivial common property $q_2$ in that the two terms denote the same individual.

Instead of focusing the inference on the commonalities among corresponding relations and entities in the utterances, one may wish to draw attention to points of departure among either of these, which yields two definitions for the *Contrast* relation. In the first, the relations expressed by the utterances are contrasted:

**Contrast (i):** Infer $p(a_1, a_2, ...)$ from the assertion of $S_1$ and $\neg p(b_1, b_2, ...)$ from the assertion of $S_2$, in which for some property vector $\vec{q}$, $q_i(a_i)$ and $q_i(b_i)$ for all $i$.

Passage (22) is an example in which this definition of Contrast applies.

(22) Gephardt supported Gore, but Armey opposed him.

In the second definition, a set of parallel entities is contrasted.

**Contrast (ii):** Infer $p(a_1, a_2, ...)$ from the assertion of $S_1$ and $p(b_1, b_2, ...)$ from the assertion of $S_2$, where for some property vector $\vec{q}$, $q_i(a_i)$ and $\neg q_i(b_i)$ for some $i$.

Passage (23) is an example in which this definition applies.

(23) Gephardt supported Gore, but Armey supported Bush.

The difference between Parallel and Contrast is determined primarily by whether the similarities or differences among a set of entities or events are highlighted. Thus, the same set of clauses can often participate in either relation. In many cases, the one intended by the author is

indicated by the choice between the conjunctions *and* (Parallel) and *but* (Contrast).

Other relations in the Resemblance class derive from a membership or subset relationship between elements in a set of clauses. For instance, the *Exemplification* relation holds between a general statement followed by an example of the generalization.

**Exemplification:** Infer $p(a_1, a_2, ...)$ from the assertion of $S_1$ and $p(b_1, b_2, ...)$ from the assertion of $S_2$, where $b_i$ is a member or subset of $a_i$ for some $i$.

Example (24) illustrates the Exemplification relation.

(24) Young aspiring politicians often support their party's presidential candidate. For instance, Bayh campaigned hard for Gore in 2000.

Analyzing (24) as an Exemplification requires that one infer that Bayh is a young aspiring politician and that Gore was his party's presidential candidate. While not directly expressed in the above definition, the subset relationship can also hold between the relations instead of one or more pairs of entities (or both). The same is true for the next two relations.

The *Generalization* relation is similar to Exemplification, except that the ordering of the clauses is reversed.

**Generalization:** Infer $p(a_1, a_2, ...)$ from the assertion of $S_1$ and $p(b_1, b_2, ...)$ from the assertion of $S_2$, where $a_i$ is a member or subset of $b_i$ for some $i$.

An example of Generalization is shown in sentence (25).

(25) Bayh campaigned hard for Gore in 2000. Young aspiring politicians often support their party's presidential candidate.

Because the constraints are essentially the same as for Exemplification, establishing Generalization here requires that the same inferences be drawn.

Recall that the main difference between Parallel and Contrast was that the latter involved negation. Similarly, we can introduce negation within the constraints for Exemplification and Generalization to derive two definitions for the *Exception* relation, depending on the clause order.

**Exception (i):** Infer $p(a_1, a_2, ...)$ from the assertion of $S_1$ and $\neg p(b_1, b_2, ...)$ from the assertion of $S_2$, where $b_i$ is a member or subset of $a_i$ for some $i$.

**Exception (ii):** Infer $p(a_1, a_2, ...)$ from the assertion of $S_1$ and $\neg p(b_1, b_2, ...)$ from the assertion of $S_2$, where $a_i$ is a member or subset of $b_i$ for some $i$.

Examples in which these two definitions apply are given in (26) and (27) respectively.

(26) Young aspiring politicians often support their party's candidate. However, Rudy Guiliani supported Mario Cuomo in 1994.

(27) Rudy Guiliani supported Mario Cuomo in 1994. Nonetheless, young aspiring politicians often support their party's candidate.

In order to establish Exception in these two cases, one must infer that Mario Cuomo was *not* Rudy Guiliani's party's candidate.

Finally, the *Elaboration* relation can be seen as a limiting case of the Parallel relation, in which the parallel entities $a_i$ and $b_i$ are in fact identical.

**Elaboration:** Infer $p(a_1, a_2, ...)$ from the assertions of $S_1$ and $S_2$.

Elaborations are generally restatements; thus while the corresponding relations and entities are constrained to be the same, the perspective from or level of detail at which they are described will generally be different. An example is given in (28).

(28) A young aspiring politician was arrested in Texas today. John Smith, 34, was nabbed in a Houston law firm while attempting to embezzle funds for his campaign.

The inference that only one event is being described instead of two is crucial for understanding this passage.

These relations and the constraints associated with them are summarized in Table 1. Note that I have characterized the constraints involving set membership and subset relationships in terms of sets of properties over the participating individuals and sets. Thus, in each case the properties of the more general argument comprise a subset of the properties of the more specific one, which has the effect of reversing the direction of the subset relation. The conjunctions shown are examples of sentence connectives that are consistent with the given relation, however these may not always be mandatory, they may not be unique in being compatible with the relation, and they may be compatible with other coherence relations also.

Table 1 highlights the similarities among the Resemblance relations in terms of the nature of the constraints they impose and the types of arguments over which they apply. The inference processes that are used to apply these constraints are in fact manifestations of our more general cognitive ability to reason analogically. In particular, they require that we be able to categorize entities and events within some common (and perhaps novel) domain or perspective, draw out correspondences

| Relation | Constraints | Conjunctions |
|---|---|---|
| Parallel | $p(p_1)$ and $p(p_2)$, $q_i(a_i)$ and $q_i(b_i)$ | and |
| Contrast | $p(p_1)$ and $\neg p(p_2)$, $q_i(a_i)$ and $q_i(b_i)$<br>$p(p_1)$ and $p(p_2)$, $q_i(a_i)$ and $\neg q_i(b_i)$ | but |
| Exemplification | $p(p_1)$ and $p(p_2)$ ; $q_i(a_i) \subset q_i(b_i)$ | for example |
| Generalization | $p(p_1)$ and $p(p_2)$ ; $q_i(b_i) \subset q_i(a_i)$ | in general |
| Exception | $p(p_1)$ and $\neg p(p_2)$ ; $q_i(a_i) \subset q_i(b_i)$<br>$p(p_1)$ and $\neg p(p_2)$ ; $q_i(b_i) \subset q_i(a_i)$ | however<br>nonetheless |
| Elaboration | $p_1 = p_2$, $a_i = b_i$ | that is |

TABLE 1 Resemblance Relations

with respect to salient commonalities and distinctions among them, and see certain entities or events as special cases of others along varying dimensions.

There are also several steps that need to be carried out before (and in some respects in tandem with) this inference. First, one must identify the number and identity of arguments to a Resemblance relation, as it is not known *a priori* how many arguments there are: The common relation $p$ to be inferred can be of any arity, including zero. Furthermore, in addition to identifying the appropriate argument vectors $\vec{a}$ and $\vec{b}$ from their respective utterances, one must also determine which members of $\vec{a}$ are parallel to which members of $\vec{b}$. With these steps completed, the constraints associated with a Resemblance relation can be applied to the parallel arguments.

In its most complex form, the determination of Resemblance can require arbitrarily deep and knowledge-intensive reasoning. For instance, Hobbs (1990) gives the following example, from a physics textbook.

(29) The ladder weighs 100 lb with its center of gravity 20 ft from the foot, and a 150 lb man is 10 ft from the top.

In this context, these clauses are parallel because they both express forces on objects at some location. Identifying this parallelism requires a certain degree of nontrivial inference, and certain knowledge of the domain of physics. Indeed, this nontriviality is what makes this passage a suitable part of a physics problem.

Nonetheless, it is common for clauses in a Resemblance relation to wear their parallelism on their 'syntactic sleeves' as it were, making relation and parallel element identification easier for the hearer. For instance, the parallel arguments to the Parallel relation for passage (30) are congruent with the syntactic structure: *Hillary* is parallel to *Bill*, *Eleanor Roosevelt* to *Jack Kennedy*, and *admire* to *look up to*.

(30) Bill looks up to Jack Kennedy, and Hillary admires Eleanor Roosevelt.

Indeed, reducing the degree of *syntactic* parallelism appears to cause a corresponding reduction in the ease with which parallelism can be identified at the coherence level:

(31) Bill looks up to Jack Kennedy, and Eleanor Roosevelt is admired by Hillary.

Passage (31) is identical to passage (30) except that the second clause has been passivized, a change that does not affect the semantic relations expressed. Nonetheless, most would agree that (31) has diminished coherence: The speaker has made it more difficult to identify which arguments are parallel, increasing the processing burden on the hearer.

To summarize this section, to establish a Resemblance relation the hearer identifies a common relation $p$ that applies over a set of entities $a_1, ..., a_n$ from the first sentence and a set of entities $b_1, ..., b_n$ from the second sentence, and performs operations based on categorization, comparison, and generalization on each pair of parallel elements. While the reasoning underlying the establishment of Resemblance is a purely semantic process, the process of argument identification and alignment utilizes cues from the syntactic structure of the utterances, and thus speakers can aid hearers' comprehension by structuring their utterances accordingly.

### 2.1.2 Cause-Effect Relations

Next I consider the Cause-Effect category. The establishment of a Cause-Effect relation is based primarily on a different type of reasoning than that for Resemblance, in which the hearer draws a path of implication connecting a pair of propositions $P$ and $Q$ identified from the first and second sentences $S_1$ and $S_2$ respectively. For my purposes here I use the term 'implication' in a loose sense, to mean roughly "could plausibly follow from", rather than the stronger relation found in classical logic. In particular, the applicability of the implication relation to particular examples might be contingent on other properties being true of the world.

The canonical case of a Cause-Effect relation is *Result*, exemplified by sentence (32).

**Result:** Infer $P$ from the assertion of $S_1$ and $Q$ from the assertion of $S_2$, where normally $P \rightarrow Q$.

(32) George is a politician, and therefore he's dishonest.

Here, $P$ corresponds to the meaning of *George is a politician*, and $Q$

corresponds to the meaning of *he's dishonest.* The information that must be presupposed in order to establish coherence is that *being a politician implies being dishonest.* This same presupposition is required for establishing the coherence of the next three examples.

The *Explanation* relation is *Result* with reversed clause ordering.

**Explanation:** Infer $P$ from the assertion of $S_1$ and $Q$ from the assertion of $S_2$, where normally $Q \rightarrow P$.

An Explanation relation is often indicated by the conjunction *because* as in sentence (33), but not necessarily so, per sentence (34).

(33) George is dishonest because he's a politician.

(34) George is dishonest. He's a politician.

The *Violated Expectation* relation is used to contrast an actual effect with an expected or desired effect in light of a potential cause, as exemplified in sentence (35).

**Violated Expectation:** Infer $P$ from the assertion of $S_1$ and $Q$ from the assertion of $S_2$, where normally $P \rightarrow \neg Q$.

(35) George is a politician, but he's honest.

Finally, *Denial of Preventer* is Violated Expectation with reversed clause ordering, exemplified in sentence (36).

**Denial of Preventer:** Infer $P$ from the assertion of $S_1$ and $Q$ from the assertion of $S_2$, where normally $Q \rightarrow \neg P$.

(36) George is honest, even though he's a politician.

These relations are summarized in Table 2.

| Relation | Presuppose | Conjunctions |
|---|---|---|
| Result | $P \rightarrow Q$ | and (as a result) therefore |
| Explanation | $Q \rightarrow P$ | because |
| Violated Expectation | $P \rightarrow \neg Q$ | but |
| Denial of Preventer | $Q \rightarrow \neg P$ | even though despite |

TABLE 2 Cause-Effect Relations

Table 2 highlights the similarities between the four Cause-Effect relations. Establishing each relation requires that the hearer identify propositions $P$ and $Q$ from sentences $S_1$ and $S_2$ respectively and infer an implicational relationship between them. Cause-effect relations therefore

contrast with Resemblance relations in that they focus on the identification of the clause-level semantics for each expression (the $P$ and $Q$) instead of the (possibly) subclausal arguments $p$, $a_i$, and $b_i$. It is therefore unsurprising that sentences in a Cause-Effect relation are often not syntactically parallel, since there would be no reason to expect that such parallelism would aid the process of establishing coherence.

### 2.1.3 Contiguity Relations

The final class I consider is Contiguity. This category is a bit murkier than the other two in several respects, and I will tentatively posit only one relation for it: Hobbs's (1990) *Occasion* relation. Occasion allows one to express a sequence of eventualities centered around some system of entities. Hobbs offers two versions (slightly reworded below).

**Occasion (i):** Infer a change of state for a system of entities from $S_1$, inferring the final state for this system from $S_2$.

**Occasion (ii):** Infer a change of state for a system of entities from $S_2$, inferring the initial state for this system from $S_1$.

An example of the Occasion relation is given in passage (37).

(37) George picked up the speech. He began to read.

Occasion can be seen as a mechanism for communicating a complex situation in a multi-utterance discourse by using states of affairs as points of connection between partial descriptions of that situation. Much of what makes for a coherent Occasion is thus based on knowledge gained from human experience about how eventualities can enable (or otherwise set the stage for) other eventualities in the world and the granularity with which people conceptualize such eventualities and change resulting from them. The definitions of this relation are given in less formal terms than the others because precise constraints that utilize this knowledge prove difficult to state explicitly.

One attempt to encode such knowledge was manifest in the 'scripts' approach pursued by Roger Schank and colleagues. Scripts are data structures that encode representations of a set of event-types that typically co-occur in a relatively predictable order. The following example is from Samet and Schank (1984).

(38) Larry went into a restaurant. The baked salmon sounded good and he ordered it.

Passage (38) is perfectly coherent despite the fact that a number of intermediate events that presumably occurred are not mentioned at all. For instance, we expect that between entering the restaurant and deciding on a dish, Larry was seated at a table (either by a waitperson or

himself). We also assume that after being seated but before deciding, he looked at a menu or was told about the salmon by a waitperson. We likewise assume that he spoke to the waitperson, and so forth. A restaurant script that encodes our knowledge about normal sequences of events when dining out allows for these inferences to be made during the establishment of Occasion.

The scripts approach leaves many questions open, however, regarding the level of detail at which eventualities need to be encoded, under exactly what conditions can various intermediate eventualities be left unsaid, and whether scriptal representations will prove too limited to capture our ability to interpret less predictable (but nonetheless coherent) series of events, among many others (see Samet and Schank (1984) for further discussion of this last point). I will not attempt to answer these questions here; the important point for my purposes is that *some* encoding of such knowledge, along with an inferential system for reasoning with it, is necessary for establishing Occasion relations. As such, past treatments of coherence relations that have equated Occasion with temporal progression (e.g., Halliday and Hasan (1976), Longacre (1983), inter alia) are too underconstraining. Hobbs (1990) in fact demonstrates this with an example similar to passage (19), repeated below as (39).

(39) A flashy-looking campaign bus arrived in Iowa. Soon afterward, George W. Bush gave his first speech of the primary season.

As we have seen, understanding passage (39) as a coherent Occasion requires inferences beyond the asserted information that the events occur in temporal progression, such as that Bush was on the bus and the speech was delivered in Iowa. In general, assumptions will be required that allow the final state of the first sentence to be identified as the initial state of the second, and hence temporal progression in the absence of a common scenario connecting the events is insufficient in and of itself.

Although many details remain to be investigated, we can assume for our purposes that hearers have a set of principles with which to establish coherent Occasions and a knowledge store of experience that these principles can utilize.

## 2.2 Basic Principles of Coherence

The set of coherence relations I have just presented is but one of many that have been proposed in the literature (Halliday and Hasan 1976, Longacre 1983, Mann and Thompson 1987, Hobbs 1990, Martin 1992, Sanders et al. 1992, inter alia). I will not attempt here to describe the myriad of ways in which these proposals both overlap and differ. Nonetheless, the existence of competing proposals brings to light a more

fundamental question that needs to be addressed, that is, on what basis should theories of coherence relations be evaluated and compared.

Sanders et al. (1992) pinpoint two primary criteria: *descriptive adequacy* and *psychological plausibility.* An analysis is descriptively adequate to the extent to which its relation set covers the diversity of naturally-occurring data. While all of the aforementioned analyses were undoubtedly informed by data analysis to some degree, some pursue the goal of descriptive adequacy to a greater extent than others. One that considers it to be the primary motivating factor is Rhetorical Structure Theory (Mann and Thompson 1987, henceforth RST). RST posits a set of 23 relations that can hold between two adjacent spans of text, termed the *nucleus* (the more central text span) and *satellite* (the span containing less central, supportive information).[1] RST relation definitions consist of five fields, the first three of which place constraints on these spans: *Constraints on Nucleus, Constraints on Satellite, Constraints on the Combination of Nucleus and Satellite, The Effect,* and *Locus of the Effect.* Consistent with their goal of providing a tool for analyzing texts rather than a precise scientific theory, these fields contain textual descriptions as opposed to more formal characterizations. The definition of their Evidence relation is given below as an exemplar.

**Relation name:** EVIDENCE
**Constraints on nucleus:** The hearer might not believe nucleus to a degree satisfactory to the speaker.
**Constraints on satellite:** The hearer believes the satellite or will find it credible.
**Constraints on the combination of nucleus and satellite:** The hearer's comprehending the satellite increases the hearer's belief of the nucleus.
**The Effect:** The hearer's belief of the nucleus is increased.
**Locus of the Effect:** The nucleus.

Mann and Thompson claim that their relations are suitable for describing a large and varied set of texts, but ultimately suggest that the set is open to extension:

> There are no doubt other relations which might be reasonable constructs in a theory of text structure; on our list are those which have proven most useful for the analysis of the data we have examined. (p. 8, fn. 5)

[1] A small set of relations are actually multi-nuclear and can relate more than two spans of text, such as the JOINT relation.

Knott and Dale (1994) point out several problems associated with positing a subjective and open-ended list of relations in this way. They note, for instance, that without *a priori* constraints on relation definitions one could just as easily define relations that describe incoherent texts. They suggest the possibility of defining an *Inform-Accident-and-Mention-Fruit* relation that would cover example (40).

(40) ? John broke his leg. I like plums.

If one can add relations to the theory when necessary, the claim that the theory is sufficient for analyzing a large and varied set of texts is not particularly meaningful, especially if there is nothing in the theory to prevent the analogous description of arbitrary incoherent texts.

Likewise, Knott and Dale point out that the descriptive adequacy criterion alone could be used to support any of a number of sets of relations, thus leaving us without a way to compare RST against other analyses. Taking this to the extreme, they suggest that one could posit only the relations *causal* and *non-causal*. While this set is obviously too general; the descriptive adequacy criterion alone says nothing about what the correct level of granularity for relations is, thus leaving no grounds for the proponents of RST to claim that their relations are not also too general, or for that matter, too specific. Thus, if we are to have a scientific theory of relations, it cannot be not up to us to concoct them for our own purposes – our job instead is to uncover the pre-existing ground truth. On scientific grounds, it is difficult to see how an unconstrained and potentially unbounded catalog of relations could give rise to an explanatory account of coherence.

This leads us to the second criterion for evaluating a theory of coherence, psychological plausibility. We expect that there are fundamental cognitive principles at work which will serve both to constrain the set of possible relations, and to provide an explanation for why a particular set of relations is to be preferred to one containing more, fewer, or different relations. What are these principles? Once again, I take my cue from Hobbs, who says:

> It is tempting to speculate that these coherence relations are instantiations in discourse comprehension of more general principles of coherence that we apply in attempting to make sense out of the world we find ourselves in, principles that rest ultimately on some notion of cognitive economy. [...] Recognizing coherence relations may thus be just one way of using certain very general principles for simplifying our view of the world. (Hobbs 1990, p. 101)

My categorization is based on this view, being rooted in a small set of more general and basic cognitive methods for establishing connections between ideas suggested by Hume. As I have already indicated, the inference processes underlying Resemblance and Cause-Effect relation recognition can be seen as being based on two familiar operations from artificial intelligence: categorization and subsumptive reasoning with respect to a semantic classification, and implication based on axioms contained in a knowledge base. After distinguishing top-level categories on this basis, other factors then differentiate the relations within each class, such as the order of propositions and the existence of negation within the relation constraints. (As we will soon see, my analysis has these factors in common with the analysis of Sanders et al. (1992).) In the case of the Resemblance category, additional distinctions result from whether parallel entities and relations stand in an identity or set membership/inclusion relationship. Other factors that make even more fine-grained distinctions could conceivably be added to enlarge the relation set within each category. Because such extensions would keep my overall categorization intact, they would likely not affect the claims I will make regarding the behavior of linguistic phenomena.

Although my claims will remain preliminary, additional evidence for the hierarchy might ultimately come from the meanings that are typically associated with conjunctions. In general, there is no one-to-one relationship between conjunctions and coherence relations, indeed the mapping is many-to-many. On the other hand, conjunctions do typically constrain the type of relation that can be inferred. Consider first the meanings of the conjunction *and* that are commonly cited in the literature, exemplified by sentences (41a-c).

(41) a. Bill went to the movies, and Hillary went to the store. (Parallel)
b. Bill went to the movies, and (then) he came home. (Occasion)
c. Bill went to the movies, and (as a result) Hillary got upset. (Result)

The operative relations in these passages are the canonical exemplars of the categories Resemblance, Contiguity, and Cause-Effect respectively; they are the ones that correspond to standard clause ordering without any of the arguments to the relation being negated. Thus, an analysis of *and* in my framework would need not attribute any ambiguity to it. It could simply be marked for standard clause order and positive polarity; the three 'meanings' would then result from the fact that it leaves the category of coherence relation unspecified.

Similarly, the conjunction *but* has meanings consistent with both Resemblance and Cause-Effect, as seen in examples (42a-b).

(42) a. Bill went to the movies, but Hillary went to the store. (Contrast)
b. Bill went to the movies, but (nevertheless) Hillary didn't get upset. (Violated Expectation)

Violated Expectation and Contrast are both characterized by standard clause order and negative polarity. Thus, *but* can simply be ascribed these properties, again leaving it unspecified for relation type. In this case, a Contiguity relation is ruled out since this class does not contain a relation that is associated with negative polarity. Other connectives that constrain the possible relations more narrowly (e.g., *because, although, despite*) would specify the category of the coherence relation in addition to these other features.

**Comparison with Sanders et al. (1992)** The 'molecular' style of approach to coherence relations that I have sketched out – in which relations are actually composites of more basic underlying features – is shared with the account of Sanders et al. (1992), who likewise consider the psychological plausibility criterion to be the primary motivating factor of their analysis (see also Sanders et al. (1993) and Sanders (1997)). They posit four basic features:

**Basic Operation:** If the connection is fundamentally rooted in implication the basic operation is *causal*, otherwise it is *additive*.

**Order of Segments:** In a causal relation, which by definition involves an implication of the form $P \rightarrow Q$, the order is *basic* if the first segment expresses $P$ and the second expresses $Q$. Otherwise, the order is *nonbasic*. Order of segments is not distinguished for additive relations, which are considered to be, for the purposes of coherence, insensitive to order.

**Polarity:** A relation has *positive* polarity if it connects the content of each segment as they stand, whereas it has *negative* polarity if the negation of the content of one of the segments participates directly in the connection.

**Source of Coherence:** A relation has a *semantic* source of coherence if the segments are related at the level of propositional content, whereas the source of coherence is *pragmatic* if they are related at the level of illocutionary meaning.

From four two-valued parameters one would expect to obtain 16 relations, but there are actually only 12 since their Additive relations do not

distinguish between order of segments (i.e., these relations are symmetric).[2] By breaking down relations into more primitive features, Sanders et al. take a step toward a more principled and explanatory account of coherence of the type I have sought here. Although this approach does not offer an exhaustive account of all the different coherence relations that researchers have proposed, the resulting set of relations is economic, cognitively motivated, and leaves open the possibility that other factors interact with these features to yield a more comprehensive set of distinctions. As such, Sanders et al.'s and my proposals share many characteristics;[3] both see coherence relations as arising from more fundamental semantic and cognitive primitives that place them in naturally organized classes.

Several substantive differences remain, however. First, their top-level organization only distinguishes between two categories, causal and additive, whereas I distinguish three. Sanders et al. in fact treat additive relations as a rather weak relation:

> The first question in identifying the coherence relation is therefore: Is the relation between P and Q a causal relation? If it is not, then the relation is additive. (p. 6)

> An additive operation exists if only a conjunction relation P & Q can be deduced between two discourse segments, that is, if all that can be deduced is that the discourse segments are true for the speaker. (p. 7)

This criterion is problematic, primarily because it appears to be open to the criticism that Knott and Dale levied at RST: By defining additive relations as everything that is not causal, the relation is left so open-ended as to not rule anything out. For instance, Knott and Dale's *Inform-Accident-and-Mention-Fruit* relation meets this definition, assuming the two statements in example (40) are true for the speaker. So we are left with no mechanism to predict why some discourses are incoherent. In contrast, I do not find any important sense in which the constraints required for any class of relations are inherently weaker than for any other. Furthermore, the constraints for Resemblance and Conti-

[2]Strictly speaking, Sanders et al. actually use the primitives to generate *classes* of relations. Sanders et al. (1992) list 17 relations, in which two of the 12 classes each contain three relations, another class contains two, and the rest each contain one. Sanders et al. (1993) also list 17 relations, but in this case two relations are categorized within each of five classes.

[3]Their paper was brought to my attention shortly after Kehler (1994a) appeared. The fact that these two independently developed analyses arrived at very similar conclusions increases my belief in their shared claims.

guity are fundamentally distinct – the crucial properties that make cases of Resemblance coherent (inference of common and contrasting properties) are not necessary to establish Contiguity, and vice versa – and as such they belong in separate categories.[4] My stronger focus on formally specifying the constraints imposed by relations and the nature of the inference processes underlying their establishment – a focus that will prove crucial for the analyses posited in the remainder of this book – helps here to clarify these distinctions.

Certain other differences are smaller in scope. Beyond the fact that I use different relation names, I offer a richer set of relations in the Resemblance category, in particular I include Generalization, Exemplification, and Elaboration. Including these relations would presumably be a minor adjustment to the Sanders et al. framework, since they already allow for the Exception relation on the basis of it exhibiting a different degree of specificity than their Opposition (Contrast) relation. Similar factors distinguish the three aforementioned relations from Parallel.

The major remaining difference is that I have not made the distinction that Sanders et al. refer to as *Source of Coherence*. The distinction is exemplified by two possible meanings of their sentence (43), paraphrased in (44a-b).

(43) Maybe John is home because he is ill.

(44) a. {That John is at home because he is ill} may be the case.
b. Because he is ill, it may be the case {that John is at home}.

The distinction corresponds to whether the segments are related at the locutionary or illocutionary level. Under the reading given in (44a), the source of coherence in (43) is *semantic*, since the segments are related at the level of their propositional (locutionary) content. On the other hand, the source of coherence in (43) under reading (44b) is *pragmatic*, because the clauses are related at the speech act (illocutionary) level. That is, the speaker's *reason for saying* that maybe John is home is driven by the fact the speaker knows that he is ill.

As Sanders et al. note, other authors have made similar distinctions (van Dijk 1979, Halliday and Hasan 1976, Martin 1983, Redeker 1990, Sweetser 1990, inter alia). Sweetser actually draws a three-way division between *content*, *epistemic*, and *speech act* readings, exemplified in (45a-c) respectively, as discussed by Sanders (1997).

(45) a. John came back because he loved her.

[4]Sanders et al. specifically reject distinguishing the category of 'temporal relations', but we have already seen that constraints other than temporal progression also underlie the establishment of Contiguity.

b. John loved her, because he came back.

c. What are you doing tonight, because there's a good movie on.

The source of coherence distinction is often considered to be less clear cut than the others; see Sanders (1997) for an extensive discussion. By excluding this distinction from my categorization, however, I do not mean to suggest that it does not exist. Rather, I have no reason to believe that the constraints underlying such relations and the corresponding inference processes that establish them, other than operating at different levels of force, are different in any respect that would impact the analyses I posit in the remainder of this book. Therefore, for my purposes I set this issue aside.

**Comparison with Hobbs (1990)** As I have indicated, the majority of the relations and definitions that I have used here are either identical to or variants of relations proposed by Hobbs (1990). The theory offered in this chapter owes a large debt to his important work; a brief comparison highlighting places in which the two accounts diverge is therefore in order. The differences apply both to the relations themselves and to the manner in which they are categorized.

The differences with respect to the relation set itself are minor. First, I have added a relation, *Denial of Preventer*, to round out the Cause-Effect category, as well as others not specifically listed but suggested in his text (e.g., *Result, Exception*). I also consider his *Ground-Figure* relation to be a subcase of *Occasion*, and do not include his *Evaluation* relation. In describing the latter relation, Hobbs says that "$S_1$ tells you why $S_0$ was said", and thus it is perhaps better treated as a causal relation with a pragmatic source of coherence (in Sanders et al.'s terminology). Finally, I have altered the specification of several of his relation definitions in ways that highlight the relevant commonalities among them.

As I have already indicated, Hobbs was the first to note that coherence relations could be classified with respect to Hume's categorization (Hobbs 1990, pp. 101-102). He does not pursue this idea in depth, however, and the cursory classification he does mention along these lines differs from mine with respect to the placement of several relations. He opts instead to categorize his relations with respect to the following question: Why do we want to call a sequence of utterances a single discourse rather than simply a sequence of utterances? Viewing this question in terms of the situations in which discourses take place, he identifies four considerations that may hold when a speaker and hearer engage in a discourse, listed in (46).

(46) a. The speaker wants to convey a message.

b. The message is in service of some goal.

c. The speaker must link what he says to what the listener already knows.

d. The speaker should ease the listener's difficulties in comprehension.

These considerations respectively give rise to four classes of coherence relations in his analysis: the *Occasion* relation, the *Evaluation* relation, the *Ground-Figure* and *Explanation* relations, and the *Expansion* relations. The categorization is thus based on the claim that each relation addresses one of these four needs. I do not find this motivation for grouping relations convincing, however, as I would submit that *all four* of these considerations hold for any discourse segment, and not just one that would in turn suggest a relation from a particular category. I therefore favor an approach in which the categorization is based on the types of coherence that we extract from the external world.

## 2.3 Identifying Relations in Examples

To summarize to this point, I have categorized a set of coherence relations into three categories: Resemblance, Cause-Effect, and Contiguity. These categories differ systematically in two respects: in the type of arguments over which the coherence constraints are applied, and in the central type of inference process underlying this application. The question remains as to how the correct relation can be identified for a given example for the purpose of testing the predictions of a linguistic analysis. Of course, one needs to use tests that are distinct from the various phenomena for which one is trying to account.

In the best of worlds, one could use a mechanical procedure for determining the relations. Computational algorithms for establishing relations have in fact been proposed. For example, Hobbs et al. (1993) outline a procedure for utterance interpretation and coherence establishment based on the inference rule of logical abduction. In their framework, an utterance is interpreted by "proving" its logical form, in which assumptions can be made at a specified cost where necessary. This approach is then scaled up to the multi-utterance level, in which world and domain knowledge are used to determine the most plausible coherence relation holding between utterances in a discourse segment. Those interested in further details of the system are referred to their paper; see also Kehler (2000b) for a walkthrough on a fairly simple example.

Unfortunately, rather hefty obstacles stand in the way of applying such a procedure to arbitrary examples. First, just about any fact about the world could conceivably be necessary for interpreting any given ex-

ample with which a mechanical system might be confronted. The number of axioms that would be required to encode all of this knowledge would obviously be enormous. Second, even if we had such a knowledge store, we lack robust mechanisms for effectively managing such knowledge and constraining the inference process. As such, neither Hobbs et al.'s system nor any other proposed mechanical procedure currently approaches the robustness necessary to reliably determine the correct relations for unconstrained passages. Therefore, we must use some other method for the purpose of linguistic study.

Ultimately we will have to rely on our intuitions when applying the constraints dictated by the relation definitions. If we proceed with some care, however, we can also get an indication of the relation by applying paraphrase tests using conjunctions and other indicator words. Of course, simple connectives do not always constrain the possibilities to a single coherence relation; we have already seen that the meaning of *and*, for instance, is compatible with the Parallel, Occasion, and Result relations. We can instead use more complex connectives that constrain the possible relations to a single instance. For instance, if the clauses are (or can be) conjoined by *and*, then an ability to paraphrase with connectives such as *and similarly*, *and likewise* or *and ... too* signals Parallel, whereas tests using conjunctions and other indicator words. Of course, simple connectives do not always constrain the possibilities to a single coherence relation; we have already seen that the meaning of *and*, for instance, is compatible with the Parallel, Occasion, and Result relations. We can instead use more complex connectives that constrain the possible relations to a single instance. For instance, if the clauses are (or can be) conjoined by *and*, then an ability to paraphrase with *and therefore* or *and as a result* signals Result. An ability to paraphrase with *and then* may signal Occasion, but often not to the exclusion of other relations, as I discuss in Chapter 3. See also Lakoff (1971) for an insightful discussion of the relations associated with conjunctions.

Using these tests, we see that passage (21), repeated below as (47a), is an instance of the Parallel relation, since it is perhaps best paraphrased by using the connective *and likewise* as shown in sentence (47b).

(47) a. Dick Gephardt organized rallies for Gore, and Tom Daschle distributed pamphlets for him.

b. Dick Gephardt organized rallies for Gore, and likewise Tom Daschle distributed pamphlets for him.

This paraphrase can be compared with the one using the connective *and as a result* in (48); this version results in a meaning that hearers would normally not assign to (47a) outside of context.

(48) Dick Gephardt organized rallies for Gore, and as a result Tom Daschle distributed pamphlets for him.

In contrast, consider the modification of passage (14) shown in (49).

(49) The domestic pharmaceutical industry does not want to reveal the true costs of their proprietary medicines. They fear the institution of a Medicare drug benefit.

The reading that hearers normally assign to this passage is best paraphrased by using the connective *and as a result*, as shown in (50), and not *and likewise* per example (51).

(50) The domestic pharmaceutical industry does not want to reveal the true costs of their proprietary medicines, and as a result they fear the institution of a Medicare drug benefit.

(51) The domestic pharmaceutical industry does not want to reveal the true costs of their proprietary medicines, and likewise they fear the institution of a Medicare drug benefit.

Similarly, the conjunction *but* will generally signal a Contrast relation when it can be paraphrased with *but in contrast*, whereas it signals a Violated Expectation relation when it can be paraphrased with *but surprisingly* or *but counter to expectation*. The conjunctions *even though* and *even when* typically signal a Denial of Preventer relation, and *because* typically signals an Explanation relation. For more in-depth discussions of the relationship between connectives and coherence, see papers by Knott and colleagues (Knott and Dale 1994, Knott and Mellish 1996, Knott and Sanders 1998), inter alia.

We also have to consider what happens when a passage simultaneously satisfies the constraints of more than one relation, with neither being clearly dominant. We would presumably not expect any constraints imposed by one of the relations to be affected by the existence of others. As I will discuss at appropriate points in the book, however, it appears that certain such cases may actually have a more intermediate status.

It is important to bear in mind, however, that although connectives can serve to constrain the set of coherence relations that can hold between two or more utterances, connectives in and of themselves do not *create* coherence. For example, passage (52a) is of marginal coherence assuming an Explanation relation, and including the connective *because* in (52b) does nothing to change this fact.

(52) a. The domestic pharmaceutical industry fears the institution of a Medicare drug benefit. Former president George Bush hates broccoli.

b. The domestic pharmaceutical industry fears the institution of a Medicare drug benefit because former president George Bush hates broccoli.

Any coherence relation indicated by a connective must still be established just as if the coherence relation were not signalled. The establishment of Explanation in both (52a) and (52b) fails because of the lack of causal knowledge that would explain how the state of former president George Bush hating broccoli could cause the domestic pharmaceutical industry to fear the institution of a Medicare drug benefit.

Having said all of this, there are still many open questions regarding the manner in which hearers establish coherence. For instance, I will have little to say about the actual process by which a particular coherence relation is chosen over another; for my purposes I will assume that the analyses developed here operate within a framework in which interpretations for each possible relation are attempted in parallel. None of my analyses rely critically on this assumption, however. I briefly discuss several other remaining issues in the conclusion to the book. For now, however, suffice it to say that I do not think that the gaps in our current understanding regarding coherence establishment processes prohibit their effective use within linguistic theory.

## 2.4 Conclusion

To conclude this chapter, I have presented a theory in which coherence relations are classified into three general categories originally suggested by David Hume (1748): *Cause-Effect*, *Resemblance*, and *Contiguity*. The categories are distinguished by the types of arguments over which the constraints imposed by the relations apply, and the inference mechanisms underlying this application. I have also provided further motivation for the theory by suggesting ways in which the relations in each class can be derived from more primitive notions.

The categorization of relations given here goes beyond previous ones in two respects. First, along with Sanders et al. (1992), it presents one of the few categorizations based on particular aspects of the formal definitions of such relations, as opposed to subjective judgments about what relations intuitively belong together. Second, previous studies of coherence relations have generally operated within the confines of the field of text coherence itself. In the next five chapters, I will provide *applications* of my theory, in which it is utilized to predict otherwise puzzling behavior of five diverse linguistic phenomena.

# 3

# Coherence and VP-Ellipsis

The first linguistic phenomenon I address is the VP-ellipsis construction, exemplified by sentence (53).

(53) George likes his mother, and Al does too.

The hallmark of VP-ellipsis the appearance of a stranded auxiliary verb, such as *does* in the second clause of (53), which indicates the 'elision' of a verb phrase. Interpreting a clause with VP-ellipsis therefore requires that the meaning of the missing VP be recovered, typically from the meaning of another clause, in this case the first clause of (53). Following the terminology of Dalrymple et al. (1991), I refer to the antecedent clause in such examples as the *source* clause, and the clause in which the ellipsis is manifest as the *target* clause. VP-ellipsis may give rise to certain ambiguities when the source clause contains a pronoun or other context-dependent form. Sentence (53) may receive one of two readings, for example, one in which Al likes *George's* mother, and one in which Al likes *his own* mother. These meanings are termed *strict* and *sloppy* readings respectively.

As I indicated in Chapter 1, there is an ongoing debate concerning the level of language processing at which VP-ellipsis is resolved. Two major positions have been staked out in the literature. According to the first, VP-ellipsis is resolved at some level of syntactic structure (Sag 1976, Williams 1977, Haïk 1987, Hellan 1988, Lappin 1993b, Fiengo and May 1994, Hestvik 1995, Lappin 1996, inter alia), whereas in the second it is resolved at a purely semantic level of representation (Dalrymple et al. 1991, Hardt 1992, Kehler 1993a, Hardt 1999, inter alia). This question has remained a point of contention for a considerable time, largely because proponents in either camp have been able to offer seemingly definitive evidence to support their particular view.

In this chapter, I show that the apparently contradictory VP-ellipsis data actually exhibit a systematicity, particularly with respect to the

type of coherence relation that is operative between the source and target clauses. Specifically, we will see that the data primarily support syntactic accounts when a Resemblance relation is operative, whereas they support semantic accounts when a Cause-Effect relation is operative. (A discussion of the data involving Contiguity will be postponed until Section 3.3.3.) I will argue for a theory which makes a variety of predictions based on two requirements: (i) that the referent of VP-ellipsis be anaphorically resolved, and (ii) that there be syntactic parallelism at the VP level (and in some cases, below the VP level) in light of the inference processes that underlie the establishment of Resemblance relations. The distribution of the data will be shown to result from the combination of these independently-motivated aspects of discourse interpretation.

## 3.1 An Abstract Characterization of the Approaches

Particular syntactic approaches (and likewise, semantic approaches) to VP-ellipsis interpretation vary in their details, of course. However, in order to create a more coherent foundation for the remaining sections of this chapter, I begin by presenting abstract characterizations of these two types of approach to use as bases of comparison. I will then return to discuss how certain instances of past work deviate from these characterizations in Section 3.5.

### 3.1.1 Syntactic Approaches

Syntactic accounts posit that VP-ellipsis is resolved at a level of syntactic representation, imposing the requirement that a suitable syntactic structure for the source be available as an antecedent for the ellipsis. There is substantial evidence to support this view. For instance, Lappin (1993b) discusses example (54).

(54) * The lawyer defended $\text{Bill}_i$, and $\text{he}_i$ did too. [ defend $\text{Bill}_i$ ]

Syntactic accounts predict that (54) is unacceptable assuming the indicated coreference between *he* and *Bill*. This prediction results from the fact that if the elided syntactic structure were present in the target clause, the pronoun *he* would c-command the full NP *Bill* with which it corefers, resulting in a *Condition C* (Chomsky 1981) violation. A semantic approach does not predict such a violation, since Condition C is a constraint on syntactic representations.

In a similar vein, Kitagawa (1991) offers example (55).

(55) * $\text{John}_i$ blamed $\text{himself}_i$, and Bill did too. [ blamed $\text{him}_i$ ]

Syntactic approaches predict the unacceptability of sentence (54) for similar reasons, except in this case *Condition A* is violated. That is, assuming the strict reading in which *Bill* blamed *John* as indicated by

the indices, the reflexive that would appear in the target if the syntactic structure for the elided VP was present would lack the c-commanding antecedent it requires.

Finally, Haïk (1987) gives example (56).

(56) * John read everything which Bill believes the claim that he did. [ read $\phi$ ] (from Haïk (1987))

In this case, syntactic approaches predict unacceptability on the basis of a *subjacency* constraint violation, since the gap that would be in the target clause if the VP were present would be two bounding nodes away from the NP on which it is dependent.

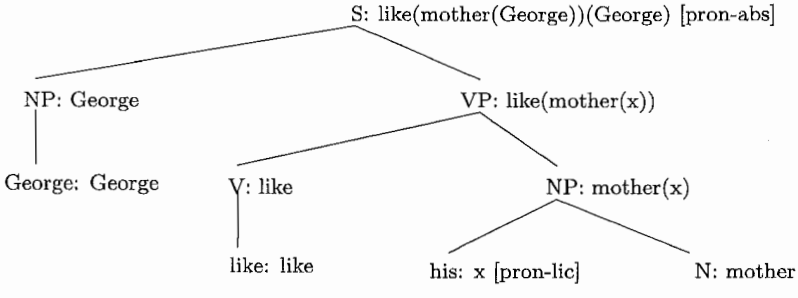


FIGURE 1 Syntactic and Semantic Representations for *George likes his mother.*

One manner in which to characterize the process of recovering the meaning of the elided VP in a syntactic account is by the reconstruction

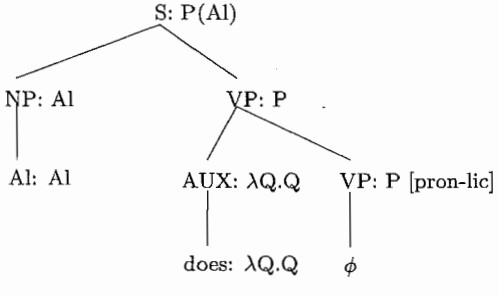


FIGURE 2 Syntactic and Semantic Representations for *Al does [too].*

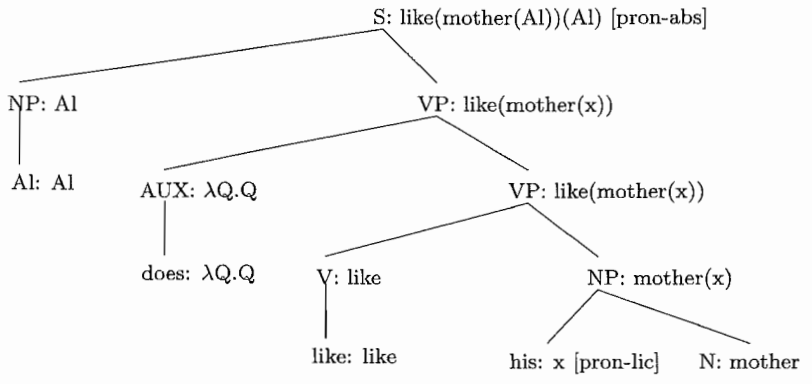


FIGURE 3 Syntactic and Semantic Representations for *Al does [too]* after reconstruction.

of a copy of the syntactic structure of the source VP at the site of the missing target VP. Syntactic representations for the source and target clauses of sentence (53) are shown in Figures 1 and 2 respectively. The symbol $\phi$ at the leaf node of the VP in Figure 2 indicates that the VP constituent is empty. The result of performing this reconstruction for sentence (53) is shown in Figure 3. In a syntactic reconstruction account, the semantics of the target clause is recovered by computing it independently after the VP has been reconstructed from the source.

A brief aside is in order concerning the formalism I use here. The formalism has its roots in the version of Categorial Semantics described by Pereira (1990), and employs a standard Montagovian pairing between rules combining constituents in the syntax and rules for combining the meanings of those constituents in the semantics. Syntactic analyses are used to control the semantic derivation in a proof system, along with the constraints imposed by types in the semantic representations. In Pereira's paper, derivation trees are shown with semantic representations and their corresponding types, whereas I will omit the types in my depictions. I nonetheless follow Pereira in using a curried semantic notation in the figures when appropriate; thus a sentence such as *Bill likes Al* will be represented as $likes(Al)(Bill)$, in accordance with the standard compositional analysis by which the meaning of a transitive verb is first combined with the meaning of its object and then its subject.

The leaf nodes of derivation trees are labeled by *assumptions* which may or may not need to be subsequently *discharged* for a valid derivation. One type of assumption is introduced by lexical items (such as by *George*, *likes*, and *mother* in Figure 1); these assumptions are not

discharged. Other assumptions that express a dependency between one form and another, such as those introduced by pronouns and traces, need to be licensed and subsequently discharged by a suitable construction or assignment. The process of introducing and discharging assumptions for pronominal binding is encoded by two rules: pronoun licensing [**pron-lic**] and pronoun abstraction [**pron-abs**]. The appearance of an unbound pronoun introduces an assumption (using [**pron-lic**]) that is subsequently discharged when the pronoun is bound by an antecedent entity, a trace, or a quantifier (using [**pron-abs**]). Similarly, trace assumptions are licensed by the occurrence of a trace in the syntactic representation using the rule [**trace-lic**], and are discharged by a syntactic construction taking a constituent containing a trace using the rule [**trace-abs**]. Additional details are not necessary for my purposes, but further discussion of these topics may be found in Pereira (1990). I take care to note, however, that while this formalism is convenient for the analyses presented in this chapter and the next, these analyses do not depend on it in any crucial way.

As I indicated earlier, the details of actual syntactic accounts to VP-ellipsis resolution differ in various respects. While some theories posit the type of reconstruction mechanism that I illustrated (Williams 1977, Kitagawa 1991, Lappin 1993b, Hestvik 1995, Fiengo and May 1994), others view VP-ellipsis as a process of deletion under suitable conditions (e.g., Sag (1976)). Furthermore, some theories operate purely at the level of surface syntax (Lappin 1993b, Lappin 1996), whereas others operate at some level of syntactic logical form (Sag 1976, Williams 1977, Kitagawa 1991, Hestvik 1995, Fiengo and May 1994, inter alia). Common to all of these analyses, however, is the requirement that a suitable syntactic source representation be available for reconstruction in the target.

### 3.1.2 Semantic Approaches

In contrast to syntactic accounts, semantic accounts posit that VP-ellipsis is resolved at a purely semantic level of representation. After witnessing the syntactic effects evident in examples (54–56), one might wonder why we should even consider such an analysis. The reason is that there is equally impressive evidence that the semantic approach is the correct one.

For instance, Dalrymple (1991) cites example (57).

(57) In March, four fireworks manufacturers asked that the decision be reversed, and on Monday the ICC did. [ reverse the decision ]
(from text of Rosenthal (1988))

Syntactic approaches predict example (57) to be unacceptable. A syntactic structure for the VP required for resolution, *reverse the decision*, is not available because the source clause is passivized. On the other hand, this example is perfectly compatible with semantic analyses, since they operate at a level of representation at which voice distinctions are lost.

Likewise, Hardt (1993) discusses example (58).

(58) Harry used to be a great speaker, but he can't anymore, because he lost his voice. [ speak ]

Sentence (58) is acceptable even though the referent is evoked by a nominalization. Again, this case is problematic for a syntactic approach, as it would require that one commit to the (dubious) claim that a syntactic verb phrase headed by *speak* underlies the syntactic representation for *speaker*. Again, there is no conflict here for a purely semantic approach, as long as the semantics of the nominalized verb makes its underlying eventuality sufficiently salient for subsequent reference. (I will have more to say about this issue in Section 3.3.1.)

Finally, Dalrymple (1991) offers sentence (59).

(59) I expected $\text{Bill}_i$ to win even when $\text{he}_i$ didn't. [expect $\text{Bill}_i$ to win]

This example is perfectly felicitous despite the fact that Condition C predicts unacceptability under a syntactic account. Again, no such constraints apply at a purely semantic level of representation. Note that example (59) is directly at odds with example (54), in which a Condition C violation presumably causes the sentence to be unacceptable.

For my purposes here, I will use the representation and resolution method of Dalrymple et al. (1991, henceforth DSP) as an exemplar for semantic analyses of VP-ellipsis. (See also Shieber et al. (1996).) Again, my analysis does not depend on this choice.[5] The semantic representation for the source clause of sentence (53) is given in (60).

(60) $likes(\underline{George}, mother\_of(George))$

Underlining a term signifies that it is a *primary occurrence*, meaning that it is parallel to a term that is overt in the target. It must therefore be abstracted over in generating a valid solution. The representation of the target clause contains an uninstantiated relation that applies over the overt target material, as shown in (61).

(61) $P(Al)$

In representation (61), $P$ stands proxy for the missing property that corresponds, in this case, to the missing VP in the syntax of the target.

[5]Several issues regarding DSP's analysis will be discussed in Section 3.6.

*George* and *Al* are pragmatically determined to be parallel elements in these equations. We solve for $P$ by computing the relation(s) that when applied to *George* results in representation (60); that is, one solves for $P$ in equation (62).

(62) $P(George) = likes(\underline{George}, mother\_of(George))$

A possible solution to this equation is given in (63), which when applied to *Al* in the target representation results in meaning (64).

(63) $P = \lambda x.like(x, mother\_of(George))$

(64) $likes(Al, mother\_of(George))$

This result represents the strict reading for the target. Another possible solution to equation (62) is shown in (65).

(65) $P = \lambda x.like(x, mother\_of(x))$

This case results in meaning (66) when applied to *Al*, which corresponds to the sloppy reading for the target.

(66) $likes(Al, mother\_of(Al))$

Some semantic accounts derive a property by abstracting over the propositional meaning of the entire source clause (Dalrymple et al. 1991, Kehler 1993a, inter alia), whereas others recover a property resulting solely from the meaning of a VP (Hardt 1992, Hardt 1999). Common to these approaches, however, is the characteristic that they require only a suitable semantic source representation for resolution.

In sum, there are sets of data to motivate both syntactic and semantic approaches to VP-ellipsis resolution, which therefore appear to be contradictory when taken together. Upon closer inspection, however, there is a difference between examples (54–55) on the one hand and examples (57–59) on the other. (I will set aside example (56) for the time being.) Examples (54–55) are examples of the Resemblance relation Parallel, as indicated by the use of *and...too*. On the other hand, examples (57–59) are instances of Cause-Effect relations; in particular, Result, Violated Expectation, and Denial of Preventer, respectively. This pattern offers a preliminary suggestion that coherence establishment may play a role in accounting for these data. I flesh out the details of an analysis that takes coherence establishment into account in the following section, and then evaluate it with respect to a more comprehensive set of data in Section 3.3.

## 3.2 VP-Ellipsis and Coherence

At this point, one might be inclined to view the implicit gridlock between syntactic and semantic approaches to VP-ellipsis as a manifestation of

Maslow's hammer-and-nail maxim that I discussed in Chapter 1. Based on my preliminary survey, it appears that syntactic or semantic tools alone are not likely to be sufficient for constructing an empirically adequate theory. As I will discuss in Section 3.5, past attempts at a theory that operates solely within one of these modules of language processing have either failed to address the full breadth of data, or have required significant extensions to the set of theoretical constructs beyond those which are intuitively plausible and independently motivated.

It does seem apparent nonetheless that there are both syntactic and semantic factors that need to be accounted for in a theory of VP-ellipsis. As I have already alluded to, I will offer an account in which two sets of properties interact: the syntactic and semantic properties of VP-ellipsis, and the properties of the establishment mechanisms that underlie the recognition of different classes of coherence relations. I consider the first part in Section 3.2.1 and the second in Section 3.2.2.

### 3.2.1 Syntactic and Semantic Properties of VP-Ellipsis

The first question we need to ask is what properties VP-ellipsis appears to have when viewed independently of the behaviors for which we are attempting to account. Syntactic reconstruction approaches typically rely on the premise that VP-ellipsis leaves behind a stranded auxiliary that governs an empty node at which a VP could be reconstructed. Semantic theories, on the other hand, typically imply that VP-ellipsis has an anaphoric quality which would trigger a search for a purely semantic representation in the discourse context. The evidence suggests that VP-ellipsis has both of these properties.

First, there is no doubt that the word "does" in examples like (53) is an auxiliary and not the main verb form found in various superficially similar forms of event reference, including *do it*, *do that*, and *do so* anaphora (Halliday and Hasan 1976, inter alia) . First, sentences (67a-d) show that VP-ellipsis is possible with other auxiliaries, but event referential forms are not.

(67) a. George announced his victory, and Al did too. (auxiliary *did*)
b. George announced his victory, and Al did it too. (main verb *did*)
c. George will announce his victory, and Al will too.
d. # George will announce his victory, and Al will it too.

Second, the main verb *do* requires a non-stative direct object. This constraint applies for pronominal event referential forms as shown by examples (68a) and (68c), but not in cases of VP-ellipsis, as shown by examples (68b) and (68d).

(68) a. Al likes the internet, and George does too.
b. # Al likes the internet, and George does it too.
c. Al wants to be president, and George does too.
d. # Al wants to be president, and George does it too.

Thus, in accordance with syntactic approaches, it is reasonable to analyze the syntax of VP-ellipsis as consisting of an auxiliary that commands an empty verb phrase constituent.

In accordance with semantic approaches, however, the evidence also indicates that VP-ellipsis is anaphoric. This can be seen from the manner in which it patterns with other types of anaphora, such as pronouns. For instance, as described by Lakoff (1968) and Jackendoff (1972), VP-ellipsis and pronouns may be cataphoric in similar circumstances; consider sentences (69a-d).

(69) a. # Al will $\phi$, if George will claim victory.
(where $\phi$ = *claim victory*)
b. If George will claim victory, Al will $\phi$.
c. If George will $\phi$, Al will claim victory.
d. Al will claim victory if George will $\phi$.

Whereas sentence (69a) is unacceptable under the reading in which Al may claim victory, sentence (69c) is acceptable with this interpretation. Examples (70a-d) show that this situation is analogous to the one for pronouns that refer to entities.

(70) a. # He will make a fool of himself, if Al claims victory.
(where *He* = *Al*)
b. If Al claims victory, he will make a fool of himself.
c. If he claims victory, Al will make a fool of himself.
d. Al will make a fool of himself, if he claims victory.

Cataphora is allowable when the pronoun is embedded as in sentence (70c), as it is for VP-ellipsis in sentence (69c). Likewise, cataphora is not allowable when the pronoun is not embedded as in sentence (70a), as is the case for VP-ellipsis in sentence (69a). This pattern caused Lakoff (1968, p. 332) to state that "clearly these [the ellipsis and pronoun cases] are related phenomena, and [(69a)] should be blocked by the same constraint that blocks [(70a)]". Jackendoff (1972, p. 268) similarly addressed both examples using a precede-and-command rule that blocks coreference.

Another similarity between VP-ellipsis and pronouns is that each can access referents that are evoked from more than one clause back. (We will see in Chapter 4 how this contrasts with other types of ellipsis, such

as gapping.) Such reference is not uncommon for VP-ellipsis. Hardt (1990), for instance, reports that five percent of the examples in the Brown corpus (Francis 1964) have an antecedent that is at least two sentences back in the discourse. He gives example (71).

(71) The thought came back, the one nagging at him these past four days. He tried to *stifle it.* But the words were forming. He knew he *couldn't.*

Pronominal reference is also well known to allow local but non-immediate referents. This property is likewise demonstrated by passage (71), in which the referent of the subject pronoun *he* in the final sentence is last mentioned two sentences prior. These similarities therefore lead us to the conclusion that VP-ellipsis is anaphoric. The idea that VP-ellipsis is a "proform" dates at least as far back as Schachter (1978), and has been adopted by a variety of researchers since (Chao 1987, Hardt 1992, Lobeck 1999).

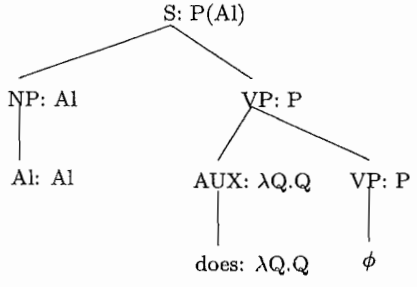


FIGURE 4 Syntactic and Semantic Representations for *Al does.*

To summarize, Figure 4 illustrates the syntactic and semantic properties of the elliptical clause *Al does.* The $\phi$ indicates the empty constituent in the syntax, and $P$ represents the anaphoric form in the semantics. As it stands, the clause-level semantics for this sentence is incomplete, and admits of two possible ways of recovering the remaining material. First, the syntactic VP could be copied from the source with its corresponding semantics, from which the semantics for the newly completed structure in the target can be derived. This option is advocated by the syntactic reconstruction approach. In this case, the anaphoric expression is constrained to have a semantics that is consistent with the copied constituent. Alternatively, the anaphoric expression could be resolved by purely semantic means. This option is advocated by the purely semantic

approach; in this case the sentence-level semantics is recovered without copying any syntactic material. I discuss the scenarios in which either or both of these resolution processes take place in the next section.

### 3.2.2 Interaction with Coherence Establishment

Now that we have pinned down a few basic syntactic and semantic properties of VP-ellipsis, we are ready to consider how these properties interact with various aspects of discourse interpretation. I begin by stepping back for a moment and considering the purpose that ellipsis serves in discourse production and understanding, as the results will shed light on how we would expect it to interact with the processes underlying the establishment of coherence.

One question that immediately comes to mind is why a speaker would ever choose to employ ellipsis. The answer is not at all obvious. After all, by eliding material a speaker is choosing to use an underspecified expression that requires that the hearer invoke a 'search' process to recover the information that is left missing, in lieu of using a fully specified expression. One might expect that such a choice would increase the interpretation burden on the hearer, and thus explicitness would always be preferred.

In reality, the felicitous use of ellipsis appears to have the opposite effect, in which the computational burden on the hearer is reduced rather than increased. Indeed, a failure to employ ellipsis when it is licensed can sometimes even seem redundant, if not outright confusing. Why would this be so? A plausible hypothesis is that ellipsis is a signal to the hearer that the missing material has already been computed and is readily recoverable from the discourse context. In situations in which this is so, avoiding the need to recompute this information could offset the cost of accessing a suitably salient referent. This logic has often been cited as a reason why natural languages have pronouns, and indeed the repetition of a full lexical noun phrase or proper name in a context that licenses a pronoun can be misleading and lead to unwanted inferences. (See also Gordon et al. (1993) for psycholinguistic evidence that repeated names lead to longer processing times.) We would expect the same to be true of VP-ellipsis in light of the anaphoric properties it shares with pronouns. Of course, if the elided information that is needed for further processing is not sufficiently precomputed and readily recoverable, the hearer will most likely be misled, and the increased computational burden that results will render the passage marginal or even uninterpretable.[6]

So what is the nature of the missing information that needs to be re-

[6]I will discuss similar effects for pronoun interpretation in Chapter 6.

covered for VP-ellipsis? My analogy with pronoun interpretation would suggest that it is simply the identity of the referent – in this case, the representation of an eventuality – that needs to be recovered. This recovery is accomplished by both syntactic and semantic theories, albeit again by different methods: syntactic theories access the meaning by copying syntactic material from which the semantic representation can be derived, and semantic theories recover it through a form of anaphora resolution. Given the fact that we already know that VP-ellipsis behaves anaphorically, however, one has to wonder why syntactic reconstruction would ever be required to access this meaning. In light of its anaphoric behavior, we would therefore expect *a priori* that semantic accounts provide the correct characterization of VP-ellipsis resolution.

However, we saw in Chapter 2 that there is more to discourse understanding than recovering the semantics of each sentence. In particular, the coherence of the passage must also be established, which requires that two steps be taken: first the correct arguments to a coherence relation must be identified, and then the constraints associated with that relation must be applied to them. The identity of any elided arguments to the relation must therefore also be recoverable for VP-ellipsis to be felicitous. From this emerges a purpose for a syntactic reconstruction process: not as a means for recovering the meaning of an elided clause, but as a means for recovering arguments to coherence relations that are unavailable due to their having been elided.

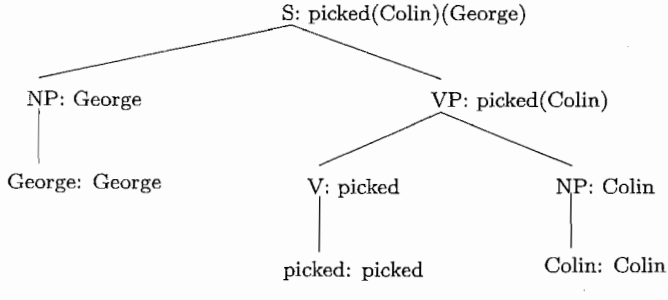


FIGURE 5 Syntactic and Semantic Representations for *George picked Colin.*

The extent to which coherence establishment will require access to elided constituents, and thus require reconstruction, depends on the needs of the inference processes used to establish the operative relation. We saw in Chapter 2 that the arguments to Cause-Effect relations are simply the propositional semantics of each utterance. Assuming that

the meaning of the elided VP is successfully resolved anaphorically, the proposition-level semantics will be complete, and thus reconstruction of syntactic material at lower-level nodes will not be necessary. In the case of Resemblance relations, in contrast, the more complex process of identifying parallel arguments and their relative pairings attempts to utilize the syntactic structure of the utterances. As a result, in addition to signaling that the semantics of the missing VP is recoverable, elision in this context signals that the syntactic parallelisms necessary for computing coherence are also recoverable. As I will discuss in greater detail below, this signal, in a situation in which a relation requires access to a VP-level argument, will force a requirement that the source VP be parallel to the target VP. Furthermore, if the relation also requires access to arguments below the VP level, a reconstruction process to recover them will be triggered. This requirement will ultimately explain why sentences such as (54–56) display syntactic effects, despite the fact that the meaning of VP-ellipsis is resolved anaphorically.

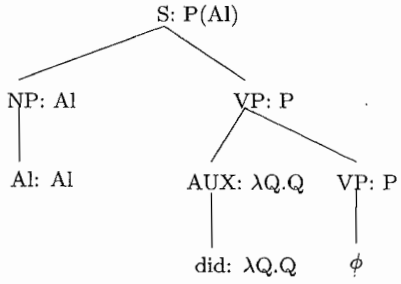


FIGURE 6 Syntactic and Semantic Representations for *Al did.*

Let us illustrate this process for both types of relation on a common example, involving the relatively simple source and target clause fragments shown in (72).

(72) George picked Colin...Al did...

The syntactic and semantic representations for the clauses are shown in Figures 5 and 6 respectively (the latter is repeated from Figure 4).

I first consider the case in which these clauses are related by a Cause-Effect relation, as in (73).

(73) George picked Colin because Al did.

The anaphoricity of the VP-ellipsis in the second clause triggers a search for a referent, in this case, the representation evoked by the first clause.

I will again use the Dalrymple et al. approach to resolution. Reading the semantic representation of the source clause from the top of Figure 5 (and reverting to a non-Curried notation), we have:

(74) $picked(\underline{George}, Colin)$

Likewise, reading the representation the target clause from the top of Figure 6 we have:

(75) $P(Al)$

*George* and *Al* are parallel elements. We solve for $P$ by computing the relation that when applied to *George* results in representation (74); that is, we solve for $P$ in equation (76).

(76) $P(George) = picked(\underline{George}, Colin)$

The solution to this equation is given in (77), which when applied to *Al* in the target representation results in meaning (78).

(77) $P = \lambda x.picked(x, Colin)$

(78) $picked(Al, Colin)$

After resolution, the syntactic and semantic representations for the target are as shown in Figure 7, in which the semantic representation of the sentence node has been updated.

Next the Explanation relation, indicated by *because*, is established.[7] The definition of Explanation is repeated below from Chapter 2.

**Explanation:** Infer $P$ from the assertion of $S_1$ and $Q$ from the assertion of $S_2$, where normally $Q \rightarrow P$.

The first step is to identify the arguments to the relation, which are simply the sentential-level propositions of each sentence. Reading from the top-level sentence node of the representations, we have *picked(George, Colin)* and *picked(Al, Colin)* as the arguments $P$ and $Q$ to the inference procedure respectively. At this point the constraints of the relation can be applied to these arguments.

The key property of this process as it pertains to the constraints on VP-ellipsis is that no part of the syntactic information in the target had to be reconstructed for the coherence establishment process to proceed. The anaphoric resolution of the meaning of the VP, and the full propositional semantics for the target sentence that results, provides all the information necessary to compute coherence. While the source and target sentences in this example happen to be syntactically parallel,

[7] In reality, the processes of anaphora resolution and coherence establishment are likely intertwined, as I suggest they are for pronoun interpretation in Chapter 6. This question does not affect us here in any crucial way, however.

this was not necessary. Even if the source clause was passivized, for instance, the semantically-driven resolution procedure would still succeed since the semantic relations would be the same.

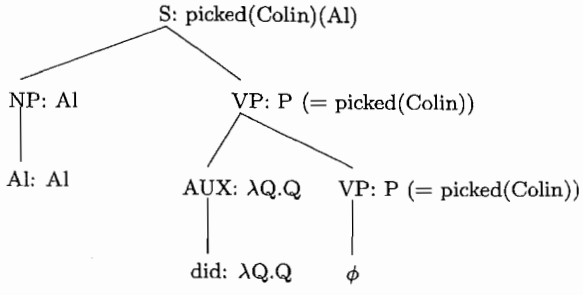


FIGURE 7 Syntactic and Semantic Representations for *Al did* after anaphora resolution.

I now consider the example again, except where the clauses in (72) participate in a Resemblance relation, such as example (79).

(79) George picked Colin and Al did too.

First, the resolution of the anaphoric form occurs just as it did for example (73), resulting again in the representations in Figure 7. At this stage, there is no reconstruction of elided material. The representations in Figures 5 and 7 then serve as input to the coherence establishment procedure. In this case the Resemblance relation *Parallel* holds; the definition of which is repeated from Chapter 2 below.

**Parallel:** Infer $p(a_1, a_2, ...)$ from the assertion of $S_1$ and $p(b_1, b_2, ...)$ from the assertion of $S_2$, where for some property vector $q$, $q_i(a_i)$ and $q_i(b_i)$ for all $i$.

The first step in establishing this relation is to identify its arguments: that is, the parallel relations from which a common $p$ can be inferred ($p_1$ and $p_2$ respectively), and the $a_i$'s and $b_i$'s over which those relations predicate. In the source clause, the process identifies $p_1$ as *pick*, $a_1$ as *George*, and $a_2$ as *Colin*. In the target clause, it finds the syntactically-parallel subject *Al* as the argument $b_1$, but hits upon an empty VP node, and thus does not find the corresponding structure for locating the other parallel arguments.

Two types of constraints become manifest at this point. The first pertains to the VP level, in particular, the constraint is that the source

and target VPs be parallel. We would in fact expect VP-ellipsis to enforce such a requirement in the context of a Resemblance relation. My evidence again lies in the analogy to the situation for pronoun interpretation, for which the effect of parallelism has been well documented (Sidner 1983, Kameyama 1986, Kehler 1995, Kehler 1997, see also Chapter 6). Consider the case of Resemblance in example (80), which was discussed briefly in Chapter 1.

(80) Margaret Thatcher admires Hillary Clinton, and George W. Bush absolutely worships her.

Theories of pronoun interpretation that incorporate a preference for referents evoked from subject position suggest *Margaret Thatcher* as the referent for the pronoun, as do semantically-based preferences, given our knowledge about the political orientations of the people involved.[8] Nonetheless, assuming that the pronoun remains deaccented, hearers consistently interpret the pronoun as referring to *Hillary Clinton*, and in fact will not even 'backtrack' to force a reinterpretation. Although some researchers have cast this as a preference for grammatical role parallelism in any context (Kameyama 1986, inter alia), in Chapter 6 I argue that the preference is strongly associated with Resemblance relations, in which the default assignment of a pronoun in the second clause is its parallel element in the first (see also Kehler (1995)). Consider example (81).

(81) Colin Powell defied Dick Cheney, and George W. Bush punished him.

Here the preferred referent of the pronoun depends directly on the coherence relation that is inferred. If a Parallel relation is inferred for passage (81), then, as in (80), the pronoun must be interpreted to refer its parallel element, in this case Cheney. However, if a Cause-Effect relation is inferred, then Powell is the preferred referent for the pronoun. This difference exists despite the fact that the syntactic conditions remain constant, and that Cheney is the grammatically parallel referent in both cases.

Thus, given that VP-ellipsis patterns with pronominal reference in terms of its anaphoric behavior, we would expect to see a similarly strong disposition for it to refer to the meaning of a parallel VP when a Resemblance relation is operative. We would likewise expect the effect of parallelism to be so strong that a failure to identify a suitably parallel antecedent will not be enough to force a reinterpretation.

However, merely recovering the meaning of the parallel VP will not

[8]Thatcher and Bush are political conservatives, whereas Clinton is liberal.

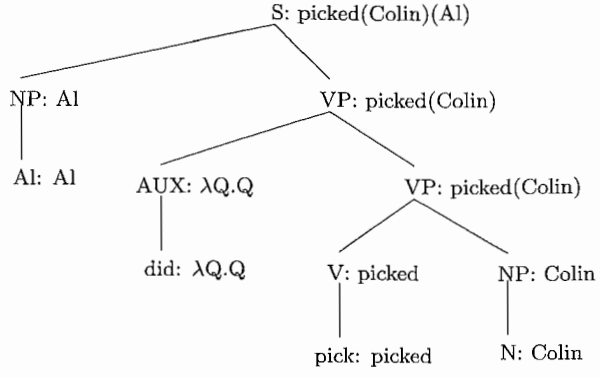


FIGURE 8 Syntactic and Semantic Representations for *Al did [too]* after reconstruction.

be enough if the coherence relation seeks to identify maximal parallelism, which is a reasonable property to attribute to the *and...too* construction used in example (79). Thus the second constraint becomes manifest, in which the need to recover parallel elements below the VP level triggers the reconstruction of the missing syntactic material at the site of the elided node. There is a suitable syntactic source VP in this case, and so it is reconstructed as shown in Figure 8.

The parallel entities and relations can now be identified and retrieved. Applying the constraints to them successfully establishes the Parallel relation, along the lines of the procedure described in Chapter 2, in which $p_1$ and $p_2$ (and thus $p$) are *picked*, $a_1$ and $a_2$ are George and Al respectively, and $b_1$ and $b_2$ are both Colin. Note that while the reconstructed arguments have the same denotation as they do in the source, this will not always be the case. For instance, if the source clause had contained a pronoun that was coreferential with the subject, the target clause may have received either a strict or a sloppy interpretation. In the sloppy case, the denotations for the pronoun would be different in the source and target representations.

Note that in cases which require that the source syntactic representation be reconstructed, the anaphora resolution process is effectively constrained to a referent which corresponds to a source syntactic structure that has an appropriate form. If the source clause of example (79) had been passivized (*# Colin was picked by George and Al did too*), reconstruction of the passive VP at the active VP site in the target would fail, thus eliminating that clause as a possible source for the ellipsis. In light of the need to establish coherence, the speaker should not have

elided the VP with this choice of antecedent, since the signals the ellipsis sends to the hearer regarding parallelism in this context are not met.

To sum, the facts regarding the syntactic and anaphoric properties of VP-ellipsis on the one hand, and the establishment of Cause-Effect and Resemblance relations on the other, combine to yield a theory of VP-ellipsis resolution that makes several predictions. First, VP-ellipsis data in Cause-Effect relations should accord with the predictions of the prototypical semantic analysis, since there is no process that would cause syntactic reconstruction in this case. Second, it predicts that VP-ellipsis data in Resemblance relations will at a minimum require parallelism at the VP level, because the source and target VPs must be parallel elements. Third, it predicts that reconstruction will take place if there is any additional syntactic information below the VP that is required for establishing further parallelism between the source and target clauses. In the following sections, we will see how these facets of the analysis explain several patterns in the data.

## 3.3 Explaining the Data

There are essentially two types of data to consider to determine whether VP-ellipsis is resolved at the level of syntax or semantics. The first type includes examples for which a suitable semantic representation for the source is available for resolution, but for which there is a mismatch of surface-syntactic form between the source and target syntactic representations. In such cases, syntactic approaches are supported if ellipsis shows a sensitivity to syntactic form, as this difference should not be manifest at the level of semantic representation. Conversely, an apparent lack of such sensitivity would provide evidence for semantic approaches. The second type of data include examples for which there is no syntactic form mismatch, but for which reconstruction of the source VP would result in a syntactic constraint violation in the target. In these cases, if the source and target pair is unacceptable because of such a constraint violation, then the syntactic theories are supported; again a lack of such an effect supports a semantic analysis.

In this section, I examine data involving the different types of coherence relation within five types of contexts for VP-ellipsis: two involving a mismatch of syntactic form (specifically, *voice alternation* and *nominalized antecedents*), and three involving syntactic constraint violations (specifically, *Condition A, B, and C violations).* (Additional ellipsis contexts involving traces, such as those capable of giving rise to subjacency violations, will be discussed in Section 3.4.) We will see that the data are consistent with the analysis just given.

### 3.3.1 Ellipsis and Cause-Effect Relations

The analysis proposed here predicts that the behavior of VP-ellipsis will correspond to semantic analyses when a Cause-Effect relation is operative, since there is no mechanism that would invoke the reconstruction of missing syntactic material. The data appears to support this prediction. We can begin by considering cases of nonparallel syntactic form, such as examples (82)–(84).

(82) In March, four fireworks manufacturers asked that the decision be reversed, and on Monday the ICC did. [reverse the decision]
(from text of Rosenthal (1988), cited in Dalrymple (1991))

(83) This problem was to have been looked into, but obviously nobody did. [look into the problem]
(Vincent Della Pietra, in conversation)

(84) Of course this theory could be expressed using SDRSs, but for the sake of simplicity we have chosen not to. [express this theory using SDRSs]
(from text of Lascarides and Asher (1993))

In each of these cases, a target clause in the active voice receives its interpretation from a source clause in the passive voice. Conversely, examples (85–86) have a target clause in the passive voice that receives its interpretation from a source in the active voice.

(85) Actually I have implemented it [= a computer system] with a manager, but it doesn't have to be. [implemented with a manager]
(Steven Ketchpel, in conversation)

(86) Just to set the record straight, Steve asked me to send the set by courier through my company insured, and it was. [sent by courier through my company insured]
(posting on the Internet)

These examples are problematic for syntactic analyses, since the VP needed for reconstruction in the target is not available in the source clause. On the other hand, semantic analyses predict these cases to be acceptable, since voice distinctions are presumably lost at the level of semantic representation. What all of the above cases have in common is that they are instances of Cause-Effect relations. In particular, a Result relation is operative in examples (82) and (86), in which *and* can be paraphrased roughly as *and as a result* (see Section 2.3). A Violated Expectation relation is likewise operative in examples (83–85), in which *but* can be paraphrased (roughly) as *but nonetheless*.

The second type of syntactic form mismatch I consider occurs when the source semantic representation is evoked by a nominalized antecedent. Examples include (87) and (88).

(87) This letter deserves a response, but before you do, .... [respond] (Gregory Ward, personal communication)

(88) Today there is little or no OFFICIAL harassment of lesbians and gays by the national government, although autonomous governments might. [harass lesbians and gays] (Hardt 1993)

These data are also problematic for syntactic approaches, since there is no verb phrase representation for the source available for reconstruction.

A semantic approach predicts the acceptability of such examples as long as the nominalization makes its underlying event representation available for anaphoric reference. This is a critical issue, because it is relatively easy to find cases of VP-ellipsis with nominalized antecedents that are infelicitous regardless of the operative coherence relation. Some informants have even considered examples (87) and (88) to be marginal as compared to other types of mismatches. There are two possible reasons for this. First, it could be that semantic approaches to VP-ellipsis are wrong, and there is some other reason for the (at least marginal) acceptability of examples like (87) and (88). Second, these judgments could result from the low level of salience generally associated with the semantic representations of events that are evoked by nominalizations. That is, if VP-ellipsis is anaphoric we would actually expect some examples to be more degraded than others, corresponding to differences in the level of activation of the referent in the discourse context.

To distinguish these two possibilities, we can simply consider examples analogous to (87) and (88) in which the ellipses are replaced with event referential expressions that are uncontroversially considered to be resolved anaphorically. *Do it* anaphora is such a form; consider (89) and (90).

(89) This letter deserves a response, but before you do it, .... [respond]

(90) Today there is little or no OFFICIAL harassment of lesbians and gays by the national government, although autonomous governments might do it. [harass lesbians and gays]

I and several informants do not find any marked difference between the acceptability of examples (89) and (90) on the one hand and their counterpart examples (87) and (88) on the other. As such, any marginality attributable to (87) and (88) appears not to be due to a mismatch of syntactic form. Importantly, most speakers find any degradation in the acceptability of these examples to be notably distinct from the unac-

ceptability of similar sentences which are discussed in Section 3.3.2.

I now consider the second class of examples, in which there is no mismatch of syntactic form. Instead, reconstruction is structurally possible in these cases but would result in a syntactic constraint violation in the target. Again, syntactic approaches predict that such violations would occur, whereas semantic approaches do not.

I first consider cases in which a syntactic approach would predict a Condition A violation, such as sentences (91) and (92), adapted from Dalrymple (1991).

(91) $\text{Bill}_i$ defended $\text{himself}_i$ against the accusations because his $\text{lawyer}_j$ couldn't. [defend $\text{himself}_i$]

(92) $\text{John}_i$ voted for $\text{himself}_i$ even though no one $\text{else}_j$ did. [vote for $\text{himself}_i$]

These examples are perfectly felicitous, thereby supporting a semantic account. Cause-Effect relations are operative in these cases, particularly Explanation and Denial of Preventer respectively.

Next I consider potential Condition B violations, such as in examples (93) and (94).

(93) $\text{John's}_j$ mother introduced $\text{him}_j$ to everyone because $\text{he}_j$ wouldn't. [introduce $\text{him}_j$ to everyone]

(94) $\text{John}_i$'s lawyer defended $\text{him}_i$ because $\text{he}_i$ couldn't. [defend $\text{him}_i$]

Again, the expected violation appears to be absent when a Cause-Effect relation is operative (Explanation in each case).

Finally, Condition C effects are likewise not observed. Consider (95) and (96), in which Denial of Preventer and Explanation are operative respectively.

(95) I expected $\text{Bill}_i$ to win even when $\text{he}_i$ didn't. [expect $\text{Bill}_i$ to win]

(96) The lawyer defended $\text{Bill}_j$ against the accusations because $\text{he}_j$ couldn't. [defend $\text{Bill}_j$ against the accusations]

Once again, these examples are perfectly felicitous.

In sum, after considering five different contexts for VP-ellipsis, the data strongly support a semantic approach to ellipsis interpretation. Each case is acceptable despite the fact that a syntactic approach predicts unacceptability.

### 3.3.2 Ellipsis and Resemblance Relations

I now consider examples of VP-ellipsis in which Resemblance relations are operative. As in the last section, I begin by examining cases in which there is a mismatch of syntactic form between the source and target VPs.

First, unlike the acceptable cases of voice alternation discussed in Section 3.3.1, similar examples in Resemblance relations are unacceptable.

(97) # This problem was looked into by John, and Bob did too. [look into the problem]

(98) # This theory was expressed using SDRSs by Smith, and Jones did too. [express this theory using SDRSs]

(99) # John implemented the computer system with a manager, but it wasn't by Fred. [implemented with a manager]

These sentences, at a minimum, have an extremely jarring quality. (The question of whether they should be considered to be ungrammatical is a separate issue. I personally consider them to be grammatical, since their unacceptability results not from being disallowed by grammatical rules, but from their infelicity assuming a certain source representation. Thus, I use the "#" marking instead of "*".) In contrast to the data discussed in Section 3.3.1, we now see the effects predicted by a syntactic account of VP-ellipsis. In each of the above examples, a Resemblance relation is operative; Parallel in examples (97) and (98), and Contrast in example (99).

It is worth noting here that, as explained in Section 2.3, there may be more than one coherence relation that is operative in certain examples. In particular, merely using a connective that indicates a Cause-Effect relationship may not be enough to avert the recognition of a Parallel relation also; witness example (100).

(100) ? This problem was looked into by John, even though Bob did. [look into the problem]

The fact that the intended meaning of (100) does not change markedly when the adverb *too* is added, as in examples (101a-b), is an indication that a Parallel relation is still operative in (100).

(101) a. John looked into this problem even though Bob did too. [look into the problem]

b. ? This problem was looked into by John, even though Bob did too. [look into the problem]

However, we can change other aspects of example (100) that reduce the parallelism, and in so doing, make the VP-ellipsis acceptable. Consider example (102), in which the change in auxiliary also diminishes the parallelism found in example (100).

(102) This problem was looked into by John, even though Bob already had (# too). [looked into the problem]

Example (102) is considerably more acceptable than (100). This transition of judgments provides further evidence that the different degrees of acceptability are directly related to the coherence relationships that are operative between the clauses.

The other type of syntactic mismatch that I considered in Section 3.3.1 included examples in which the intended referent of a VP-ellipsis was evoked from a nominalized form. Examples (103) and (104), which are similar to (87) and (88) except that a Resemblance relation is operative, are far less acceptable that their Cause-Effect counterparts.

(103) # This letter provoked a response from Bush, and Clinton did too. [respond]

(104) # There is unofficial harassment of lesbians and gays by the American government, and the Canadian government does too. [harass lesbians and gays]

As before, we see an improvement in acceptability as the examples display less parallelism between the clauses. Example (105) is marginal; this case remains consistent with the adverbial *too.*

(105) ? This letter provoked a response from Bush because Clinton did (too). [respond]

Example (106) is much more acceptable, however, and yields a reading that is not consistent with *too.*

(106) This letter provoked a response from Bush because Clinton already had. (# too) [responded]

In sum, the data in which there is a mismatch of syntactic form between the source and target clauses in the context of a Resemblance relation supports the existence of at least one of the constraints associated with a syntactic approach to VP-ellipsis resolution, in particular, that the source VP be syntactically parallel to the target VP.

However, the data examined so far stops short of providing evidence that a syntactic reconstruction process is actually occurring. We can shed light on this question by considering the second class of examples, in which there is syntactic parallelism between the source and target, but in which reconstruction would result in a syntactic constraint violation. The theory proposed here predicts that reconstruction will be necessary for those coherence relations that attempt to establish parallelism at a level below the (missing) verb phrase node. While we would expect that Resemblance relations generally do, we will see that there is a subclass of constructions in this category that apparently do not.

In my earlier discussion, I suggested that the canonical case of Re-

semblance is the *and...too* construction. As one would expect, these cases appear to require reconstruction. Evidence for this comes from their potential to yield binding theory violations. First, I consider the effect of Condition A in sentences (107) and (108).

(107) * John$_i$ defended himself$_i$, and Bob$_j$ did too. [defend himself$_i$]

(108) * Fred$_i$ voted for himself$_i$, and Gary$_j$ did too. [vote for himself$_i$]

Almost all informants find it difficult to obtain a strict reading for these sentences, a result attributable to the fact that the pronoun in the source clause is reflexive. Condition A accounts for this fact since a reconstructed reflexive would require a locally-bound antecedent in the target, a restriction that permits only a sloppy reading in these examples.

Likewise, sentences (109) and (110) are also unacceptable.

(109) * John's$_j$ mother introduced him$_j$ to everyone, and he$_j$ did too. [introduce him$_j$ to everyone]

(110) * John$_i$'s lawyer defended him$_i$, and he$_i$ did too. [defend him$_i$]

This unacceptability is the expected result of Condition B, since the copied pronoun is not a reflexive, and thus cannot be bound to the subject pronoun in the target upon reconstruction.

Finally, I consider Condition C.

(111) * John defended Bob$_i$, and he$_i$ did too. [defended Bob$_i$]

(112) * Mary introduced John$_j$ to everyone, and he$_j$ did too. [introduced John$_j$ to everyone]

Examples (111) and (112) are similarly unacceptable, as would be expected if Condition C is operative, since the reconstructed proper name is prohibited from coreferring with the c-commanding pronominal in the target representation.

These judgments contrast sharply with those for similar examples in which Cause-Effect relations are operative as discussed in Section 3.3.1. There are cases, however, in which judgments for some informants appear to improve within suitable semantic contexts. Many informants, for instance, find sentences (113) and (114) to be at least marginally acceptable, although a majority of them report that they are not completely natural.

(113) ? The alleged murderer$_i$ defended himself$_i$, and his lawyer$_j$ did too. [defended himself$_i$]

(114) ? George W. Bush$_i$ voted for himself$_i$, and his campaign manager$_j$ did too. [voted for himself$_i$]

I actually find these examples to be markedly odd under a strict interpretation, but informant judgments do vary. For almost all speakers I have consulted, however, the Cause-Effect cases given in Section 3.3.1 are more acceptable under the strict interpretation than the Resemblance cases cited above.

Hestvik (1995) also notes that a strict reading for reflexives in examples similar to (113) and (114) may be only marginal rather than completely unacceptable, and proposes that although VP-ellipsis is always syntactically reconstructed, hearers can reinterpret it "off-line" as an anaphoric expression. Anaphoric interpretation would thus bypass binding theory constraints. This proposal is problematic since, as we have seen, independent evidence suggests that VP-ellipsis is already anaphoric; no reinterpretation as an anaphoric form is necessary. Furthermore, this explanation predicts that other types of violations, such as the cases of mismatching syntactic form I have discussed, should be able to be similarly mitigated by context. I have not found empirical support for that prediction.

Nonetheless, data such as those above suggest that any theory of VP-ellipsis must ultimately be equipped with a way of accounting for marginal data. In (113) and (114), I would speculate that the preference to identify maximal parallelism enforced by the *and...too* construction can be offset by other factors that come into play in the semantic interpretation of such examples, albeit at the cost of reduced felicity. In these cases, the strict readings can be seen to result from a process of backing off of the determination of maximal parallelism – in which parallelism is only established at the VP level, for instance – in order to accommodate the strong semantic bias toward the strict reading. Such tradeoffs could potentially be modeled in a cost-based inference system, such as the weighted abduction system of Hobbs et al. (1993) mentioned in Chapter 2, although I will not attempt to pursue the details of such an analysis here. In any case, while some degree of parallelism is still being enforced in these examples, the failure to establish it below the VP level has the effect of not requiring reconstruction, which leaves open the possibility of a strict reading.

As I noted earlier in this section, however, there are other constructions in which Resemblance relations are operative that appear not to require parallelism below the VP level. This is the case with comparatives and related constructions. (In what follows, I will use the term *comparative* to include temporal subordination constructions, such as those with adjuncts headed by the adverbials *before* or *after*.)

I first consider example (115), from Dalrymple (1991).

(115) John$_i$ defended himself$_i$ against the accusations better than his lawyer$_j$ did. [defend himself$_i$]

Example (115) is felicitous despite the fact that reconstruction would cause a Condition A violation, in contrast to sentences (107) and (108). Likewise, examples (116) and (117) do not display the expected Condition B and C violations respectively.

(116) John$_i$'s lawyer defended him$_i$ better than he$_i$ did. [defend him$_i$]

(117) Sue defended John$_i$ better than he$_i$ did. [defend John$_i$]

Thus, there is no indication that reconstruction is occurring during the interpretation of these examples.

In light of these data, we might be led to ask if comparatives pattern completely with Cause-Effect examples, despite the fact that their interpretation is clearly rooted in the establishment of parallelism and contrast. They in fact do not, since they still require the source and target VPs to be parallel. Consider examples (118) and (119).

(118) # Sue was defended by John more competently than Bob did. [defend Sue]

(119) # Sue introduced John to everyone more quickly than Bill was. [introduced to everyone]

Unlike similar cases in which Cause-Effect relations are operative, the voice mismatch in these examples renders them infelicitous.

I would speculate that the reason why maximal parallelism is not required for comparatives is related to the fact that the construction itself distinguishes the central point of contrast between the two clauses it connects. As such, comparatives express one-place Contrast relations, in which the remainder of the material is the predicate $p$ that is inferred to be common across the two clauses. As such, establishing parallelism below the VP level is unnecessary, and hence reconstruction is not invoked.

The comparative data might also suggest that the critical determinant of whether or not binding theory violations are in force is not the type of coherence relation between the clauses, but whether the syntactic relation between the clauses is one of coordination or subordination. This distinction places the Cause-Effect and comparative data in the same class, which would appear appropriate since neither gives rise to such violations. Indeed, this proposal has been made previously by Hestvik (1995), although his analyses only addresses the case of Condition A violations.

We can address this question by considering examples which contain

clauses that are syntactically coordinated yet give rise to Cause-Effect relations. Consider (120) and (121).

(120) The alleged murderer$_i$ didn't want to defend himself$_i$, and so his lawyer did. [defend himself$_i$]

(121) John$_i$ hated the idea of introducing himself$_i$ to everyone, and therefore Mary had to. [introduce himself$_i$ to everyone]

Contra Hestvik's generalization, I find these cases to be acceptable despite the predicted Condition A violation.

Presumably alluding to this type of data, Hestvik states that "*and* also has a 'consequence' reading, as in *Mary hit him, and John cried*, which may result in syntactic subordination, leading to the expectation that the strict reading would be facilitated by this interpretation" (p. 216). He leaves the issue for future exploration, but we can analyze this idea a little further. Basically what is needed is independent evidence that shows that *and* is actually a syntactic subordinator in these examples. (Relying on the ellipsis data as a basis for this argument would render the analysis circular, of course.) Indeed, a standard test to distinguish subordination from coordination – the ability to move the clause headed by the connective to the front of the sentence – supports just the opposite claim. Consider examples (122a-d).

(122) a. John cried because Mary hit him.

b. Because Mary hit him, John cried.

c. Mary hit him, and John cried.

d. * And John cried, Mary hit him.

The clause headed by *and* in the Cause-Effect sentence (122c) still cannot be preposed as shown in (122d); compare this with the ability to prepose the subordinate clause in (122a) as shown in (122b). Thus, these facts support the claim that it is the operative coherence relation is the critical factor, and not the nature of the syntactic relationship between the clauses.

In sum, the data indicate that VP-ellipsis is only felicitous in the context of a Resemblance relation when the source VP is syntactically parallel to the target VP. Furthermore, certain constructions that are strongly rooted in parallelism, such as the *and...too* construction, appear to invoke the reconstruction of the source VP, which may in turn give rise to syntactic constraint violations in the target. These restrictions contrast to what we found for Cause-Effect relations, in which such constraints do not appear to be manifest.

### 3.3.3 Ellipsis and Contiguity Relations

As I discussed in Chapter 2, the constraints underlying Contiguity relations are perhaps not as well understood as the other relations. Recall that I included only one relation in the Contiguity class, Hobbs's (1990) *Occasion* relation, which allows one to express a sequence of eventualities centered around some system of entities.

**Occasion (i):** Infer a change of state for a system of entities from $S_1$, inferring the final state for this system from $S_2$.

**Occasion (ii):** Infer a change of state for a system of entities from $S_2$, inferring the initial state for this system from $S_1$.

Establishing Occasion thus requires that one be able to infer a situation from a sequence of partial descriptions that are suitably connected through their initial and final states.

While I have not pinned down the mechanics of this inference process in formal detail, in Section 2.2 I described a number of ways in which they are qualitatively different from those involved in establishing Resemblance relations. As such, there is no *a priori* reason to expect that reconstruction of elided material would be necessary during the establishment of an Occasion relation. As such, the expectation is that the data behave in a way consistent with semantic analyses of VP-ellipsis.

It turns out to be difficult to find interesting uses of VP-ellipsis in the context of an Occasion relation, and indeed examples of this sort are not often found in the literature. Let us consider the examples in (123a-c), which are variants of examples I have previously discussed.

(123) a. ?? The problem was solved by John, and then Bill did. [solve the problem] (voice mismatch)

b. ?? This letter evoked a response from Bush, and then Clinton did. [respond] (nominalized antecedent)

c. ?? Sue went to John's$_i$ apartment, and then he$_i$ did. [go to John's$_i$ apartment] (Condition C violation)

I find all of these examples to be marginal; somewhat more acceptable than the corresponding Resemblance cases, but perhaps not as felicitous as the analogous Cause-Effect cases.

The main factor that intervenes with this data is the fact they almost certainly lead to the recognition of a Parallel relation also. Like other examples that I discussed in the previous sections, their marginal character results from the fact that the connector *then* is not enough to force the establishment of Occasion to the exclusion of Parallel, which may be further reinforced by the fact that the source and target clauses focus on

different individuals. Note once again that the addition of the adverbial *too* in these sentences does not affect the meaning substantially.

I will not pursue cases of Contiguity further. However, my theory predicts that examples participating in a Contiguity relation will pattern with those involving Cause-Effect, as long as a Resemblance relation does not also hold.

### 3.3.4 Summary

To summarize to this point, I have characterized ellipsis as a signal from the speaker to the hearer that the information needed for further discourse-level interpretation is readily recoverable from context. In the case of Cause-Effect relations, the recovery of the meaning of the elided VP is enough to satisfy the requirements of the coherence establishment process. In the case of Resemblance relations, on the other hand, the coherence establishment process may require parallelism at the VP level and in some cases access to the structure below, which leaves VP-ellipsis subject to a variety of syntactic constraints in these contexts. Finally, the analysis predicts that VP-ellipsis in the context of a Contiguity relation will not be subject to syntactic constraints, as long as a Resemblance relation is not simultaneously operative.

The range of potential syntactic constraint violations is not limited to those that I have discussed thus far, however. Other types can be manifest, which are the topic of the next section.

## 3.4 Ellipsis and Syntactic Constraints Involving Traces

It has been noted that VP-ellipsis can also be subject to syntactic constraints involving traces. One context that gives rise to trace dependencies is antecedent-contained ellipsis (ACE), exemplified in sentence (124).

(124) John read every book that Bill did.

A straightforward application of syntactic reconstruction runs into a problem on examples like (124), since the source clause for the ellipsis – *read every book that Bill did* – contains the ellipsis site itself. Thus, a straightforward copying procedure would lead to an infinite regress, in which each iteratively reconstructed VP would contain another embedded ellipsis site to resolve.

However, a syntactic reconstruction approach can account for examples like (124) at a level of syntactic logical form (LF), as long as the quantified NP is raised prior to the resolution of the ellipsis (Sag 1976, May 1985, Fiengo and May 1994). Consider the simplified representation for (124) given in (125).

(125) [John [$_{VP}$ read every book that Bill [$_{VP}$ did]]]

The quantified NP is first raised, leaving behind a VP that contains a trace as the source representation, per (126).

(126) [[ every book$_i$ [ that [ Bill [$_{VP}$ did $\phi$ ]]]] [ John [$_{VP}$ read $t_i$ ]]]

The VP [$_{VP}$ read $t_i$ ] can then be reconstructed at the site of the auxiliary *did*, as shown in (127).

(127) [[ every book$_i$ [ that [ Bill [$_{VP}$ did [$_{VP}$ read $t_i$ ]]]]]
[ John [$_{VP}$ read $t_i$ ]]]

This results in the desired LF representation.

However, Dalrymple et al. (1991) show how cases of ACE can also be resolved within their purely semantic resolution process. Their mechanism allows elided VPs to contain bound variables in their interpretation, on analogy with the ability for elided VPs to contain traces in syntactic analyses. The quantified NP in the source clause leads to a quantifier assumption under which the remainder of the clause is interpreted, as shown in (128).

(128) $\langle$every x book(x) $\wedge$ P(bill)$\rangle \vdash$ read(john, x)

Note that the quantifier assumption contains the semantic representation of the ellipsis site. The remainder of the derivation proceeds as in earlier examples. The equation shown in (129) is set up,

(129) P(john) = read(john, x)

which generates the solution shown in (130).

(130) P = $\lambda$z.read(z, x)

Substitution of this solution for $P$ in (128) yields the representation shown in (131).

(131) $\langle$every x book(x) $\wedge$ read(bill, x)$\rangle \vdash$ read(john, x)

Discharging the assumption generates the final representation given in (132), which is the desired result.

(132) every(x, book(x) $\wedge$ read(bill, x), read(john, x))

While this analysis is capable of generating the correct interpretation of sentence (124), note that it leaves open the question of why (124) is syntactically acceptable. In particular, assuming that constraints on gap-filler dependencies fall within the domain of syntactic relations, it is not clear where the trace required by the relative clause comes from without a reconstruction process.

Indeed, a syntactic approach predicts that VP-ellipsis will be subject to syntactic constraint violations involving traces, and there is evidence

that it is. Consider examples (133) and (134).

(133) * John read everything which Bill believes the claim that he did. [read $\phi$]

(134) * John read everything which Bill wonders why he did. [read $\phi$]

As pointed out by Haik (1987), the unacceptability of these examples is predicted by constraints on subjacency, because the dependency between the gap and its antecedent spans two bounding nodes. Again, the prototypical semantic approach offers no explanation for this fact.

I can see three ways in which these facts might be addressed within the analysis I have proposed. First, the analysis would predict them without further modification if it was established that such examples involve Resemblance relations, as that would in turn be the basis for reconstruction. While it is admittedly not clear that such a relation holds between the clauses, it remains as a possible line of investigation.

A second possibility is to offer an analysis in the spirit of Chao (1987). In her account, the need to satisfy wh-trace dependencies in the target is what causes the missing syntactic material to be reconstructed. Thus, along with the need to identify missing arguments during the establishment of Resemblance, one could view the satisfaction of trace dependencies as another requirement that is capable of invoking syntactic reconstruction in my account.

Finally, these facts could conceivably be explained within syntactic theories capable of representing trace dependencies without movement or reconstruction, such as HPSG and LFG. Such an analysis would coordinate the trace dependency represented at the elided VP node in the syntactic representation with a variable within the anaphorically-resolved semantic representation of the VP. (This suggestion is due to Mark Gawron, p.c.).[9] Note that such a mechanism would not require the embedded VP to be resolved to the matrix VP. Consider (135).

(135) John didn't read every book$_i$ that Sue bought $t_i$, but he did read every book$_j$ that Fred did. [bought $t_j$]

The final ellipsis in (135) is not ACE; its source is the embedded VP containing a trace in the previous clause. (See also Fiengo and May (1994, p. 239) for discussion of a similar example.) This interpretation is effectively a sloppy reading involving a trace rather than a pronoun.

[9]Lappin (1999) provides an *in situ* analysis of ACE in HPSG in which syntactic material is reconstructed at the target site. He notes, however, that "nothing in principle excludes the possibility of an *in situ* analysis relying on a purely semantic procedure for ellipsis resolution" (p. 69).

In sum, there are several ways in which the data involving trace dependencies and the syntactic violations associated with them might be handled within the current framework. I will leave the question of which analysis is best as a subject for future work.

However, there is evidence that suggests that coherence establishment plays a role in determining whether or not other types of trace violations will be manifest, particularly with respect to parasitic gap configurations (Rooth 1981, Fiengo and May 1994, Lappin 1999). First, consider sentences (136) and (137), from which nothing has been elided.

(136) * Which problem did you think John would solve because of the fact that Susan solved? (from Rooth (1981))

(137) * These are the McNuggets that Hillary wouldn't eat despite the fact that Bill ate.

These examples are unacceptable because the second gap, which is parasitic on the wh-trace object in the first clause, appears in a complex NP. The sensitivity to the parasitic gap violation is not evident in the elided counterparts of these cases, however, shown in (138) and (139).

(138) Which problem did you think John would solve because of the fact that Susan did? [solved] (from Rooth (1981))

(139) These are the McNuggets that Hillary wouldn't eat despite the fact that Bill did. [ate]

Cause-Effect relations are operative in both sentences, *Explanation* and *Denial of Preventer* respectively. A syntactic analysis of VP-ellipsis predicts that these cases will be unacceptable, since reconstruction will effectively restore them to the structure of their unelided counterparts.[10]

These cases are different than the subjacency examples in (133) and (134) in that there is no dependency within the sentence that requires there to be a trace within the elided VP. For instance, example (140), which is similar to example (138) except that the second clause contains a full VP, is perfectly acceptable.

(140) Which problem did you think John would solve because of the fact that Susan solved Rubik's cube?

In contrast, example (141), which is likewise similar to (133) but with a full VP in the second clause, results in a trace violation.

[10]Kennedy (1997) presents a syntactic analysis that relies on the assumption that reconstructed target VPs in examples such as these contain a pronoun rather than the expected trace. His account requires that the *vehicle change* proposal of Fiengo and May (1994) be adopted, which is discussed in Section 3.5.

(141) * John read everything which Bill believes the claim that he read *Moby Dick*.

A semantic theory is capable of predicting the acceptability of examples (138) and (139), as long as it allows missing VPs to contain bound variables (as does DSP). My account likewise predicts these facts, since these examples participate in Cause-Effect relations.

## 3.5 Comparison with Past Work in VP-Ellipsis

To facilitate the presentation of my theory, I have contrasted my analysis of VP-ellipsis resolution with the prototypical syntactic and semantic approaches I outlined in Section 3.1. In this section, I compare the analysis with several specific instances of past work. VP-ellipsis has been so well-studied that a comprehensive discussion of the literature would prove intractable. Thus, I will focus primarily on a small number of theories that either deviate from the prototypical accounts in an important respect, or otherwise present arguments and data that bear directly on the arguments I have made.

Lappin (1993a, 1993b, 1996) has outlined a theory of VP-ellipsis resolution in which a copy of the surface syntactic representation of the source VP is reconstructed within the target clause representation. His analysis is essentially the the prototypical syntactic approach, except that it operates at the level of surface structure rather than LF.

Lappin argues for a surface structure analysis based in part on the type of binding condition violations discussed in Section 3.3. However, he does acknowledge the existence of cases that are acceptable despite the fact that binding conditions would be expected to apply. These cases include examples (142) and (143), taken and adapted from Dalrymple (1991).

(142) The lawyer defended $\text{Bill}_i$ against the accusations because $\text{he}_i$ couldn't.

(143) I expected $\text{Bill}_i$ to win even when $\text{he}_i$ didn't.

Lappin addresses this data by appealing to Evans's (1980) observation that binding condition effects may be overridden in part by placing contrastive accent on the pronoun. Indeed, the pronouns in examples such as sentences (142) and (143) do generally receive additional accent. Evans gives examples such as (144).

(144) Everyone has finally realized that Oscar is incompetent. Even $\text{HE}_i$ has finally realized that $\text{Oscar}_i$ is incompetent.

The claim is that the second sentence in passage (144) is acceptable as long as the pronoun is contrastively accented, despite the expected

Condition C violation. Lappin argues that sentences (142) and (143) are acceptable for the same reason, thus making this data consistent with a purely syntactic account.

I find this argument to be problematic, however. First, while the discourse context and intended discourse effect help mediate the acceptability of passage (144), I do not find it to be perfectly acceptable. A certain degree of stiltedness remains, which has the effect of drawing out the intended parallelism. In contrast, I find sentences (142) and (143) to be perfectly acceptable, with no hint of the stiltedness or resulting discourse effect of passage (144). Furthermore, under Lappin's argument, the unelided versions should sound as natural as the elided versions. The unelided versions of sentences (142) and (143) are given in (145) and (146), respectively.

(145) The lawyer defended $\text{Bill}_i$ against the accusations because $\text{HE}_i$ couldn't defend $\text{Bill}_i$ against the accusations.

(146) I expected $\text{Bill}_i$ to win even when $\text{HE}_i$ didn't expect $\text{Bill}_i$ to win.

I find these versions to be considerably less acceptable; in fact, these have the very properties of Evans's example (144) that are absent in (142) and (143).

Finally, under Lappin's argument, the Condition C violation should be suspended by the use of accent regardless of the type of coherence relation that is operative between the clauses. Consider examples (147) and (148), which are similar to sentences (142) and (143) but participate in the Parallel relation.

(147) The lawyer defended $\text{Bill}_i$ against the accusations, and $\text{HE}_i$ did too.

(148) I expected $\text{Bill}_i$ to win, and $\text{HE}_i$ did too.

Again, the added accent improves the acceptability of these examples somewhat, but does not make them completely felicitous, and again the effect is similar to that for Evans' example (144). This effect is notably distinct from the lack of such an effect in sentences (142) and (143). Furthermore, the effect remains similar to those for the unelided versions of (147) and (148), shown in (149) and (150).

(149) The lawyer defended $\text{Bill}_i$ against the accusations, and $\text{HE}_i$ defended $\text{Bill}_i$ against the accusations too.

(150) I expected $\text{Bill}_i$ to win, and $\text{HE}_i$ expected $\text{Bill}_i$ to win too.

As a result, we have a fairly clear distinction: the Resemblance examples (147) and (148) share the properties of their unelided counterparts (149) and (150), whereas the elided Cause-Effect examples (142) and

(143) do not share these properties with their unelided counterparts in (145) and (146). This difference suggests that reconstruction is only occurring in the Resemblance cases. Lappin's analysis cannot account for this fact since coherence relations play no role in his analysis, whereas this difference is exactly what my analysis predicts. Furthermore, Lappin's analysis leaves open the question of how to account for the syntactic mismatch data discussed in Section 3.3.2.

In his extensive study of VP-ellipsis, Sag (1976) provides an account in which VP-ellipsis results from deletion under identity (more specifically, identity under "alphabetic variance") at a level of syntactic logical form. Because of this identity constraint, the predictions of his analysis correspond to that of the prototypical syntactic analysis presented here, even though there is no reconstruction. An exception is the case of Condition A violations; in his account reflexives can always receive either strict or sloppy readings.

Webber (1978) offers several examples of VP-ellipsis that are problematic for Sag's account, for which she argues that inference is required to create the necessary referent. Such examples include (151) and (152).

(151) Wendy is eager to sail around the world and Bruce is eager to climb Kilimanjaro, but neither of them can because money is too tight.

(152) Irv and Martha wanted to dance together, but Martha's mother said that she couldn't.

The recovered material in (151) corresponds roughly to *do what (s)he is eager to do*, and the recovered material in (152) likewise corresponds to *dance with Irv*. In neither case is the requisite source syntactic material present, and thus these examples provide further support for a semantic analysis of VP-ellipsis resolution. However, she ultimately follows Sag in not allowing cases of voice mismatch, based on examples of a now familiar sort, given in (153a-b).

(153) a. # The aardvark was given a nut by Wendy, and Bruce did too.

b. # Wendy avoided the aardvark, and the axolotl was too.

In a footnote, she gives examples of equi sentences that "seem to approach acceptability" despite a voice mismatch, given in (154a-b), admitting that "it is not yet clear to me why such examples should approach acceptability" (Chapter 4, p. 30, fn. 26).

(154) a. Although the steaks were ready to eat at 6pm, by 7pm they still hadn't been.

b. Usually John is easy to please, but by this play, he wasn't.

As I have argued in this chapter, the important difference between examples (153a-b) and (154a-b) is that they participate in Resemblance and Cause-Effect relations respectively; we have seen a variety of other felicitous cases of voice mismatch that are not equi constructions.

Hestvik (1993, 1995) offers an account in which reconstruction occurs at the level of LF. Focusing primarily on Condition A violations, he deviates from the prototypical analysis by proposing an operation of *reflexive raising*, which is performed for reflexive pronouns when they appear in a subordinate clause. After being raised, the reflexive is c-commanded by the matrix subject and thus can be bound by it, thus allowing for a strict reading without violating Condition A. Because reflexive raising only occurs in cases of subordination, the account predicts that no strict reading is available in cases of coordination.

The current analysis and Hestvik's make many of the same predictions with respect to Condition A, because Cause-Effect relations and comparatives usually co-exist with syntactic subordination. As we saw in Section 3.3.2, however, examples in which Cause-Effect relations coexist with syntactic coordination suggest that the current analysis makes a more adequate distinction. With respect to other binding constraints, his approach patterns with the prototypical syntactic approach. He invokes the same argument as Lappin with respect to Evans's work, which I have already addressed, and does not address cases that involve a mismatch of syntactic form.

Fiengo and May (1994) provide an extensive reconstruction account of VP-ellipsis at the level of LF. While their analysis also patterns with the prototypical syntactic account presented here, they do address certain examples that have been used to support semantic analyses. For instance, they discuss the example of voice mismatch shown in sentence (155), under the reading in which *it* refers to Congress.

(155) This law restricting free speech should be repealed by Congress, but I can assure you that it won't. [repeal this law restricting free speech]

They address the issue by considering the question of whether a trace, such as one left behind by passivization, can serve as an antecedent of the argument of an elided VP. They conclude that it can, so that examples like sentence (155) are rendered acceptable. A problem with this account is that it predicts that *all* cases of passive-active voice alternation should be acceptable, which we have already established is not the case in examples involving Resemblance relations. Also, it is not clear how their account could be extended to other cases of structural mismatch, such as those with nominalized antecedents.

With respect to binding theory constraints, Fiengo and May posit a process called *vehicle change*, which can allow pronouns, reflexives, and full NPs to be allowable reconstructions of each other. While binding conditions still apply at the level of LF, vehicle change will allow certain examples to be acceptable that the prototypical reconstruction approach does not. Their system cannot predict a difference in readings between examples in Resemblance and Cause-Effect relations that otherwise have equivalent syntactic conditions, however.

Finally, I discuss two previous studies that explicitly tie ellipsis resolution to an account of discourse structure and coherence, specifically those of Prüst (1992, see also Prust et al. (1994)) and Asher (1993). Prüst's account is articulated within the Linguistic Discourse Model theory of discourse structure (Polanyi 1988, Scha and Polanyi 1988). He defines a mixed representation, referred to as *syntactic/semantic structures*, that amounts to (unapplied) semantic functions and arguments arranged in a syntactic configuration similar to the surface syntax of the sentence. He then gives a method for inferring parallel and contrast relationships by computing the *Most Specific Common Denominator* over these structures. However, following Sag (1976), he assumes that VP-ellipsis always requires a syntactically-matching antecedent. Because of the extremely tight integration of syntactic and semantic information in his mixed representation, it is not clear how his operations could be adapted so as to allow for cases that support purely semantic theories of VP-ellipsis resolution.

Asher (1993) also provides an analysis of VP-ellipsis within a larger account of discourse structure and coherence, in this case working within Discourse Representation Theory (Kamp 1981, Kamp and Reyle 1993). However, he also follows Sag in requiring that elided VPs be syntactically parallel to their referents. While the framework he is working in is semantic, the semantic forms for VPs are represented distinctly from those for sentences, which has the effect of ruling out cases of syntactically-mismatched antecedents. In fact, his accounts of VP-ellipsis and of other event referential forms (e.g., *do it* and *do that* anaphora) differ in that only the latter allow for the abstraction necessary to handle such antecedents.

## 3.6 Linguistic Form and Readings for VP-Ellipsis

Up until this point, I have used Dalrymple et al. (1991) as my exemplar of semantic analyses. Recall that this analysis operates at a level of representation at which all remnants of the linguistic form of the source and target have been lost. I have shown that such an analysis can be

integrated with an account in which reconstruction is triggered by the need to establish coherence that explains why VP-ellipsis interpretation appears to be syntactically-mediated in certain contexts.

However, there are other respects in which the linguistic form of the source clause can affect the readings available for a subsequent VP-ellipsis. This is particularly the case with respect to the set of strict and sloppy readings that are available. There are two types of linguistic form distinctions to consider: those pertaining to choice of referential expression, and those pertaining to certain syntactic relations.

I first consider the effect of referential form. As we have seen, the appearance of an intrasententially-referring pronoun in the source clause can potentially lead to strict and sloppy interpretations in the target clause.[11] It is therefore crucial that the representations used by a semantic analysis have a means for distinguishing pronouns from other types of coreferential phrases that do not result in such ambiguities. I first illustrated the issue in Kehler (1993a) with respect to stripping constructions rather than VP-ellipsis; consider example (156).

(156) John likes his mother, and Mary's too.

The target clause only has one reading, according to which John likes Mary's mother. The DSP system generates two readings, including a reading in which Mary likes Mary's mother. The representation of the source of (156) is shown in (157).

(157) $likes(John, mother\_of(\underline{John}))$

The second occurrence of *John* is primary, since *Mary* is parallel to *his*. The equation to be solved is shown in (158).

(158) $P(John) = likes(John, mother\_of(\underline{John}))$

The correct analysis is achieved with the solution given in (159), resulting in the reading in which John likes Mary's mother, per (160).

(159) $P = \lambda x.likes(John, mother\_of(x))$

(160) $P(Mary) = likes(John, mother\_of(Mary))$

However, equation (158) is also consistent with the solution shown in

[11] Other context-dependent phenomena such as implicit arguments also appear to give rise to sloppy readings. Consider examples (i) and (ii), adapted from Partee (1989).

(i) John went to a local bar to watch the Superbowl, and Bob did too.

(ii) George drove to the nearest hospital, and Fred did too.

The target clause of sentence (i) can be interpreted as Bob having gone to a bar that is local to John (strict) or himself (sloppy); likewise in sentence (ii) Fred may have gone to the hospital nearest to George (strict) or himself (sloppy).

(161), which yields the reading in which Mary likes Mary's mother, per (162).

(161) $P = \lambda x.likes(x, mother\_of(x))$

(162) $P(Mary) = likes(Mary, mother\_of(Mary))$

Sentence (156) does not have this reading because the pronoun in the source is parallel to overt material in the target, and hence is not part of the material recovered for the target. Since pronouns are represented by terms denoting their referents in DSP's representation, they have no mechanism for capturing this fact.

A related problem can be illustrated in the context of VP-ellipsis with examples like (163a-b).[12]

(163) a. Bob Dole likes Bob Dole's mother, and Bill Clinton does too.
b. Bob Dole likes his mother, and Bill Clinton does too.

Whereas the target clause in example (163b) has strict and sloppy readings, the target clause in example (163a) has only a strict reading, that is, it can only mean that Clinton likes Dole's mother. DSP also generates a sloppy reading, in which Clinton likes his own mother. The derivation follows that of example (156). The semantic representations for the source and target clauses are shown in (164) and (165) respectively.

(164) $likes(\underline{Dole}, mother\_of(Dole))$

(165) $P(Clinton)$

Equation (166) has (167) as a possible solution.

(166) $P(Dole) = likes(\underline{Dole}, mother\_of(Dole))$

(167) $P = \lambda x.like(x, mother\_of(x))$

[12]Webber (1978) makes essentially the same point, noting that the following examples only have strict readings:

(i) Only John wanted Mary to kiss John. Fred didn't.

(ii) The king may hunt on the king's land. The prince may too.

I did not use examples like (163a-b) and (i)-(ii) in Kehler (1993a) out of fear that the Condition C violations would distract from the (largely orthogonal) point about DSP's method. Similarly, Gardent (1997) makes the same point with the following minimal pair, in which she claims that (iii) has a sloppy reading whereas (iv) does not:

(iii) $\text{Jon}_i$'s mother loves $\text{him}_i$ and Bill's mother does too.

(iv) $\text{Jon}_i$'s mother loves $\text{Jon}_i$ and Bill's mother does too.

While there is no Condition C violation in (iv), the claim that examples like (iii), in which *Jon* does not c-command the pronoun *him*, have a sloppy reading has proven controversial (cf. Reinhart (1983), Hirschberg and Ward (1991)). This is again an orthogonal issue however; each of these pairs of examples suffices to demonstrate the point at hand.

The application of $P$ to *Clinton* in the target representation yields representation (168).

(168) $likes(Clinton, mother\_of(Clinton))$

The corresponding analysis for sentence (163b), which has this reading, is essentially identical to that of (163a). Again, this results from the fact that pronouns are not distinguished from the terms denoting their referents in the semantic representation. The lack of a sloppy reading for (163a) is due to the fact that, unlike (163b), there is no pronoun in the recovered material to give rise to such an ambiguity. Thus, if a semantic approach to ellipsis resolution is to be maintained, distinctions between different forms of reference must be manifest at the level of semantic representation.

Likewise, there are ways in which apparently syntactically-governed dependencies among pronouns and their antecedents in a source clause can limit the number of readings available in the target clause. As DSP acknowledge, sentence (169) has only three of the four readings that their system derives, and sentence (170) has only five of the six expected readings.

(169) Bill believed that he loved his wife, and Harry did too. (Dahl 1974)

(170) John revised his paper before the teacher did, and Bill did too. (Gawron and Peters 1990)

The missing readings appear to be eliminated by certain types of dependencies between pronouns and their antecedents in the source clause, which in turn govern the possible ways that the recovered material can be applied in the target representation.

Kehler (1993a, 1995) and Hobbs and Kehler (1997) provide semantic analyses that successfully address both types of data discussed above. In the case of examples (156) and (163a), the semantic representations used in these analyses manifest a distinction between entities described by full NPs and those described by pronouns; the resolution algorithm in turn expresses a sensitivity to this distinction that results in the correct behavior for these examples. The representation likewise manifests certain syntactically-driven dependency relationships between the terms denoting coreferential expressions; again, a sensitivity to these relationships allows the correct set of readings to be derived for examples (169) and (170).

A detailed discussion of these accounts would take us too far afield, so the reader is referred to those works. The important point for my analysis here, however, is that the existence of such constraints imposed

| Phenomenon | Example |
|---|---|
| 'Do It' Anaphora | John revised his paper before Bill did it. |
| 'Do So' Anaphora | John revised his paper and Bill did so too. |
| Stripping | John revised his paper, and Bill too. |
| Comparative Deletion | John revised his paper more quickly than Bill. |
| 'Same As' Reference | John revised his paper, and Bill did the same. |
| | John revised his paper, and the teacher followed suit. |
| 'Me Too' Phenomena | A: John revised his paper.<br>B: Me too./Ditto. |
| 'One' Anaphora | John revised a paper of his, and Bill revised one too. |
| Lazy Pronouns | The student who revised his paper did better than the student who handed it in as is. |
| Anaphoric Deaccenting | John said he called his teacher an idiot, and Bill *said he insulted his teacher* too. |
| Focus Phenomena | Only John revised his paper. |

TABLE 3 Phenomena Giving Rise to Sloppy Interpretations

by linguistic form – referential, as in examples (156a) and (163a-b), and syntactic, as in examples (169) and (170) – do not in and of themselves argue for a theory in which the resolution process occurs at the level of syntax. On the other hand, they do argue that information encoded within linguistic form needs to be communicated through the syntax/semantics interface, and that the semantic representation chosen must be capable of representing such information.

The claim that these data do not argue for a syntactic approach is also supported by the fact that strict and sloppy readings are not restricted to VP-ellipsis. They are in fact common to a wide range of constructions that rely on parallelism between two eventualities. A selection of these are listed in Table 3, taken from Hobbs and Kehler (1997). (See also Kehler (1993a, 1995) and Gardent (1997)). All of these forms of reference give rise to essentially the same sets of strict and sloppy ambiguities. However, many of them do not involve elided material, and thus are not candidates for a syntactic reconstruction account of their interpretation. Indeed, any account of strict and sloppy readings that applies only to forms of ellipsis – as analyses based on reconstruction inherently do – is almost certainly missing an important generalization.

## 3.7 Situationally-Evoked Referents, Event Reference, and the Sag and Hankamer Dichotomy

Hankamer and Sag's (1976, henceforth H&S) classic study of reference argues for a categorical distinction between two types of anaphora: *deep* and *surface*. Surface anaphoric expressions are 'syntactically controlled', which imposes a requirement for an antecedent of an appropriate syntactic form. This class includes VP-ellipsis and gapping (discussed in the next chapter), among others. Deep anaphoric expressions, on the other hand, require only an appropriately constructed semantic referent in the discourse model. This category includes pronominals and, by extension, event referential forms such as *do it* and *do that* anaphora. This dichotomy is revised to a distinction between two types of anaphoric process in Sag and Hankamer (1984), *ellipsis* (surface anaphora) and *model-interpretive anaphora* (deep anaphora).

The analysis of VP-ellipsis presented here likewise predicts that deep anaphoric expressions like *do it* and *do that* will not display a sensitivity to syntactically-governed constraints. Syntactic constraints arise only as result of reconstructing elided material, and thus they will not affect the acceptability of expressions that do not involve ellipsis. Nonetheless, my analysis demonstrates that the division between forms for which resolution is syntactically versus semantically mediated is not as clear cut as the deep and surface anaphora distinction would suggest. Although H&S categorize VP-ellipsis as surface anaphora, there is ample evidence to suggest that it patterns more like deep anaphora with respect to the level of representation at which its meaning is recovered. The complicating factor is that the requirement for a syntactically suitable referent, a property of all surface anaphors in H&S's account, is only manifest in certain contexts in the current analysis.

However, the existence of syntactic effects was not the only basis for H&S's decision to categorize VP-ellipsis as surface anaphora. In particular, the requirement that there be a syntactic antecedent for surface anaphora in their account implies that the antecedent must be linguistic, that is, surface anaphora cannot specify referents that are situationally evoked (or in their terminology, 'pragmatically controlled'). They illustrate the infelicity of VP-ellipsis with situationally-evoked referents with example (171).

(171) [ Hankamer points gun offstage and fires, whereupon a blood-curdling scream is heard. Sag says: ]

a. * Jorge, you shouldn't have! [ VP-ellipsis (surface) ]

b. Jorge, you shouldn't have done it! [ *do it* anaphora (deep) ]

This behavior is left unexplained by an account that simply categories VP-ellipsis as deep anaphora.

The claim that VP-ellipsis is incompatible with situationally-evoked referents is not uncontroversial, however. Schachter (1977b) provides a number of cases which appear to be felicitous, including examples (172) and (173).

(172) [John tries to kiss Mary. She says:]
John, you mustn't.

(173) [John pours another martini for Mary. She says:]
I really shouldn't.

Schachter argues for a proform theory of VP-ellipsis based in part on these examples, and others have followed suit (Chao 1987, Hardt 1992, Lobeck 1999). Lappin (1993b) responds to Hardt with respect to this argument, stating that syntactic and semantic approaches are on equal ground in handling these cases. Neither author cites Hankamer's (1978) response to Schachter, however, which argues convincingly that such cases of VP-ellipsis are, in his terms, either formulaic or conventionalized, occurring only as "illocutionally charged expressions" and not generally as declarative statements or informational questions. Consider examples (174) and (175).

(174) [John tries to kiss Mary. She says:]
# John, you're the first man who ever has.

(175) [John pours another martini for Mary. She says:]
# John, are you aware that no one else has?

These uses of VP-ellipsis are infelicitous, despite the fact that the contexts are identical to those for Schachter's examples (172) and (173). Hankamer therefore argues that the ability to refer to situationally-evoked antecedents does not extend to VP-ellipsis in general.

Although I maintain that VP-ellipsis resolution is essentially a (semantically based) anaphoric process, I agree with Hankamer that it is not productive in its ability to refer to situationally-evoked antecedents. I would therefore argue that, contra Hankamer and Sag (1976), the questions of whether a linguistic form requires a syntactically-parallel antecedent and whether it allows for reference to situationally-evoked antecedents need to be distinguished.

Indeed, VP-ellipsis is not the only referential expression that demonstrates this need. Another form that H&S treat as a surface anaphor, the *do so* construction, also displays this behavior (Kehler and Ward

1995, Kehler and Ward 1999). First, the use of *do so* in the contexts I have been considering demonstrate that it is infelicitous with situationally-evoked referents.

(176) [ Hankamer points gun offstage and fires, whereupon a blood-curdling female scream is heard. Sag says: ]
# Jorge, you shouldn't have done so!

(177) [ John tries to kiss Mary. She says: ]
# John, you mustn't do so.

(178) [ John pours another martini for Mary. She says: ]
# I really shouldn't do so.

On the other hand, like VP-ellipsis in certain contexts, *do so* does not satisfy the other characteristic of surface anaphora, namely the requirement that there be a syntactically suitable antecedent. In all of the following (naturally occurring) examples, taken from Kehler and Ward (1999), there is no surface-syntactic VP available to serve as an antecedent for *do so*.

(179) Section 1 provides the examples to be derived by Gapping, and a formulation of Gapping capable of *doing so*. [=deriving the examples] (from text of Neijt (1981))

(180) As an imperial statute the British North America Act could be amended only by the British Parliament, which *did so* on several occasions. [=amend an imperial statute] (Groliers Encyclopedia)

(181) The defection of the seven moderates, who knew they were incurring the wrath of many colleagues in *doing so*, signaled that it may be harder to sell the GOP message on the crime bill than it was on the stimulus package. [=defecting] (Washington Post)

(182) There was a lot more negativity to dwell on, if anyone wished to *do so*. [=dwell on more negativity] (Wall Street Journal)

(183) With or without the celebration, Belcourt is well worth seeing, and you can *do so* year round. [=see Belcourt] (Wall Street Journal)

Therefore, *do so* cannot be appropriately classified as deep or surface anaphora either, and thus the requirement for syntactic parallelism and the ability to specify situationally-evoked antecedents should be distinguished in any theory of anaphora. As such, the fact that VP-ellipsis cannot be productively used with situationally-evoked referents does not in and of itself provide evidence that VP-ellipsis resolution is a syntactically-governed process.

## 3.8 Conclusion

In this chapter, I classified past approaches to VP-ellipsis interpretation according to whether the resolution procedure operates at a level of syntactic or semantic representation, and presented data that is problematic for each type of analysis. This data was shown to exhibit a systematicity, however, with respect to the coherence relation that is operative between the source and target clauses. In particular, the predictions of a syntactic account were borne out when a Resemblance relation was operative, whereas semantic accounts made the correct predictions when a Cause-Effect relation was operative. I presented a theory in which two independently-motivated aspects of language interpretation interact: the syntactic and anaphoric properties of VP-ellipsis itself, and the properties of the inference processes that underlie the establishment of different classes of coherence relations. The result predicts the distribution of the data found, making it more empirically adequate than approaches based on either syntax or semantics alone.

# 4

# Coherence and Gapping

In this chapter we direct our attention to the gapping construction, exemplified in sentence (184).

(184) Dick supports George, and Joe, Al.

Gapping is characterized by an antecedent clause, which I will again refer to as the source clause, and the elision of all but two (and in some cases, more than two) bare constituents in one or more subsequent target clauses. While bearing certain similarities to VP-ellipsis, gapping is also different in several ways that will become clear shortly.

Sag's (1976) seminal study catalogs a wealth of interesting facts concerning gapping constructions; I recount some of these that are relevant for the analysis I present here. First, as have several other researchers, he notes that gapping applies only to coordinate (and not subordinate) structures:

(185) a. Sandy played the guitar, and Betsy the recorder.

b. * Sandy played the guitar, while/after/before/since/although Betsy the recorder.

Second, gapping can apparently apply only to the highest sentence node, and not to an embedded one:

(186) * Alan went to New York, and Bill met a man who claimed (that) Betsy to Boston.

Third, gapping can cross speaker boundaries only in a very restricted way. Sag considers such examples to be cases in which the conversational participants are 'collaborating' to construct what can conceptually be thought of as a single sentence, as in (187).

(187) A. Jorge is peeling an apple.

B. And Ivan an orange.

Fourth, Sag notes that "Gapping remnants must also, in some poorly understood sense, be parallel to corresponding elements in the left conjunct." The following sentences are unacceptable due to a lack of such parallelism:

(188) a. * Sam hates reptiles, and Sandy to talk to Oh.
b. * Beth ate yogurt, and Norma at midnight.

Hudson (1976) makes a similar point, noting that sentence (189) is unacceptable because "there is no direct contrast between *in the bath* and *arias at the top of her voice*".

(189) * John sings in the bath and Mary arias at the top of her voice.

Fifth, Sag points out that gapping can apply iteratively in sentences containing more than two conjuncts:

(190) Ray plays the clarinet, Lois the oboe, John the piano, Sandy the guitar,...

Sixth, he notes that more than two stranded constituents can appear in each target clause in at least some cases. While many speakers find (191a) to be unacceptable, most speakers find (191b) to be perfectly felicitous; in both cases there are three remnants in the target clause.

(191) a. *Alan gave Sandy a book, and Peter Betsy a magazine.
b. Peter talked to his boss on Tuesday, and Betsy to her supervisor on Wednesday.

Sag notes that the acceptability of such examples seems to hinge on whether or not one or more of the remnants resides outside of the verb phrase.

Finally, Sag discusses example (192), which shows that gapping may be felicitous even when there is a mismatch of agreement information between the source and target.

(192) My brothers have all gone to the circus, and my sister to the carnival.

That is, the corresponding unelided version of (192), shown in (193), is ungrammatical.

(193) * My brothers have all gone to the circus, and my sister have all gone to the carnival.

Sag cites this example as evidence that gapping should be resolved at a level of syntactic logical form rather than surface syntax, a claim that I will likewise adopt.

A wealth of other work on gapping exists; most of which has been concerned primarily with describing and accounting for a variety of syn-

tactic constraints on its use (Ross 1970, Jackendoff 1971, Hankamer 1971, Stillings 1975, Kuno 1976a, Hudson 1976, Sag 1976, Rosenbaum 1977, Neijt 1981, Siegel 1984, van Oirsouw 1985, Prince 1986, Chao 1987, Steedman 1990, Jayaseelan 1990, inter alia). I will not attempt even a cursory survey of this literature here, nor will I attempt to offer competing analyses for the set of syntactic facts which they discuss. In fact, what I say in this chapter is largely compatible with these accounts.

I focus my attention instead on a particular phenomenon that, to my knowledge, was first noticed by Levin and Prince (1982, 1986, henceforth, L&P). L&P note that pairs of conjoined sentences such as those in (194a-c) have both *symmetric* and *asymmetric* readings (Lakoff 1971, Schmerling 1975).

(194) a. Sue became upset and Nan became downright angry.
b. Al cleaned up the bathroom and Joe cleaned up the mess.
c. One of the students was accepted at Bryn Mawr and the high school was praised on TV.

That is, each sentence has a symmetric reading in which the two events are understood as independent, as well as an asymmetric reading in which the first event is interpreted as the cause of the second event (in which *and* is can be paraphrased as "and as a result"). In my terms, the symmetric readings correspond to the Resemblance relation Parallel, whereas the asymmetric readings correspond to the Cause-Effect relation Result. L&P contrast the sentences in (194a-c) with their gapped counterparts, given in (195a-c).

(195) a. Sue became upset and Nan ∅ downright angry.
b. Al cleaned up the bathroom and Joe ∅ the mess.
c. One of the students was accepted at Bryn Mawr and the high school ∅ praised on TV.

Unlike examples (194a-c), these sentences only have symmetric readings. For instance, whereas sentence (194a) can have a reading in which Nan became angry *because* Sue became upset, this reading is unavailable in (195a). This can be seen by considering the following contexts, again due to L&P, in which gapping is acceptable in the context favoring the symmetric reading in (196), but not in the context favoring the asymmetric reading given in (197), although in both cases the non-gapped versions are acceptable.

(196) Sue and Nan had worked long and hard for Carter. When Reagan was declared the winner, Sue became upset and Nan became/∅ downright angry.

(197) Sue's histrionics in public have always gotten on Nan's nerves, but it's getting worse. Yesterday, when she couldn't have her daily Egg McMuffin because they were all out, Sue became upset and Nan became/#∅ downright angry.

The causal interpretation of the two final clauses in example (197), supported by the given context, is unavailable when gapping has applied.

While L&P limit their discussion to sentences conjoined with *and*, one might consider examples involving other coordinators that are compatible with both Resemblance and Cause-Effect relations. I start with the coordinating conjunction *or*, which like *and*, has two relevant uses. Consider example (198).

(198) John will go to New York, or Bill will go to Boston.

This sentence also has two readings: a symmetric (disjunctive) reading, and an asymmetric causal reading (e.g., to express a threat of the form *If A doesn't happen then B will!*). As is the case with *and*, gapping in clauses conjoined by *or* is allowable in the symmetric case, but not in the asymmetric case. This can again be seen by embedding each version within a context that favors a particular reading; the contexts in (199) and (200) favor the symmetric and asymmetric readings respectively.

(199) (John's and Bill's boss speaking): A meeting should not be scheduled on Thursday, since one of our people will be heading out of town. Either John will go to New York, or Bill (will go)/∅ to Boston.

(200) (John's and Bill's boss speaking): Listen, John, you might not want to be transferred to New York, and I can't force you to go. But I can transfer Bill, and you can imagine what life would be like around here for you if Bill is not here. Now either you will go to New York, or Bill will go/#∅ to Boston!

Likewise, as we saw in Section 2.2, the coordinating conjunction *but* has a Resemblance sense (Contrast) and a Cause-Effect sense (Violated Expectation). As one might anticipate, gapping with the Contrast relation is acceptable, as in example (201), whereas it is infelicitous with the Violation Expectation reading, as in example (202).

(201) The boys voted today. John voted for Gore, but (in contrast) Tom for Bush.

(202) Tom usually does everything that John does, but not today.
# John voted for Gore, but (surprisingly) Tom for Bush.

Finally, we have already seen in example (185b) that gapping is unacceptable with a variety of subordinating conjunctions; all of these

indicate that a Cause-Effect relation is operative.

(203) a. # John voted for Gore because Tom ∅ Bush.
b. # John voted for Gore even though Tom ∅ Bush.
c. # John voted for Gore despite the fact that Tom ∅ Bush.
d. # John voted for Gore although Tom ∅ Bush.

Thus, the pattern noted for the conjunction *and* by L&P appears to generalize to the broader distinction between Resemblance and Cause-Effect relations. (I will discuss Resemblance relations not covered above in Section 4.2.) In particular, gapping is felicitous when a Resemblance relation is operative, but not when a Cause-Effect relation is operative. Interestingly, this pattern is in some sense reversed from the one we found for VP-ellipsis in Chapter 3. Those contexts which allow gapping (Resemblance) actually disallow certain cases of VP-ellipsis due to syntactic constraints, whereas those contexts which allow these cases of VP-ellipsis (Cause-Effect) disallow gapping.

While previous work on gapping has been almost exclusively rooted in syntax (but cf. Kuno (1976a), in addition to L&P), the difficulty in accounting for this data using only syntactic mechanisms should be apparent. While syntactic approaches often stipulate prohibitions on gapping in various contexts (for instance, in subordinate clauses, which successfully rules out examples (203a-d)), they cannot distinguish between the Resemblance and Cause-Effect readings in examples with coordinating conjunctions such as *and, or,* and *but.* As we argued in Chapter 3, there is no evidence that these readings are associated with different syntactic structures.

As I did for VP-ellipsis in the previous chapter, I will offer an account in which these facts result from the interaction between two sets of properties: the syntactic and referential properties of gapping, and the properties of the establishment mechanisms that underlie the recognition of different classes of coherence relations.

## 4.1 Syntactic and Semantic Properties of Gapping

The first part of the analysis therefore pertains to the syntactic and referential properties of gapping. In Chapter 3, I argued on independent grounds that VP-ellipsis leaves behind an empty node in the syntax (a VP) and is also anaphoric. The evidence suggests that gapping also leaves behind an empty node in the syntax (in this case, an S), but is not similarly anaphoric.

Regarding its syntactic properties, I will follow the essential aspects of Sag's (1976) analysis. Sag argues persuasively that gapping interpretation occurs a level of syntactic logical form (or LF), at which both the

remnants in the target clause(s) and their parallel elements in the source clause have been abstracted from their respective sentence-level representations (which, in the case of the target clause, will be left empty). There are a variety of facts that argue for this choice of representation.

First, several researchers have noted that in gapping constructions, contrastive accent is generally placed on parallel elements in both the source and target clauses. This accent marks the elements as focused, and abstracting them results in an "open proposition" that both clauses share (Sag 1976, Wilson and Sperber 1979, Prince 1986, Steedman 1990). This open proposition needs to be presupposed (or accommodated) for the gapping to be felicitous. Prince says:

> Gappings are felicitous just in case they can be taken to instantiate an OP [=open proposition] corresponding to the full conjunct, where the leftmost constituents bear the same sort of anaphoric (set) relation to something in the prior context found in Topicalization and where the rightmost constituents instantiate the variable in the OP. The OP is taken to be salient shared knowledge, at least at the point in time that the first Gapped conjunct is uttered. (pp. 212–213)

Such presuppositional effects are well-known to occur in other constructions that involve contrastive accent. Sag (1976) cites examples due to Dretske (1972) that show that the placement of accent on constituents can affect the truth conditions of a sentence; consider examples (204a-b).

(204) a. If Clyde hadn't made the last shot, we would have lost.
b. If CLYDE hadn't made the last shot, we would have lost.

Sentence (204b), unlike (204a), implies that the team would have lost if someone other than Clyde had made the last shot – the open proposition evoked by accenting *Clyde* thereby feeds the interpretation of the conditional. Likewise, sentences (205a-b) presuppose that *someone supports Al.* In example (205a) the focused element is marked overtly in the surface syntax by the cleft (as well as being accented), whereas in example (205b) the effect is achieved by only placing contrastive accent on the focused element (and thus, under the proposal being made here, this element gets abstracted at LF).

(205) a. It is JOE who supports Al.
b. JOE supports Al.

Similarly, changing the element that receives accent in such constructions likewise changes the open proposition, which can cause sentence pairs that have the same surface syntactic structures to have different

meanings. Examples include the comparative constructions given in (206a-b).

(206) a. AL defended Bill more eloquently than GEORGE.
[George defended Bill; open proposition is $\lambda x.defend(x, Bill)$]
b. Al defended BILL more eloquently than GEORGE.
[Al defended George; open proposition is $\lambda y.defend(Al, y)$]

The sole difference between sentences (206a) and (206b) – the element on which accent is placed in the source clause (cf. Hankamer (1973), Napoli (1983)) – results in different interpretations being assigned to their respective target clauses.

As one would therefore expect, the same reliance on an open proposition is found in gapping. For instance, it would be infelicitous to open a conversation with a sentence such as (184), repeated as (207), whereas it is perfectly felicitous in response to the question *Who supports whom?*.

(207) Dick supports George, and Joe, Al.

In fact, one must be careful to judge the acceptability of a gapped sentence only in a context that licenses its use. Steedman (1990), citing Kuno (1976a), notes that all of the following sentences have been claimed to be ungrammatical by one or more proponents of syntactic analyses, under the readings in which the brackets indicate the elided material in the target.

(208) a. Harry [went to] London, and Barry, Detroit.
b. Harry [will give] a bone to a dog, and Barry, a flower to a policeman.
c. Harry [claimed that hedgehogs eat] mushrooms, and Barry, frogs.

As Steedman points out, however, the acceptability of each of these sentences becomes apparent when one considers them as answers to questions that specifically evoke the required open proposition, for instance, (209a-c) respectively:

(209) a. Which city did each man go to?
b. Which man will give what to whom?
c. What did each man each claim that hedgehogs eat?

Steedman says:

> Indeed, even the most basic gapped sentence, like *Fred ate bread, and Harry, bananas*, is only really felicitous in contexts which support (or can accommodate) the presupposition that the topic under discussion is *Who ate what.* (p. 248)

Gapping interpretation therefore requires that a shared open proposition be recovered from the source clause representation. Further, the contrastive accent placed on the remnants in the target and their parallel elements in the source is responsible for signaling what this open proposition is.

I will deviate from the formulation based on set abstraction that Sag proposed, and assume that the abstraction necessary to create this open proposition is achieved by fronting the appropriate constituents at LF. Necessarily, this process must preserve the linear order in which the constituents appear. Figure 9 shows the syntactic and semantic representations for the source clause of example (207) after the abstraction has taken place. Trace assumptions are left behind that are discharged when combined with their antecedents. The open proposition thus appears within the scope of lambda operators so that the trace variables can be combined compositionally with the elements on which they depend. (As in Chapter 3, these representations are written in a curried notation.) This representational scheme is chosen largely for convenience; the analysis remains compatible with other possible mechanisms and representations.

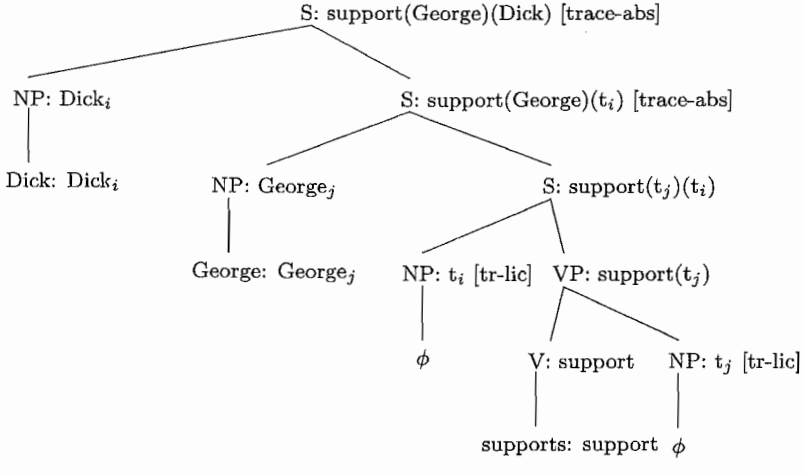


FIGURE 9 Syntactic and Semantic Representations for *Dick supports George* after abstraction.

Target clauses in gapping constructions are likewise represented with the overt constituents abstracted, in this case leaving behind an empty

sentence node. The representation of the target clause in example (207) is as shown in Figure 10. The empty constituent is reconstructed by copying the syntactic structure associated with the parallel sentence node from the source clause representation to the target, along with parallel trace assumptions to be bound within the target. The result of this process is shown in Figure 11. The semantics for the embedded sentence (in this case, $\lambda y, x.support(x, y)$) represents the open proposition that the two clauses share.

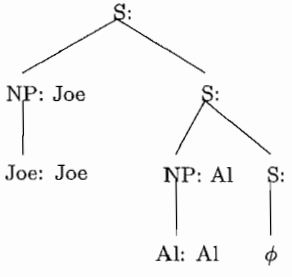


FIGURE 10 Syntactic and Semantic Representations for *Joe, Al.*

As discussed by Sag (1976), representing gapped clauses and their reconstruction in this manner has another significant advantage, in that it avoids the need to posit ambiguous syntactic structures for gapped clauses. That is, in some cases the overt elements in a gapped clause can correspond to any of a number of constituents in the source; consider the following example (due to Quirk et al. (1972, p. 580), cited by Sag (1976)).

(210) Bob will interview some candidates this morning, and Peter this afternoon.

The target clause can have either of the meanings shown in (211a-b), depending on which element in the source is identified as being parallel to *Peter* in the target. In particular, reading (211a) results if *Peter* is interpreted as being parallel to *Bob*, whereas reading (211b) results if *Peter* is interpreted as parallel to *some candidates.*

(211) a. Peter will interview some candidates this afternoon.
b. Bob will interview Peter this afternoon.

If one were to posit an analysis in which the remnants were kept *in situ* (representing all missing material as empty nodes), then a different syntactic representation for the target would be required for each of these

possibilities. This would in turn lead us to the odd conclusion that the set of allowable structures for the target could only be determined by considering where potential parallel elements reside in the source. (Indeed, the syntactic analysis of even unambiguous gapped clauses would require reference to the source.) In the current analysis, gapped clauses are syntactically unambiguous. The ambiguity in pairs such as (211a-b) arises solely from the different possibilities for abstraction in the source representation, determined at least in part from intonation, and resulting in different open propositions shared by the clauses.

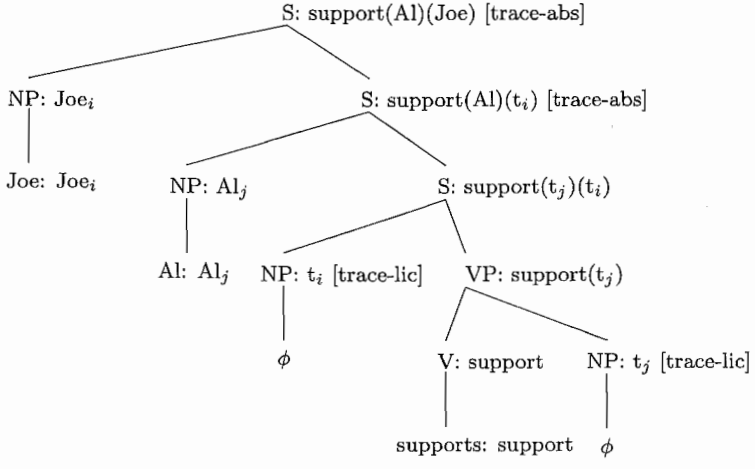


FIGURE 11 Syntactic and Semantic Representations for *Joe, Al* after reconstruction.

Having pinned down the syntactic properties of gapping, I now consider its referential properties. Unlike VP-ellipsis, which we determined to be anaphoric in Chapter 3, the evidence suggests that gapping is not similarly anaphoric. To show this, we can simply apply the same tests that were applied to VP-ellipsis in Chapter 3. First, examples (212a-d) show that gapping cannot refer cataphorically in the way that pronouns and VP-ellipsis can.[13]

[13]It is hard, if not impossible, to isolate a good test case for cataphoric reference in gapping, because of two conflicting constraints: (1) that gapping does not operate within embedded clauses, and (2) embedded clauses are necessary so as to not violate constraints on forward reference, such as those proposed by Lakoff (1968) and Jackendoff (1972). Therefore sentences (212c) and (212d) are not a minimal pair. Nonetheless, the fact that gapping does not operate in embedded clauses is in itself

(212) a. If he makes a statement blasting the press, Al will make a fool of himself. [ cataphoric reference with pronoun ]

b. If George will, Al will make a statement blasting the press. [ cataphoric reference with VP-ellipsis ]

c. * If George the newspaper reporters, Al will make a statement blasting the press. [ cataphoric reference with gapping ]

d. Al will make a statement blasting the press, and George the newspaper reporters. [ standard gapping ]

In addition, and again in contrast to pronouns and VP-ellipsis, gapping cannot locate antecedents from clauses other than the immediately preceding one; consider example (213) in comparison to example (187), repeated below as (214).

(213) A. George made a statement blasting the press. He's going to pay a big price for that.

B. # And Al the newspaper reporters. In his case the fallout will be minimal, however.

(214) A. Jorge is peeling an apple.

B. And Ivan an orange.

To conclude, the evidence suggests that gapping is like VP-ellipsis in that it is associated with an empty node in the syntax, but unlike VP-ellipsis in that it is not anaphoric. Thus, as can be seen from Figure 10, there is no sentence-level semantics for gapped clauses at all until resolution has taken place. I describe how the coherence establishment process determines when and how this semantics is recovered in the next section.

## 4.2 The Analysis Applied to Gapping

The second part of the analysis concerns how the aforementioned properties interact with the properties of the processes that underlie the establishment of coherence. Recall that in the analysis described in Chapter 3, the distribution of VP-ellipsis was predicted from the interaction of two independent phenomena: (i) the anaphoric identification of a referent, and (ii) the process of syntactic reconstruction triggered by the inference mechanisms underlying the establishment of Resemblance relations. It was demonstrated in the previous section that gapping only leaves behind an empty constituent; it is not similarly anaphoric. As such, unlike VP-ellipsis, it has no ability to invoke a mechanism that

evidence that it is not anaphoric.

would be capable of recovering the meaning of the missing material, i.e., there is no analog of item (i) above.

It turns out that these facts alone predict the pattern noted by L&P. Recall that they demonstrated that examples (195a-c), repeated below as (215a-c), are only acceptable when a Parallel relation is operative between the two clauses, and not when a Result relation is operative.

(215) a. Sue became upset and Nan ∅ downright angry.
b. Al cleaned up the bathroom and Joe ∅ the mess.
c. One of the students was accepted at Bryn Mawr and the high school ∅ praised on TV.

Because gapping is associated with an empty constituent, the reconstruction mechanism will be invoked during the establishment of a Resemblance relation (e.g., Parallel) in the same way that it was for VP-ellipsis. A complete syntactic structure will result, from which a semantic interpretation can be realized. No reconstruction is triggered during the process of establishing a Cause-Effect relation, however, and thus the theory correctly predicts that gapping will be infelicitous when such a relation is operative.

Since it is based on syntactic reconstruction, my account predicts that the reconstructed material in the target will be subject to a variety of syntactic constraints, such as binding theory violations. For instance, as with VP-ellipsis, reconstructed material containing reflexives should yield only sloppy readings in the target, in accordance with Condition A. This appears to be the case; whereas sentence (216a) exhibits a strict/sloppy ambiguity (the interpretations of interest here are only those in which George is parallel to Al, that is, Al may have bought a book for George's wife or his own wife), sentence (216b) only gives rise to the sloppy interpretation.[14]

(216) a. George$_i$ bought his$_i$ wife a book on health care, and Al a book on the environment. (strict or sloppy)
b. George$_i$ bought himself$_i$ a book on health care, and Al a book on the environment. (sloppy only)

Likewise, cases in which reconstruction would lead to a Condition B violation are unacceptable. For instance, example (217) is unacceptable with the coindexing indicated, but is perfectly felicitous if *him* in the source clause refers intersententially to someone other than George.

(217) * George$_i$'s mother bought him$_i$ a book on health care, and George$_i$, a book on the environment.

[14]This judgment for (216a) is counter to those of Chao (1987) for similar sentences; she claims that even nonreflexives in gapping result only in sloppy readings.

Finally, as one would expect, examples in which reconstruction would lead to a Condition C violation are likewise unacceptable, as in (218).

(218) * Dick bought George$_i$ a book on health care, and George$_i$ a book on the environment.

Also, in Chapter 3 we found evidence that comparatives with VP-ellipsis do not invoke reconstruction. This fact predicts that gapping is infelicitous in comparatives, which appears to be the case:[15]

(219) # Dick supports George more than Joe, Al.

(220) # George blasted the media before Al the newspaper reporters.

Finally, my analysis predicts the facts noted by Sag (1976) and Hudson (1976) concerning the need for a suitable degree of parallelism between the remnants in the target and their parallel elements in the source. Recall that Sag illustrates this point with sentences (188a-b), repeated below as (221a-b).

(221) a. * Sam hates reptiles, and Sandy to talk to Oh.

b. * Beth ate yogurt, and Norma at midnight.

This fact falls out straightforwardly from the constraints on establishing the Parallel relation in gapping constructions, which require that the abstracted remnants in the target be semantically similar to their parallel elements in the source. In particular, the abstracted elements are the arguments $a_i$ and $b_i$ to the Parallel relation, and the common relation $p$ inferred is the open proposition that the two clauses share. In (221a-b), however, there is no reasonably specific degree of similarity that can be established for the second pair of arguments.

---

[15] Of course, the *pseudo-gapping* construction, in which an auxiliary is overt in the target, is acceptable in comparative and temporal subordination constructions:

(i) Dick supports George more than Joe does Al.

(ii) George blasted the media before Al did the newspaper reporters.

Chris Kennedy (p.c.), however, offers a naturally-occurring example of gapping in a comparative:

(iii) I suspect that they have more to fear from us than we from them. (from "Mars Attacks")

Kennedy also indicates that languages that do not have pseudo-gapping may allow gapping in comparatives.

Jackendoff (1971) discusses cases similar to (iii), but ultimately excludes them from consideration:

> Since a rather free deletion rule is known to be associated with comparative constructions, we will assume that [(iii)] is a special case of comparative deletion rather than Gapping. (Jackendoff 1971, p. 22)

Given all of the questions that these issues raise, I leave them for further exploration.

The fact that this constraint is inherently semantic rather than syntactic can be seen by considering example (222).

(222) ?? Beth ate in the classroom, and Norma at midnight.

Unlike sentences (221a-b), the second remnants in (222) are syntactically parallel. They are not semantically similar, however, and thus the example at best has an awkward or zeugmatic character.

There are two lingering issues that remain to be addressed. First is the question of whether gapping is felicitous when a Contiguity relation is operative. Recall from Chapter 3 that we saw no reason to expect that the establishment of Contiguity would invoke reconstruction, which predicted that VP-ellipsis patterns with cases involving Cause-Effect relations unless a Resemblance relation is also operative. Likewise, the prediction for gapping is that it is unacceptable when a Contiguity relation is operative, unless a Resemblance relation is also inferred.

Once again, the data are not entirely clear; consider (195a-c) again, repeated as (223a-c) below, when interpreted as a 'natural' (but not causal) sequence of events in accordance with the Occasion relation.

(223) a. ? Sue became upset, and (then) Nan ∅ downright angry.
b. ? Al cleaned up the bathroom, and (then) Joe ∅ the mess.
c. ? One of the students was accepted at Bryn Mawr, and (then) the high school ∅ praised on TV.

I find these cases to be odd when keeping to an Occasion interpretation, but judgments vary. I would speculate that the degree to which speakers find such examples acceptable correlates with the extent to which their interpretations favor Parallel over Occasion as the primary means by which the clauses cohere, but this issue requires further investigation.

The second issue stems from the fact that, like previous work, I have been focusing on examples in which the conjunction used is either *and, or,* or *but,* corresponding to the Resemblance relations Parallel and Contrast. The correlation with Resemblance would suggest that gapping should also be acceptable with the other relations in this category discussed in Chapter 2, including Generalization, Exemplification, Exception, and Elaboration. This does not appear to be the case, however, as illustrated by the unacceptability of examples (224a-d).

(224) a. # Gingrich supports Bush, and in general, a politician, his party's presidential candidate. (Generalization)
b. # A politician normally supports his party's presidential candidate, for instance, Gingrich, Bush. (Exemplification)
c. # A politician normally supports his party's presidential candidate, however Guiliani, Cuomo. (Exception)

d. # An aspiring politician was arrested carrying drugs today; in particular, John Smith, cocaine.

A possible explanation for this behavior results from the fact that the accent placed on the remnants and their parallel elements in gapping is contrastive. These four coherence relations, by definition, cannot involve contrast among the parallel elements. In the cases of Exemplification, Generalization, and Exception, the constraints specify that the entities in one clause denote instances (or subsets) of classes denoted by their parallel elements in the other clause, and thus cannot contrast with them. (An element of contrast does come into play for Exception, but only with respect to the predication.) Likewise, the parallel entities are the same in the case of Elaboration, and thus cannot contrast. Therefore, the contrastive accent on parallel entities in gapping that licenses their abstraction in the logical form – the crucial feature that permits gapping to be an allowable form of ellipsis – is inherently ruled out for these relations, and as a result gapping in these contexts is infelicitous.

## 4.3 Interaction of Gapping and VP-Ellipsis

Finally, I address cases in which gapping and VP-ellipsis interact, as exemplified by sentences (225) and (226), adapted from similar examples in Sag (1976, page 291).

(225) Laura supports George, and Tipper, ∅ Al, although she doesn't know why she does.

(226) ?? Laura supports George, and Tipper, ∅ Al, and Mary does too.

Sag's account correctly predicts that sentence (226) is infelicitous, but incorrectly predicts that sentence (225) is also. His account requires that the source and target clauses for VP-ellipsis satisfy his alphabetic variance condition (a syntactic identity constraint) for his deletion operation to be allowable, and in these cases they are not. The representations for sentence (226) are shown in (227), using Sag's representational system.

(227)

$$\{Laura, George\} \subset \hat{x}\hat{y}[x, \lambda r(r\ support\ y)]\ \&$$
$$\{Tipper, Al\} \subset \hat{w}\hat{z}[w, \lambda s(s\ support\ z)]$$
$$...[Mary, \lambda t(t\ support\ him_j)]$$

The lambda expressions in the first two clauses are alphabetic variants, which predicts the acceptability of gapping. However, the expressions in the second and third clauses are not alphabetic variants, which predicts that the VP-ellipsis in each case is infelicitous.

Sag then suggests a weakening of his alphabetic variance condition, with the result that both of the above examples would be predicted to be acceptable. He does not consider any solutions that would predict the different judgments as stated.

The respective felicity and infelicity of examples (225) and (226) are exactly what my analysis predicts. The representation for the second clause in each case after gapping is resolved is given in Figure 12. In example (226), the third clause is in a Resemblance relationship with the second (and the first, for that matter), so the coherence establishment process attempts to retrieve the antecedent syntactic structure and reconstruct it within the target. Reconstruction fails, however, since one of the constituents in the VP has been extracted out of the source, leaving an unbound trace in target. On the other hand, the third clause in example (225) is related to the second by a Cause-Effect relation. Therefore, there is no requirement to reconstruct the syntax of the VP, and the anaphoric resolution of the VP-ellipsis succeeds by only making reference to the sentence-level semantics of the source clause. Thus, the apparent paradox between examples (225) and (226) is just as predicted.

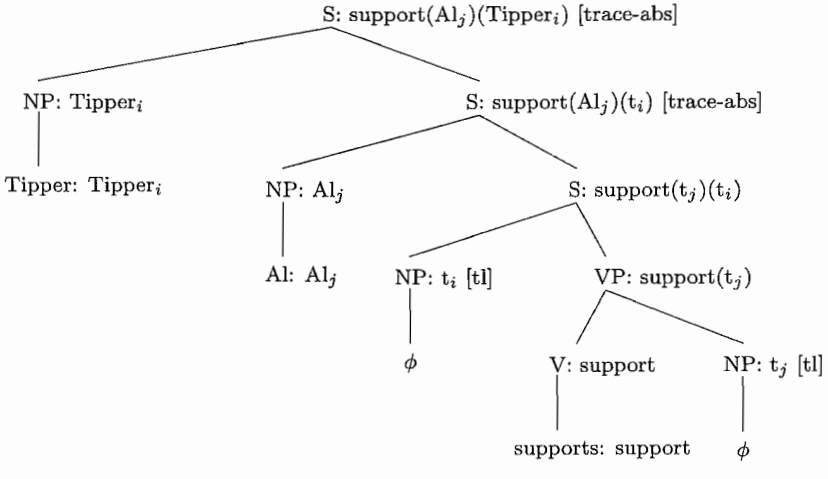


FIGURE 12 Syntactic and Semantic Representations for *Tipper, Al* after reconstruction.

## 4.4 Comparison to Past Work

I have not attempted to offer a comprehensive analysis of all of the facts concerning gapping, and so I will forgo a detailed comparison between my analysis and the long list of syntactic approaches cited at the beginning of this chapter. Because the approach is based on syntactic reconstruction, it predicts that gapping is subject to a variety of syntactic constraints that have been previously noted in these works, such as the binding conditions discussed in Section 4.2. As previously mentioned, however, it is also important that such studies be sensitive to the underlying pragmatic facets of gapping. As Kuno (1976a) and Steedman (1990) have noted (see Section 4.1), many of the sentences cited as ungrammatical in the literature become acceptable within a discourse context which licenses the open propositions they require. Therefore, one must take care to not overly constrain an analysis of gapping by positing new and unwarranted syntactic principles.

One point of contrast with previous approaches is that my account may provide a partial explanation of why gapping does not apply in subordinate clauses, such as in sentences (203a-d). The fact that gapping is also unacceptable with coordinating conjunctions indicating Cause-Effect relations suggests that the purely syntactic split between coordinating and subordinating conjunctions may not ultimately prove to be the best one to make. While I believe the analysis presented here accounts for these cases as well as the cases of syntactic subordination I discussed in a theoretically-motivated and non-stipulatory way, further research is still necessary to determine whether the need to explicitly stipulate a prohibition on gapping in all subordinating contexts can be eliminated entirely.

I will instead compare my account with that of L&P, which is the only analysis of the facts they presented of which I am aware. L&P articulate their analysis within the *ordered entailment* framework of Wilson and Sperber (1979). In that framework, processing a sentence results in a computation of foreground and background entailments. The background entailments are those "presupposed" propositions that result from applying rules that replace focused constituents with variables to the propositional representation of the sentence. Specifically, the *First Background Entailment* (FBE) is the open proposition resulting from replacing a minimal tonically stressed (or clefted) constituent with a variable. For instance, sentence (228a), with the indicated stress on *Bill*, has expression (228b) as its FBE, along with other background entailments (228c) and (228d).

(228) a. BILL'S father writes books.

b. Someone's father writes books.
c. Someone writes books.
d. Someone does something.

As L&P put it, an utterer of sentence (228a) is taking (228b) to be in the hearer's consciousness at speech time (i.e., "given").

L&P postulate the following rule applying to gapping:

(229) Discourse Function of Gapping:
Upon hearing a gapped sentence, a Hearer infers that the Speaker intends that both (all) the conjuncts of the Gapped sentence share a single open proposition as their First Background Entailment, i.e., as that which is appropriately in the hearer's consciousness at that point in the discourse. The open proposition consists of (the representation of) the material deleted in the second (through nth) conjunct, with variables replacing (the representation of) the constituents remaining in the second (through nth) conjunct. The foreground is, of course, the new information.

To see how this principle is used in predicting the facts concerning gapping and causal implicature, consider again sentence (194a), repeated in (230).

(230) Sue became upset and Nan became downright angry.

Because the corresponding elements in both the source and target clauses are contrastively accented, under the symmetric reading the two clauses share the FBE given in (231).

(231) Someone became something. [ open proposition is: $become(X, Y)$ ]

On the other hand, they claim that under the causal reading there are a number of possibilities for what the FBEs are, for example, one set of possibilities for the two clauses in sentence (230) are (232a-b) respectively.

(232) a. Something happened.
b. Nan did something.

Because two FBEs are required for causal implicature, the Discourse Function of Gapping rule accounts for why gapped sentences do not yield causal implicatures.[16]

My analysis improves upon the L&P account in several respects. First, where L&P have to stipulate the Discourse Function of Gapping

[16]It seems that both clauses could still share the same FBE in a causal implicature, specifically one of the form *something happened*. But even in this case, the Discourse Function of Gapping rule is not satisfied, since this open proposition does not contain two variables standing proxy for the non-elided constituents.

rule, in my account the behavior of gapping is the result of the interaction of more fundamental and independently-motivated aspects of language interpretation. Second, the data is accounted for with an analysis that is unified with one of VP-ellipsis, a form which does not share the same pragmatic effects with gapping. Finally, I have extended the argument to show why gapping succeeds or fails in cases in which a variety of other conjunctions participate. However, both accounts share the property that the data are handled by appeal to discourse-level factors, and not only surface-syntactic facts.

As a final note, in Chapter 3 I discussed two works that address VP-ellipsis resolution in the context of broader theories of discourse structure and coherence, particularly those of Prüst (1992) and Asher (1993). Prüst addresses gapping, but does not acknowledge the infelicity of gapping with Cause-Effect relations, and therefore provides no account for it. Furthermore, it appears that neither Asher nor Prüst can account for the mixed gapping/VP-ellipsis cases discussed in Section 4.3.

## 4.5 Conclusion

In this chapter, I have provided an analysis of gapping that accounts for the facts noted by Levin and Prince (1982, 1986), as well as additional facts relevant to examples involving conjunctions other than *and*. The correct predictions result from applying the account of VP-ellipsis presented in Chapter 3 without modification; the differences between gapping and VP-ellipsis result primarily from the independently-motivated fact that gapping is not anaphoric whereas VP-ellipsis is. The analysis was also shown to handle the cases of mixed gapping/VP-ellipsis constructions noted by Sag (1976). The account is compatible with and extends previous syntactic accounts of gapping, although it may ultimately provide more adequate divisions among the data than the purely syntactic distinctions sometimes stipulated in those studies.

# 5

# Coherence and Extraction

In the previous two chapters, I discussed different linguistic forms a speaker can select to denote an eventuality that has already been introduced into the discourse. I investigated the contexts in which these alternatives can be felicitously employed, showing that the answer requires an appeal to the type of coherence relation that is operative between the clauses.

In this chapter, I address another phenomenon relating to choice of syntactic form, although unlike the previous phenomena, it does not involve ellipsis. In particular, I investigate the contexts in which a speaker can use a syntactic structure involving *extraction*. (My use of the term "extraction" for this phenomenon is for historical purposes, and is not meant to imply that such sentences are derived by an explicit movement operation.) Examples of such structures are given in sentences (233a-d), in which the noun phrase *the magazine* has been extracted from its canonical position as sentential object.

(233) a. This is the magazine which John bought. (relative clause)
b. Which magazine did John buy? (wh-question)
c. It is this magazine that John bought. (it-cleft)
d. This magazine, John bought. (topicalization)
e. John bought, and Bill read, this magazine. (right node raising)

In keeping with my focus on linguistic phenomena that apply interclausally, I will be concerned with extraction from coordinate clauses. As is well known, the interaction between extraction and coordination led Ross (1967) to posit one of the most commonly cited constraints in syntactic theory, the so-called *Coordinate Structure Constraint* (CSC). I begin this chapter with a review of the CSC, as well as some known counterexamples to it. While previous studies have often relegated these examples to the linguistic 'periphery', I find that, as with the other phenomena discussed in this book, a pattern emerges which corresponds to

my neoHumian categorization of coherence relations. Once again, a set of apparently contradictory data will be shown to be perfectly consistent when the type of coherence relation is taken into account. I, as have several authors before me, therefore hope to convince the reader that there is no CSC in universal grammar. Instead, the data supporting this purported constraint arise from independently motivated factors that apply in only a particular subset of the possible scenarios that involve extraction from coordinate structures.

## 5.1 The Coordinate Structure Constraint

Ross (1967) proposed the CSC as a basic constraint in universal grammar:

> In a coordinate structure, no conjunct may be moved, nor may any element contained in a conjunct be moved out of that conjunct.

Grosu (1973) makes a convincing case, which has been commonly adopted by researchers since, that two components of the CSC should be differentiated: the *Conjunct* Constraint and the *Element* Constraint. The Conjunct Constraint bars the movement of whole conjuncts out of coordinate structures, ruling out sentences such as (234).

(234) * This is the magazine which John bought the book and.

The Conjunct Constraint is extremely robust; none of the examples I will be discussing in this chapter are of this type. This constraint has been argued to result from independently motivated constraints in several theories of grammar; for example, Ross (1967) claims that it results independently from his A-over-A Principle, and Pollard and Sag (1994) from their Trace Principle. (But cf. Johannessen (1998), who says that even the Conjunct Constraint has principled exceptions.)

I instead address facts concerning the Element Constraint, which bars the movement of elements contained within a conjunct as opposed to the conjunct itself. The Element Constraint rules out sentences such as (235a-b), because extraction has taken place out of a conjoined verb phrase (VP).[17]

(235) a. * What book did John buy and read the magazine?
b. * What magazine did John buy the book and read?

Past researchers have often assumed that the Element Constraint is a valid generalization, and typically have sought to explain it solely at the

[17]When citing examples discussed by previous authors, I will mark examples in accordance with the judgments that these authors provided. See also footnote 19.

level of syntactic representation. For instance, Schachter (1977a) argues for his Coordinate Constituent Constraint (CCC) based on the fact that it predicts the CSC along with its so-called 'across-the-board' exceptions (discussed below). Likewise, it has been considered a success of the Generalized Phrase Structure Grammar (GPSG) analysis of extraction phenomena (Gazdar et al. 1985) that the CSC and the across-the-board exceptions are predicted from the interaction of independently motivated mechanisms in the grammar (e.g., Slash Categories, the Foot Feature Principle, the Head Feature Convention). Similarly, Steedman (1985) has considered the same property to be a feature of his proposed extensions to Categorial Grammar, as has Goodall (1987) with respect to his formulation of the union of phrase markers in coordination.

The thesis of this chapter is that the CSC data should not be accounted for using rules of grammar. The data discussed in the next section motivates this position.

## 5.2 CSC Violations and Their Treatments

A variety of interesting CSC data have been brought to light over the course of the last several decades. In this section, I will introduce this data in more or less the order in which the works that discussed them appeared. (An exception is Kuno's (1976b) analysis, which will be discussed later.) This will not only allow all of the data to be discussed in one place, but will also offer a sense of history of the controversy that has centered around the Element constraint. I leave the problem of sorting through this data to the second half of the chapter.

**Ross's Violations** No sooner did Ross propose the CSC than did he begin listing counterexamples to it. Sentence (236a) is an example which allows extraction to occur out of a single conjunct, as shown in (236b).

(236) a. I went to the store and bought some whiskey.
b. Here's the whiskey which I went to the store and bought.

Syntactically, sentence (236b) looks a lot like (235b) in the relevant respects, yet only the former is acceptable. However, this acceptability fades away when one considers versions of (236a-b) that involve sentence-level instead of VP-level conjunction, as shown in (237a-b).

(237) a. I went to the store and Mike bought some whiskey.
b. * Here's the whiskey which I went to the store and Mike bought.

Ross claims, citing arguments by George Lakoff, that "there are clear indications that the relative clause in [(236b)] is not an instance of ordinary sentence conjunction". First, he claims that such constructions

do not allow the main verb of the second conjunct to be stative, per (238a-b).

(238) a. Tony has a Fiat and yearns for a tall nurse.
b. * The tall nurse who Tony has a Fiat and yearns for is cruel to him.

Second, examples (239a-b) suggest that the second conjunct cannot be negative:

(239) a. I went to the movies and didn't pick up the shirts.
b. * The shirts which I went to the movies and didn't pick up will cost us a lot of money.

Third, there appear to be restrictions on the tenses used; compare (236a-b) with (240a-b).

(240) a. I went to the store and have bought some excellent whiskey.
b. * The excellent whiskey which I went to the store and have bought was very costly.

Because example (236a) can be paraphrased with a purpose clause as in (241), and (237a)–(240a) cannot be analogously paraphrased, Ross suggests that (236a) is not an instance of true conjunction but is instead derived from the underlying structure of (241).

(241) I went to the store to buy some whiskey.

The second major set of exceptions discussed by Ross include those resulting from the "across-the-board" application of extraction, in which the same element is extracted from all conjuncts as shown in sentence (242).

(242) What book did John buy and read?

This class of counterexamples is quite broad; for instance, extraction is allowable across-the-board in examples analogous to (237)–(240).

(243) a. This is the whiskey which I bought and Mike drank.
b. It's a Fiat that Tony owns and yearns to drive.
c. What shirts did you drop off but not pick up?
d. This is the whiskey that Mike wants and I have bought.

Finally, Ross presents a final class of cases which Na and Huck (1992) refer to as 'idiomatic conjunctions', exemplified in pairs (244–246).

(244) a. She's gone and ruined her dress now.
b. Which dress has she gone and ruined now?

(245) a. I've got to try and find that screw.

b. The screw which I've got to try and find holds the frammis to the myolator.

(246) a. Aunt Hattie wants you to be nice and kiss your granny.

b. Which granny does Aunt Hattie want me to be nice and kiss?

As Na and Huck note, these counterexamples differ from the others Ross cites in that they do not describe divisible eventualities. That is, example (244) does not mean that two distinct events have occurred, a 'going' event and a 'dress ruining' event. The same is true for (245) and (246). Since these examples do have an idiomatic quality, and my focus is on examples that involve two or more eventualities and the type of relationship that holds between them, I will not say anything further about examples in this class. (But see Schmerling (1975) and Na and Huck (1992) for further discussion.)

Ross claims that the CSC can be used as a criterion for coordinate structure, although this embodies an obvious circularity given his claim that the counterexamples are not cases involving true conjunction (see also Na and Huck (1992)). Interestingly, Ross demonstrates his test by considering whether a sentence that has undergone gapping is still conjoined, concluding that it is on the basis of the inability to extract from a single conjunct:

(247) a. The boy works in a skyscraper and the girl in a quonset hut.

b. * Which boy works in a skyscraper and the girl in a quonset hut?

Recall that it was shown in Chapter 4 that although all gapped examples involve conjunction, gapping only succeeds when the operative coherence relation is an instance of Resemblance. As we will see later, these are just the cases in which extraction has to occur across-the-board, and thus we might conclude that Ross's test actually discriminates between the types of relation that can hold between the conjuncts.

In sum, Ross posits the existence of the CSC as a fundamental constraint of grammar, maintaining only the across-the-board proviso. He maintains that apparent counterexamples such as (236b) do not involve true conjunction, and are instead derived from a non-conjoined source.

**Schmerling** Schmerling (1972) presents a convincing argument against Ross's reanalysis account of examples like (236a-b). Ross's argument centered around a claim that example (248a) is synonymous with (248b); the reason that (248c) is acceptable is therefore due to the fact that (248d) is.

(248) a. I went to the store and bought some whiskey.

b. I went to the store to buy some whiskey.
c. Here's the whiskey which I went to the store and bought.
d. Here's the whiskey which I went to the store to buy.

As Schmerling shows, however, there is a variety of evidence that establishes that examples (248a) and (248b) are not synonymous. For instance, sentence (249b) expresses a contradiction, whereas sentence (249a) does not.

(249) a. I went to the store to buy some whiskey, but the sales clerk persuaded me to buy Ripple instead.
b. * I went to the store and bought some whiskey, but the sales clerk persuaded me to buy Ripple instead.

Furthermore, the acceptability of *and* in (250a), and the contradiction arising from the use of *to* in (250b), shows the former is not expressing a purpose relation and thus the acceptability of the extraction in (250c) cannot be due to such a relation.

(250) a. I came home and read the latest issue of Rolling Stone, even though I had intended to do the laundry.
b. * I came home to read the latest issue of Rolling Stone, even though I had intended to do the laundry.
c. The latest issue of Rolling Stone is what I came home and read, even though I had intended to do the laundry.

Schmerling (1975) also follows up on Ross's CSC observations in her study of the differences between *symmetric* and *asymmetric* conjunction, noticing that this distinction impacts the ability to extract from coordinate clauses. (Recall this terminology from the discussion of Levin and Prince (1986) in Chapter 4.) She adds examples (251) and (252) to Ross's data, saying that all of these apparent counterexamples can be plausibly argued not to involve what she terms 'logical' conjunction.

(251) a. Lizzie Borden took an axe and gave her mother forty whacks.
b. Who did Lizzie Borden take an axe and whack to death?

(252) a. Roy called a secret meeting and offended Bob and Jeff.
b. ? When was that meeting that Roy called and offended Bob and Jeff?

Furthermore, recall that Ross noticed that when the conjunction in example (236a) is not reduced to the VP level, as in (237a), the ability to extract from the coordinate structure is no longer available. Schmerling argues that this difference is accompanied by a change in the symmetry of the conjunction; for instance, compare the minimal variants given in examples (253)–(255).

(253) a. I went to the store and I bought some whiskey.
b. * This is the whiskey which I went to the store and I bought.

(254) a. Lizzie Borden took an axe and she gave her mother forty whacks.
b. * Who did Lizzie take an axe and she whack to death?

(255) a. Roy called a secret meeting and he offended Bob and Jeff.
b. * When was that meeting that Roy called and he offended Bob and Jeff?

The implicatures that arise in the original versions do not also arise in these variants. For example, sentence (253a) does not imply that the whiskey was bought at the store as sentence (236a) does, and sentence (254a) could be used to describe two unrelated events, unlike (251a).

Finally, Schmerling (1972) adds a case involving causation to the list of types of exceptions to the CSC, particularly example (256).

(256) Spiro told a little joke and infuriated Paul.

She notes that while it is possible to extract from the first conjunct, extraction from the second is impossible, as shown in (257a-b), judgments hers.

(257) a. ?? What was the joke that Spiro told and infuriated Paul?
b. * Paul is the guy who Spiro told a little joke and infuriated.

I return to cases such as these in my discussion of Lakoff's contributions below.

**Goldsmith** Goldsmith (1985) added another class of cases to the list of CSC counterexamples, pointing out that extraction out of a single conjunct can occur when what he calls the "nonetheless" use of *and* is operative between the conjuncts. This meaning can be roughly paraphrased by *and still, and nonetheless*, or *and yet*, as seen in examples (258a-e).

(258) a. How many courses can we expect our graduate students to teach and (still) finish a dissertation on time?
b. How many counterexamples can the Coordinate Structure Constraint sustain and still be considered empirically adequate?
c. How many lakes can we destroy and not arouse public antipathy?
d. Who is the most incompetent member the Commission can nominate and still preserve face in the international community?

e. How much can you drink and not end up with a hangover the next morning?

Goldsmith points out certain characteristics of these examples that appear to facilitate the felicity of violating the CSC. First, he claims that extraction is most felicitous when a scalar quantity is involved; he considers examples (259) to be marginal compared to (258).[18]

(259) a. What can you drink and not end up with a hangover the next morning?

b. Who can we nominate and still preserve face in the international community?

c. Who can this country elect and still survive?

d. Which one can we take and not get caught?

e. What king of music can you listen to and still get your work done?

The second fact is that the second conjunct must be a bare VP, and not a full infinitival VP.

(260) a. How many courses can we expect our graduate students to teach and still write a decent dissertation?

b. * How many courses can we expect our graduate students to teach and to still write a decent dissertation?

c. * How many courses can we expect our graduate students to teach and still to write a decent dissertation?

Goldsmith claims, correctly in my view, that this behavior is more generally a property of the 'despite' sense of *and* rather than a limitation on extraction. He expresses the intuition that the complement represents a single mental representation when bare infinitives are coordinated, whereas two mental representations are being reported when the infinitival VPs are coordinated. This intuition is basically the one that Schmerling had concerning the difference between examples (253–255) and their reduced counterparts; only the fully reduced forms are typically understood with an asymmetric meaning, otherwise the symmetric meaning is favored.

A third difference that Goldsmith finds is that the VPs must describe activity-types and not specific actions that took place at a given place and time, claiming that the 'despite' sense is greatly enhanced by this difference.

[18]Goldsmith admits that the judgments are extremely subtle, and in fact I find all of these cases to be acceptable.

(261) a. How many courses can we expect our graduate students to teach and still finish their dissertations on time?
b. * How many courses did Mrs. Sykes teach last year and still finish her dissertation on time?

Goldsmith concludes that the CSC "requires syntactic parallelism just in case the semantics also presents its own semantic parallelism". He nonetheless posits a syntactic reanalysis account, in which *and* is reanalyzed from a syntactic coordinator to a syntactic subordinator. Examples like (258a-e) would be immune to the CSC if this were the case, since conjunction is no longer involved. (This explanation presumably applies only to cases involving the 'despite' reading of *and*, however, and thus does not address the other types of counterexample to the CSC.) However, as I argued in Chapter 3 with respect to Hestvik's (1995) similar claim about strict readings with reflexives in VP-ellipsis, independent motivation is required for treating *and* as a subordinating conjunction. Again, the constructions in question to not pass the most basic discriminating test for subordination – the ability to front a clause headed by a subordinating conjunction – as can be seen by comparing *and* with *because* in examples (262a-d).

(262) a. We can expect our graduate students to finish a dissertation on time, because they only teach one course.
b. Because they only teach one course, we can expect our graduate students to finish a dissertation on time.
c. We can expect our graduate students to teach one course and still finish a dissertation on time.
d. * And still finish a dissertation on time, we can expect our graduate students to teach one course.

Thus, Goldsmith's interesting class of counterexamples to the CSC, like others previously presented, do not readily admit of a purely syntactic explanation.

**Lakoff** Lakoff (1986) synthesizes much of the data I have discussed in presenting what is perhaps the most pointed argument to date that a purely syntactic CSC does not exist. He categorizes the different counterexamples in terms of three 'scenarios'. The 'Type A' scenario is illustrated by examples like (263), an adaptation of (236b), in which the clauses can be interpreted as a natural, and expected, course of events.

(263) What did Harry go to the store and buy?

In contrast, the 'Type B' scenario is illustrated by examples like (264), an adaptation of Goldsmith's (258e), which expresses a course of events

that runs counter to conventionalized expectation.

(264) How much can you drink and still stay sober?

Finally, 'Type C' scenarios are illustrated by examples (265a-b), in which a cause-effect relationship holds between the clauses. Recall that Schmerling (1972) had also noted cases of this type.

(265) a. That's the stuff that the guys in the Caucasus drink and live to be a hundred. (attributed to Peter Farley)

b. That's the kind of firecracker that I set off and scared the neighbors. (attributed to William Eilfort)

Lakoff provides another piece of evidence against previous attempts to explain away such examples as involving something other than 'true conjunction', such as Ross's reanalysis of the conjoined clause in (263) to an *in order to* adverbial and Goldsmith's reanalysis of the conjoined clause in (264) to a *despite* adverbial. The argument appeals to examples in which extraction takes place from a subset of more than two conjoined phrases as in (266a-b).

(266) a. What did he go to the store, buy, load in his car, drive home, and unload?

b. How many courses can you take for credit, still remain sane, and get all A's in?

The fact that there are more than two conjuncts, marked by the comma-intonation sequence and final *and* that is characteristic of conjunction, would appear to preclude reanalysis to a structure only capable of relating two clauses. Lakoff concludes, correctly in my opinion, that the CSC simply does not exist as an independent constraint in natural language syntax. The unacceptable cases should instead be explained as violations occurring at the semantic or pragmatic levels of language processing.

The data nonetheless appear to have constraints at work which vary with the different scenario types. In contrast to the fact that Type B and C scenarios do not generally allow extraction to occur from the final conjunct, Type A scenarios appear to require such extraction. This can be seen by considering the minimal pair shown in (267a-b). Sentence (267a) is a Type A scenario – the semantics does not support a denial of expectation reading – and extraction that does not include the final conjunct is unacceptable. In contrast, the semantics of example (267b) supports the denial of expectation reading characteristic of a Type B scenario, and thus such extraction is acceptable.

(267) a. * How big a meal did he eat and feel satisfied?

b. How small a meal can you eat and feel satisfied?

Lakoff does not discuss constraints on extraction in Type B or C scenarios, but such examples appear to nonetheless resist extraction out of only the second conjunct. This fact was noted by Schmerling (1972) for her example (256), as shown in (257); consider also (268a-b).

(268) a. ?? How small of a hangover can you drink a six pack and still avoid getting?

b. ?? Those are the neighbors that I set off a firecracker and scared off.

Lakoff provides a preliminary sketch of how these data might be explained in terms of 'predication principles' within a Fillmorian theory of frame semantics, although further details of these principles and independent justification for their existence would be necessary to evaluate his proposal. In general, however, his proposal appears to be compatible with an explanation stated in terms of the discourse-level coherence relationship that is operative between the conjuncts, which I will offer later in the chapter.

**Deane** Deane (1991) attempts to make the extraction conditions in Lakoff's scenarios more precise. Instead of addressing the issue at the level of scenarios, he presents a classification of conjuncts themselves. With respect to A-scenarios, he states:

> The conjunct types which need not submit to across-the-board extraction possess special functions within a larger narrative frame, for instead of describing the main narrative sequence they provide explanations and background. (p. 23)

Deane breaks down these 'special functions' into several types. The first are *preparatory actions*, or actions which form part of an established routine for accomplishing some other action. Examples include (269a) and (270a):

(269) a. He went to the store and bought something.

b. What did he go to the store and buy?

c. ?*What store did he go to and buy groceries?

(270) a. He picked up the phone and called someone.

b. Who did he pick up the phone and call?

c. * What did he pick up and call me?

Deane categories most of Lakoff's A-scenario examples into this class. Lakoff's claim that A-scenarios require extraction from the final conjunct is then explained by the fact that the final conjunct is the main action in such cases. Hence we get the ability to extract the second conjunct

without the first in (269b) and (270b), and the lack of ability to extract the first without the second in (269c) and (270c).

Deane's second class of conjuncts that serve special functions includes *scene-setters*, which describe the scene in which an event takes place. Examples include (271a-b).

(271) a. Sam is not the sort of guy you can just sit there and listen to.

b. Who did you stand in the parlor and tell jokes about?

His third class of conjuncts includes *internal causes*, which describe an internal state which causes an agent to perform a subsequent action. Examples include (272a-b).

(272) a. Which problem did he get bored and give up on?

b. What did he lose his balance and fall on top of?

His fourth class of conjuncts includes *incidental events*, which are included to provide incidental details but which do not form part of the main narrative sequence. Examples include (273a-b).

(273) a. This is the sort of brandy that you can sip after dinner, watch TV for a while, sip some more of, work a bit, finish off, go to bed, and still feel fine in the morning.

b. This is the kind of job that you can work on all morning, take a lunch break, and finish off by 2 p.m.

Deane notes that incidental events differ from the other conjunct types in that they tend to appear between more important events in the narrative sequence, whereas the other aforementioned types occur before a main event conjunct. This leaves open the case in which such conjuncts occur after the main event conjunct. This is just what occurs in Lakoff's B- and C-scenarios.

To account for the CSC data, Deane posits that NPs that are highly activated or focal are those which must be extracted, whereas the conjuncts which resist extraction are not focal because of these special narrative functions. In the 'canonical' cases of across-the-board extraction each phrase is equally important, so extraction from one without the other would contradict the equal emphasis that is normally considered to be inherent in such structures. In the case of A-scenarios, on the other hand, the (discourse) subordinate role that certain conjuncts play makes them inherently unequal in importance, and thus no contradiction arises from failing to extract from them. Finally, the main event typically occurs first in the case of B- and C-scenarios, and thus normally extraction occurs from the first conjunct.

**Na and Huck** Na and Huck (1992) analyze CSC data in English and Korean and conclude that the CSC "is not simply a constraint on the syntactic form of conjunctions in English, but rather to a large extent a consequence of apparently universal limitations on filler-gap relationships in discourse-level structures". They characterize the constraints on extractability in terms of the notion of 'primacy' of a clause. A clause is *primary* if it cannot be deleted without seriously distorting the message expressed, otherwise the clause is *secondary*. (This difference is reminiscent of the 'nucleus' and 'satellite' distinction in the Rhetorical Structure Theory analysis of coherence (Mann and Thompson 1987), which we briefly described in Section 2.2.) For instance, in sentence (274), deleting either of the conjuncts would distort the message.

(274) Bobbie writes novels and raises goats.

In contrast, the central message in Ross's sentence (275) would remain intact even if the first conjunct was removed.

(275) I went to the store and bought some whiskey.

Na and Huck use the discourse-relative term *coordinate* to refer to cases in which each conjunct carries equal weight (i.e., each is primary); cases with one or more secondary conjuncts are thus *noncoordinate*. They couple this discourse-driven distinction with the logic-driven dichotomy between symmetric and asymmetric conjunction to classify examples into three types. Example (274) is an example of a symmetric coordinate, whereas example (275) is an asymmetric noncoordinate. The category of asymmetric coordinate is exemplified by (276a-c).

(276) a. I left the door open and the cat got in.
b. I had suspected that the solution would turn out to be elusive, and I was right.
c. Joan sings ballads and accompanies herself on the guitar.

Na and Huck categorize all of Lakoff's Type B and C examples as asymmetric noncoordinates. In these cases, exemplified by (277a-b) respectively, the first conjunct is primary and the second is secondary.

(277) a. You can eat as much beansprouts as you want and not get fat.
b. Babies always eat that and then get sick.

That is, sentence (277a) is primarily about how much beansprouts one can eat and not that you cannot get fat, and sentence (277b) is about babies always eating something and not that they get sick. In these cases, the second clause is regarded as an aftermath of the first. This contrasts with Type A scenarios, in which secondary conjuncts generally set the stage for, or otherwise modify, primary conjuncts that appear

subsequently. For example, in sentence (275), going to the store sets the scene for buying whiskey, thus the second clause is primary and the first secondary.

Finally, Na and Huck exclude a class of cases that manifest 'weak' cause-effect relationships from the set of Type C scenarios, because both clauses are primary. Examples are shown in (278a-c).

(278) a. Jean went off to Las Vegas for the weekend and cannot afford to get herself a new rug.

b. I rented out the cottage for the summer and don't have a place to stay in the area myself.

c. I slept well last night and don't have a headache anymore.

They claim that the 'and' in these cases is perhaps best paraphrased as 'and now' rather than 'and as a result'. Furthermore, unlike Type C scenarios, the meanings do not change substantially when they involve sentence coordination rather than VP coordination, as in (279a-c).

(279) a. Jean went off to Las Vegas for the weekend and she cannot afford to get herself a new rug.

b. I rented out the cottage for the summer and I don't have a place to stay in the area myself.

c. I slept well last night and I don't have a headache anymore.

Based on this data, Na and Huck propose the following condition:

> Condition on Asymmetric Conjunction (CAC): In any asymmetrical conjunction, if extraction is performed on a secondary conjunct, it must be performed across-the-board.

This condition thus captures the fact that extraction is permitted from only the second, but not only the first, conjunct in Type A scenarios (examples (280a-b)), whereas this pattern is reversed in Type B and C scenarios (examples (281a-b) and (282a-b)).

(280) a. Where's the Coors that Al just went to the store and bought?

b. * By which route did he go and buy the liquor at the store?

(281) a. How many lakes can we destroy and not arouse public antipathy?

b. * What can we destroy many lakes and not arouse?

(282) a. Which dish is it that people always order in this joint and then get sick?

b. * How sick do people order that chili dish here and then get?

Na and Huck conclude that this data provide "strong support for the CAC", having previously indicated that the CAC "predicts correctly" the extraction patterns. This way of looking at it is not quite accurate, since the CAC was designed to summarize the descriptive facts concerning extraction from coordinate clauses. That is, since it was not motivated on independent grounds, it cannot be said to *predict* the data in any meaningful sense. Furthermore, since the CAC only refers to asymmetric conjunction, it must be coupled with a separate constraint covering symmetric conjunction. Given this, a different and much simpler statement of the facts seems to be possible, one that need not make reference to the type of conjunction used. Simply put, the constraint is that extraction out of secondary conjuncts is optional. The across-the-board constraints in examples employing symmetric conjunction follow from the fact that all conjuncts in these cases are primary.

My characterization of the constraint is not identical to Na and Huck's CAC; the latter requires that extraction must occur across-the-board if it occurs from even one secondary conjunct whereas my rule does not. To distinguish these two formulations, we must look at examples that contain more than two conjuncts. Consider example (283), adapted from an example in Lakoff (1986).

(283) This is the kind of brandy that you can sip after dinner, watch TV for a while, sip some more of, read an article, finish off, go to bed, and still feel fine in the morning.

The second, fourth, sixth, and seventh conjuncts are secondary, and thus extraction is not required to take place from any of these. However, example (284) shows that extraction can nonetheless take place from one without the others, contradicting the CAC.

(284) This is the kind of brandy that you can sip after dinner, watch TV for a while, sip some more of, read the label of, finish off, go to bed, and still feel fine in the morning.

The fourth conjunct in (284), *read the label of*, is no more primary than the fourth conjunct in (283), *read an article*. Nonetheless, extraction can take place out of this conjunct without necessitating that extraction apply across-the-board.

Despite these issues, Na and Huck's analysis expresses an important insight. While previous researchers have argued against the CSC by shifting the realm of inquiry to semantic or pragmatics, Na and Huck capture the facts in terms of notions pertaining to discourse processing and understanding. The analysis I present will attempt to do the same, specifically in terms of the interaction between independently motivated

constraints on extraction and the properties of coherence relations. I discuss the coherence-level properties of the CSC data in the following section.

## 5.3 A Pattern in the Data

As can be seen from the foregoing discussion, a dizzying array of data, insights, and analyses have been presented in the literature that identify and address problems with the CSC. The reader who has followed the analyses in this book to this point, however, may well have identified an additional pattern that emerges from these data. In particular, these data admit of a fairly clean and straightforward categorization in terms of my NeoHumian trichotomy of coherence relations. This categorization, in fact, requires only a minor adaptation to the distinction between scenario types that Lakoff proposed.

The first class is exemplified by sentences (235) and (242), repeated as (285a-b) below.

(285) a. * What book did John buy and read the magazine?
b. What book did John buy and read?

In these examples the Resemblance relation *Parallel* is operative between the two conjuncts. In such cases, extraction is required to occur across-the-board.

The second class is exemplified by sentences (264) and (265a-b), repeated below as (286a-c).

(286) a. How much can you drink and still stay sober?
b. That's the stuff that the guys in the Caucasus drink and live to be a hundred.
c. That's the kind of firecracker that I set off and scared the neighbors.

In these examples the Cause-Effect relations *Violated Expectation* (example (286a)) and *Result* (examples (286b-c)) are operative. In these cases, extraction can occur out of the first (primary) clause without also occurring out of the second (secondary) clause. Hence, I have grouped together Lakoff's Type B and C scenarios, since their coherence is established using the same basic type of inference process.

Finally, the third class of cases is exemplified by sentence (236b), repeated below as (287).

(287) Here's the whiskey which I went to the store and bought.

In this example the Contiguity relation *Occasion* is operative. In such cases, extraction need not occur out of scene-setting clauses or others

that perform a supporting function. This constraint tends to rule out extraction from the final clause, since as Deane noted, supporting clauses of this sort tend to come either before or between conjuncts describing events that are central to the narrative. These are the cases that Lakoff called Type A scenarios.

Viewing the data in light of this categorization, there does not appear to be much left to argue for the CSC. In none of the three categories is extraction from a coordinate structure barred entirely; instead, there appear only to be weaker constraints at play that differ with respect to the type of coherence relation. This suggests that the data regarding violations of the CSC would be better explained as a result of principles concerning coherence establishment processes, rather than stipulated as part of a theory of grammar.

As with all the phenomena discussed in this book, my explanation of these facts derives from the interaction between two aspects of language interpretation: the properties of the linguistic phenomenon at hand, in this case extraction, and the manner in which these properties interact with the inference processes underlying the establishment of my three types of coherence relations. I discuss the first of these in the next section, with particular emphasis on *topichood* constraints that have been posited to apply to extracted elements. I then follow with a discussion of the role that topichood plays with respect to the establishment of different types of coherence relations.

## 5.4 Topichood Constraints on Extraction

Kuno (1976b, 1987) presents an assortment of intriguing data that suggests the existence of some type of topichood constraint on extraction. In his earlier paper he proposes a thematic constraint that applies to relative clause formation:

> The Thematic Constraint on Relative Clauses: A relative clause must be a statement about its head noun. (Kuno 1976b)

In his later book, he expands the constraint to apply to extraction more generally, requiring that the extracted element be able to serve as the *topic* of the clause (in some appropriate sense to be discussed below) from which it is extracted:

> *Topichood Condition for Extraction:* Only those constituents in a sentence that qualify as the topic of the sentence can undergo extraction processes (i.e., *WH*-Q Movement, *Wh*-Relative Movement, Topicalization, and *It*-Clefting). (Kuno 1987, page 23)

These constraints are posited to account for extraction effects that appear to be purely semantic or pragmatic in nature. For instance, first consider the fact that sentences (288a-b) are both perfectly natural sentences of English.

(288) a. I read a book about John Irving.
b. I lost a book about John Irving.

In contrast, sentences (289a-b), which are variants of (288a-b) in which extraction has taken place, are not equally acceptable.

(289) a. Who did you buy a book about?
b. ?? Who did you lose a book about?

In particular, sentence (289b) seems markedly odd. This difference belies the fact that sentences (289a-b) are identical syntactically, in fact the only difference between them is the verb. One must then ask what is responsible for this difference in acceptability.

As Kuno notes, there is an intuitive difference between the two cases with respect to the centrality of *John Irving.* He says:

> In a highly intuitive sense, we feel that the fact that the book under discussion was on John Irving is much more *relevant* in [(288a)] than in [(288b)]. This is undoubtedly due to the fact that one buys books, but does not lose them, because of their content. (Kuno 1987, p. 23)

Of course, we are still left with the question of how exactly to define the notion of topic. Although he does not provide an answer, Kuno (1976b) does offer tests that one can apply to try to get at the notion (see also Reinhart (1982)). One of these is place the phrase "Speaking of X" at the beginning of the sentence, in which X is the potential topic being tested, and to pronominalize the mentions of X within the sentence.

(290) a. Speaking of John Irving, I just read a book about him.
b. ?? Speaking of John Irving, I just lost a book about him.

Kuno (1976b) then addresses the CSC data by claiming that the ungrammaticality of Ross's examples are due to this constraint, thus eliminating the need to separately stipulate the element constraint of the CSC. For instance, in (291),

(291) Henry plays the lute and sings madrigals.

he claims that it is not possible to regard *the lute* as representing the theme of the entire sentence; that is, while it might be about the lute and madrigals, it cannot be understood as just being about the lute. As

a result, it is infelicitous to extract only this constituent, as shown in (292).

(292) * The lute which Henry plays and sings madrigals is warped.

Kuno contrasts sentence (291) with sentence (293).

(293) Mary bought an organ and thereby angered her husband.

Unlike *the lute* in (291), *the organ* can be the topic of the conjoined clauses in (293) because the result of Mary's buying it is still relevant to it. Therefore, it is felicitous to extract *the organ* in such circumstances, as shown in (294).

(294) This is the kind of organ that Mary bought and thereby angered her husband.

As predicted by the topichood constraint, example (293) passes the 'Speaking of X' test, whereas (291) does not:

(295) a. * Speaking of the lute, Henry plays it and sings madrigals.
b. Speaking of the organ, Mary bought it and thereby angered her husband.

Thus, there is strong evidence that the (non-syntactic) notion of topichood is a key component of an adequate analysis of CSC data.

**Related Accounts** As a brief digression, I note that other functionally-oriented accounts of extraction have been proposed. Erteschik-Shir and Lappin (1979), for instance, define a discourse property they call *dominance*, defined as follows.

> Dominance: A constituent c of a sentence S is dominant in S if and only if the speaker intends to direct the attention of his hearers to the intension of c, by uttering S.

Of course, this definition leaves open the question of what it means for a speaker to intend to direct a hearer's attention to the meaning of a constituent. They propose a 'lie test' to get at this notion, which they attribute to Ross. The idea is that a clause in a complex sentence is dominant only if the sentence can be felicitously denied by denying the content of that clause. For instance, the fact that either clause can be denied in example (296) indicates that either can be interpreted as dominant.

(296) Bill said: John believes that Orcutt is a spy.
a. which is a lie – he doesn't.
b. which is a lie – he isn't.

On the other hand, in example (297), only the matrix can be felicitously denied, and thus the complement cannot be dominant.

(297) Bill said: John carefully considered the possibility that Orcutt is a spy.
a. which is a lie – he didn't (consider it carefully).
b. * which is a lie – he isn't (a spy).

Given the notion of dominance, they propose a constraint tying it to extraction.

> Dominance Hypothesis on Extraction (DH): An NP can only be extracted out of clauses which may be interpreted as dominant or out of phrases in which the NP may itself be dominant.

Erteschik-Shir and Lappin discuss the CSC, albeit briefly, claiming that neither clause is dominant in the cases we have been calling symmetric, as in example (298).

(298) Bill said: The nurse polished her trombone and the plumber computed my tax.
Mary said: It's a lie – *she/*he/they didn't.

Thus, neither conjunct alone is dominant and thus extraction is disallowed per the DH. They do not discuss any of the examples that felicitously violate the CSC.

Takami (1988) describes a functional account of preposition stranding which utilizes 'the concept of more/less important information', based on the following hypothesis.

> An NP can only be extracted out of a PP which may be interpreted as being more important (newer) than the rest of the sentence.

Of course, this hypothesis requires a definition of what it means to be 'more important', for which Takami offers the following.

> An element in a sentence represents new (more important) information if the speaker assumes that the hearer cannot predict or could not have predicted that the element will or would occur in a particular position within the sentence. (p. 313)

Deane (1991) subsequently assimilates Kuno's notion of topichood, Erteschik-Shir and Lappin's notion of dominance, and Takami's notion of informational focus to the notion of salience. He says:

> According to the present theory, extraction occurs when the following pattern holds: (i) The extracted NP is potentially topical, hence commands attention. (ii) The matrix phrase

> for extraction is the information focus, hence commands attention. (iii) The rest of the sentence must not encourage a different construal in which the matrix NP is not intrinsically focal, hence, extraction is best when the intervening parts of the sentence present presupposed or given information. The more clearly this pattern holds, the more acceptable extraction should become. (p. 47)

I will not attempt to move any further toward pinpointing the notion of topic here. (See Deane (1991) for a more in depth comparison between the approaches described above.) For my purposes, I can proceed by appealing to Kuno's conception of topic and his corresponding 'Speaking of X' test; beyond this I will keep claims regarding the role of topics in discourse to a minimum. For instance, I will neither claim that every sentence has an associated topic, that a given sentence or discourse only has one possible topic, nor that the topic of a sentence necessarily corresponds to the semantics of a syntactic constituent within that sentence. The only commitment I make is to the notion that there is some type of topichood constraint that applies to entities which undergo extraction.

## 5.5 Toward Explaining Away the CSC: The Interaction between Topichood and Coherence

As is probably clear from the foregoing analyses, the notion of topichood in the literature still lacks a formal, concrete foundation. As Deane (1991, pp. 30-31) states, "in order to constitute genuinely predictive theories both approaches [= Kuno's and Erteschik-Shir's] require an explanation for why some NPs can qualify for dominance or topichood, but others cannot." Deane later admits that his approach also lacks this characteristic, indicating that his theory "would have to be grounded in psycholinguistic research on attention".

Nonetheless, the facts concerning extraction, and their correspondence with tests that do not involve extraction, suggest that there is a coherent linguistic phenomenon responsible for these facts. This, combined with the fact that the behavior of the extraction data correlates with my neoHumian tripartite categorization of coherence relations, in turn suggests that the explanation of the CSC data lies at the intersection between topichood and coherence establishment.

Of course, a truly independently-motivated theory of the CSC data would therefore require a theory how information packaging notions like topichood interact with processes for establishing coherence, an area of study which is currently in its infancy. Thus, I will take only some initial steps toward such an explanation here. Although my analysis will

primarily be an elaboration on the (relatively similar) insights published by the researchers discussed above, I differ in situating these insights with respect to my theory of coherence relations and the properties of the different inference procedures that underlie the establishment of these relations. As a result, while I will not provide a comprehensive solution to the problem of determining possible topics in discourse, I hope to convince the reader that an adequate account of the CSC data will likely prove to be a by-product of the solution to this larger question, in an analysis in which it is suitably integrated with independently motivated aspects of coherence establishment.

### 5.5.1 Extraction from Resemblance Relations

I first consider the case of conjoined clauses related by Resemblance. The cases cited in the literature which require extraction to be across-the-board fall into this class, as exemplified again in (299a-d).[19]

(299) a. John bought the book and Bill read the book.
b. What book did John buy and Bill read?
c. John bought the book and Bill read the magazine.
d. # What book did John buy and Bill read the magazine?

Past researchers (e.g. Schmerling, Kuno) have attributed this constraint to the fact that the clause *read the magazine* is not "about" the book in (299d), and thus the book cannot serve as a topic for a conjoined set of clauses that contains it. While this is no doubt correct in principle, I elaborate on this intuition in terms of the coherence processes that underlie the establishment of Resemblance.

Recall that the coherence of Resemblance relations is rooted in the identification of points of similarity (and contrast) between parallel entities and relations. I repeat the definition of the Parallel relation here:

**Parallel:** Infer $p(a_1, a_2, ...)$ from the assertion of $S_1$ and $p(b_1, b_2, ...)$ from the assertion of $S_2$, where for some property vector $q$, $q_i(a_i)$ and $q_i(b_i)$ for all $i$.

The establishment of these similarities yields superordinate categories defined by the $q_i$ that contain each pair of parallel elements $a_i$ and $b_i$, as well as a superordinate relation $p$ that includes the relations expressed in each clause.

These superordinate categories are ultimately what get passed to higher-level discourse structures for the purpose of determining coher-

[19]My analysis treats sentences like (299d) as infelicitous rather than ungrammatical, and thus I will begin marking unacceptability with # rather than *. However, I will continue to use * when citing examples from previous authors who used this marking.

ence with respect to those structures, and thus they are the categories that serve as candidates for the topic of the conjoined clauses. In fact, Lakoff (1971) calls such categories 'common topics' in her discussion of the constraints governing the felicity of symmetric conjunction. She says:

> ...that at least one set of paired constituents must be reducible to partial or complete identity, in one of these ways, for a conjunction to be appropriate. That is essentially what is meant by *common topic*, and further implied by this name is the notion that, if only one pair is identical, this cannot be just a random pair, but, in some sense, the identity must involve that pair of constituents in the two conjuncts that are *what the sentence is particularly about.* (p. 122)

The resulting hypothesis is that only a potential common topic identified by the inference processes underlying the establishment of the Parallel relation can be placed in a topic-denoting position that scopes over a set of conjoined clauses related by Parallel. This characterization will prove slightly too strong, but it will suffice for the moment.

I illustrate the inference process by considering the establishment of Parallel for examples (299a-b) and (299c-d). In the case of (299a), the arguments $a_1$ and $b_1$ are *John* and *Bill* respectively, and likewise $a_2$ and $b_2$ are both *the book*. The common $q_1$ will define a superset of *John* and *Bill*, roughly akin, say, to *familiar men*. In the case of $q_2$, the establishment of similarity is immediate from the coreference between $b_1$ and $b_2$, and therefore $q_2$ defines the *the book* itself as the 'superordinate' category. Thus, possible candidates for what the conjoined sentence is about include *familiar men* and *the book*.

The fact that *the book* is a potential topic means that it can be extracted to a position that establishes it as the topic, as demonstrated by (299b). Note that extraction must take place across-the-board because the parallelism between the clauses with respect to the topic must be maintained. That is, the constraint is more than that the extracted element be "about" each clause; one cannot only extract from one of the conjuncts, even though both clauses are still about *the book*:

(300) a. # This is the book which John bought and Bill read it.

b. # This is the book which John bought it and Bill read.

I now consider examples (299c-d). In (299c), $a_1$ and $b_1$ are again *John* and *Bill*, and $a_2$ and $b_2$ are *the book* and *the magazine* respectively. The common $q_i$ will define the supersets of, say, *familiar men* and *reading materials* respectively, and thus these serve as the possible topics for

the conjoined clauses. The fact that *reading materials* can serve as a topic can be demonstrated using the 'Speaking of X' construction, as in example (301).

(301) Speaking of reading materials, John bought the book and Bill read the magazine.

Because my analysis assigns what amounts to a discourse-level coherence structure to these sentences, I will also find it useful to use a test for discourse topichood, one that I will call the 'Let me tell you about X' test. Carrying out this test requires that one break up the clauses in the relevant examples into separate sentences, pronominalizing at any extraction sites, and seeing if the result is coherent within a discourse that begins "Let me tell you about X" in which the topic phrase has been substituted for X. The topic *reading materials* indeed satisfies this test also.

(302) a. Let me tell you about the reading materials.
b. John bought the book.
c. Bill read the magazine.

Importantly, I should point out that *reading materials* is just one of potentially many possible topics for this sentence; others will be allowable if sufficiently supported by context. This is the case, for instance, in (303).

(303) Speaking of junk from Joe's garage sale, I hear John bought the book and Bill read the magazine.

As long as context supports the inference that the book and magazine are from Joe's garage sale (or better yet, does not contradict it), *junk from Joe's garage sale* will be a suitable $q_2$, and hence a possible topic. In general, context no doubt often plays a prominent role in feeding the process of establishing the appropriate common topics for Parallel constructions embedded within larger discourse structures.

Nonetheless, none of the entities actually mentioned in (299c) qualify as a topic for both clauses. Therefore, neither can be extracted on its own, as shown again in (304a-b).

(304) a. # This is the book that John bought and Bill read the magazine.
b. # This is the magazine that John bought the book and Bill read.

As we would expect, neither *the book* nor *the magazine* satisfy the 'Speaking of X' test, as shown in (305a-b), nor do they satisfy the 'Let me tell you about X' test, as shown in (306) and (307).

(305) a. # Speaking of the book, John bought it and Bill read the magazine.
b. # Speaking of the magazine, John bought the book and Bill read it.

(306) a. Let me tell you about the book.
b. John bought it.
c. # Bill read the magazine.

(307) a. Let me tell you about the magazine.
b. # John bought the book.
c. Bill read the magazine.

Thus, the across-the-board extraction constraint is derivable from the fact that only entities mentioned in each clause can serve as a common topic for the conjoined set of clauses, as predicted by the properties of the inference processes that identify possible common topics during the establishment of Resemblance.

I mentioned earlier that my characterization of this constraint was slightly too strong, however. This characterization was based on the assumption that in order to get extraction across multiple clauses in a manner which both preserved topic marking and maintained parallelism, the extracted element would necessarily have to be the same in each clause. On the other hand, if it were possible to extract more than one distinct entity into a single topic-denoting position from a set of clauses related by Parallel, and then relate each of these entities to the clauses that they serve as topics of, then the result might still be acceptable since parallelism would be preserved. There is in fact a way to do this in English, in particular by using a *respectively* construction. Consider example (308), adapted from one in Postal (1998).

(308) What book and what magazine did John buy and Bill read respectively?

Obviously the issues related to the interaction of extraction, topichood, and the semantics of respective readings get complex when considering these cases. A suitable analysis would nonetheless be one in which this interaction predicts that each extracted element need only be the topic of the clause that *respectively* relates it to. Of course, the need for parallelism would require that if one topic is extracted then all need to be, and the unacceptability of example (309) would suggest that this is in fact the case.

(309) # What book and what magazine did John buy, Sue write the novel, and Bill read respectively?

Another way to place multiple entities into an equally prominent position is to use a single fronted noun phrase that obtains its denotation referentially, and then use *respectively* to relate these to their respective clauses (Dalrymple and Kehler 1995, Gawron and Kehler 2000). Sentence (310) is an example:

(310) The book is on the table and the magazine is on the chair. Those reading materials are what John bought and Bill read, respectively.

In sum, the fact that extraction needs to occur in an across-the-board fashion can be seen as a side-effect of the constraints imposed by the inference processes underlying the establishment of Resemblance. As I describe in the next two sections, the different types of inference processes underlying the establishment of the other two classes of relation lead to correspondingly different constraints on extractability.

### 5.5.2 Extraction from Cause-Effect Relations

I now consider the case of conjoined clauses in which a Cause-Effect relation is operative. Recall that the examples that Lakoff (1986) calls Type B scenarios (i.e., Violated Expectation relations) and Type C scenarios (i.e., Result relations) are not subject to across-the-board restrictions on extractability. These two cases were exemplified in (264) and (265a), repeated below as (311) and (312) respectively.

(311) How much can you drink and still stay sober?

(312) That's the stuff that the guys in the Caucasus drink and live to be a hundred.

As one would expect, both of these cases pass the 'Speaking of X' test, as shown in (313) and (314).

(313) Speaking of minikegs, John can drink one and still stay sober.

(314) Speaking of that stuff, the guys in the Caucasus drink it and live to be a hundred.

Cause-Effect relations are no different than Resemblance relations in that there are in general many different possible topics that a given passage could have. In particular, a suitable topic need not correspond to an element in the sentence at all, as illustrated by sentence (315).

(315) Speaking of mysteries of the universe, the guys in the Caucasus drink that stuff and live to be a hundred.

On the other hand, Cause-Effect relations are quite different in terms of the properties of the inference processes underlying their establishment.

Whereas the need to establish a 'common topic' in Resemblance relations has the side-effect of requiring extraction to occur across-the-board, there is no analogous constraint that arises from the need to draw the type of implicational relationship necessary to establish Cause-Effect. As a result, we do not see an across-the-board constraint.

As several researchers have noted, when extraction is not across-the-board, the extraction typically occurs from the first conjunct and not the second. This was seen in the difference between minimal pairs like (281) and (282), repeated below as (316) and (317).

(316) a. How many lakes can we destroy and not arouse public antipathy?

b. * What can we destroy many lakes and not arouse?

(317) a. Which dish is it that people always order in this joint and then get sick?

b. * How sick do people order that chili dish here and then get?

This pattern results from the fact that the first clause in Result and Violated Expectation relations is what Na and Huck call 'primary' with respect to the function the clause pair serves in the larger discourse. Put another way, the first clause typically makes the assertion that plays the dominant role in determining how the clause pair coheres with the segment in which it is embedded. Thus, we would predict that an extractable topic would have to come from the first clause.

A particularly striking example that establishes the effect of the coherence relation in determining extractability was provided to me by Gregory Ward (p.c.). Let us start by considering (318a).

(318) a. John read the book and saw the movie.
(Result or Parallel)

b. This is the book that John read and saw the movie.
(Result only)

c. Speaking of this book, John read it and saw the movie.
(Result only)

There are two ways to interpret example (318a) that are relevant here. First, there is a Parallel interpretation in which John both read a (presumably aforementioned) book and similarly saw an (also presumably aforementioned) movie that was perhaps unrelated to the book. Second, there is a Result interpretation in which the movie had not been mentioned already; instead the reading is such that John read a book and, as a result, saw *the movie based on the book*. The extracted version of example (318a) shown in (318b), however, can only have the Result

reading, which is exactly what my analysis predicts. The effect is reminiscent of the gapping examples of Levin and Prince (1986) discussed in Chapter 4, except in those cases the application of gapping ruled out the Result interpretation, leaving only the Parallel reading.

Na and Huck (1992) also discuss an example of this sort, shown in (319).

(319) Terry ran in these shoes and hurt her knee.

There are at least three relations that can hold between the clauses: (i) Parallel, in which the two events are understood as unrelated, (ii) Occasion (Type A), in which the events are interpreted as a natural sequence but the shoes are otherwise incidental, and (iii) Cause-Effect (Type C), in which the shoes were responsible for Terry hurting her knee. As predicted, example (320a) can only have the Occasion interpretation, and (320b) can only have the Cause-Effect interpretation.

(320) a. Which knee did Terry run in these shoes and hurt?

b. Which shoes did Terry run in and hurt her knee?

As would be expected in an analysis based on coherence, context can also influence the extent to which extractions that are not across-the-board are judged to be acceptable. Johannessen (1998), for instance, considers example (321).

(321) * What kind of herbs can you eat and Mary see a mouse?

This example seems quite unacceptable, but nonetheless improves considerably if one considers a situation in which

> ...Mary, who has a slight mental defect, always sees an animal when someone eats herbs. A friend of hers tells you that Mary actually sees different animals; each herb triggers its own animal. (p. 233)

Such a context supplies the information necessary to interpret (321) as a Result relation, and from this the extraction becomes acceptable.[20] (Recall similar examples from Chapter 2 in which context changed an incoherent passage into a marginally coherent one.) Thus, we see the type of sensitivity to context and interpretation that one would expect under a coherence analysis, but which undermines a purely syntactic explanation of these facts.

[20]Note that this example contradicts a claim by Schmerling (1975), cited by Na and Huck (1992), that sentence-level conjunctions cannot felicitously violate the CSC.

### 5.5.3 Extraction from Contiguity Relations

Finally, I consider cases in which the Contiguity relation Occasion is operative, such as examples (236b), (266a), and (273a), repeated below and shown with their unextracted counterparts in (322–324).

(322) a. I went to the store and bought some whiskey.
b. Here's the whiskey which I went to the store and bought.

(323) a. Harry went to the store, bought a cake, loaded it in his car, drove home, and unloaded it.
b. What did Harry go to the store, buy, load in his car, drive home, and unload?

(324) a. You can sip this brandy after dinner, watch TV for a while, sip some more of it, work a bit, finish it off, go to bed, and still feel fine in the morning.
b. This is the sort of brandy that you can sip after dinner, watch TV for a while, sip some more of, work a bit, finish off, go to bed, and still feel fine in the morning.

Like the case of Cause-Effect relations, and unlike the case of Resemblance relations, we have no reason to expect that inherent properties of the inference processes underlying the establishment of Occasion would require that the topic be mentioned in every clause. On the other hand, the notion of topic plays a greater role in the coherence of an Occasion than it does for a Cause-Effect relation, since a topic would necessarily serve as a focal point around which the sequence of eventualities being described is centered. The prediction of our analysis would therefore be that the ability to extract an entity from a set of conjoined clauses related by Occasion correlates with the extent to which that entity serves as the focal point of that Occasion.

As I discussed earlier, Deane (1991) described a variety of types of conjuncts that "possess special functions within a larger narrative frame", which he categorized as *preparatory actions, scene-setters, internal causes*, and *incidental events*. These are the types of clauses that can be included in a narrative without creating a change of topic; they provide background states, situate events with respect to time and location, and indicate changes of state necessary to connect other events in which the discourse topic participates. All of the aforementioned passages pass the "Let me tell you about X" discourse topic test; for instance, consider a discourse version of example (323a), given in (325).

(325) a. Let me tell you about the cake.
b. Harry went to the store.

c. He bought it.

d. He loaded it in his car.

e. He drove home.

f. He unloaded it.

The inclusion of utterances (325b) and (325e) is perfectly felicitous despite the fact that they do not mention the cake, since they both describe changes of state that are necessary to connect events that do involve the cake.

As such, the inclusion of these clauses also does not affect the ability to pronominalize the mention of the cake in sentences (325c) and (325f), even though in these cases it was last mentioned two sentences prior. (See also Section 6.5.3.) On the other hand, it is essential that such clauses perform an ultimately topic-relevant function. For instance, consider a version of passage (325) in which the final utterance is replaced with one that is irrelevant to the cake, as shown in (326).

(326) a. Let me tell you about the cake.

b. Harry went to the store.

c. He bought it.

d. He loaded it in his car.

e. He drove home.

f. He took a bath.

Ignoring sentence (326a) for the moment, the sentences in (326b-f) comprise a perfectly coherent Occasion (assuming, of course, that the initial use of *it* in (326c) is replaced with *the cake*). However, *the cake* is not a possible topic for this discourse. In particular, the change of state communicated by sentence (326e) does not serve as a bridge to another eventuality that is relevant to the cake. As a result, the Occasion is only coherent assuming a topic such as "what Harry did today", for example. Hence the oddness of the discourse when sentence (326a) is included.

As we would expect, therefore, extraction from the sentential equivalent of (326a) is unacceptable, as shown in (326b).

(327) a. Harry went to the store, bought a cake, loaded it in his car, drove home, and took a bath.

b. # What did Harry go to the store, buy, load in his car, drive home, and take a bath?

Hence, we can see why there appears to be a constraint operative in Occasions that requires extraction from the final clause, as Lakoff noted. When the final clause does not involve the entity which has served as

the topic up to that point, the discourse ceases to have that entity as its topic.

Of course, the unacceptability of (327) can be remedied by simply adding another clause which centers around the cake, as shown in (328).

(328) What did Harry go to the store, buy, load in his car, drive home, take a bath, and then devour in thirty seconds?

The addition of this final clause allows the two preceding it to be construed as changes of state which connect events centered around the cake, allowing the cake to serve as a topic for the entire Occasion again. (In general, we would expect the ability to insert additional scene-setting clauses to be rather limited, however, and even this example might be somewhat degraded by having two such clauses in a row.) The fact that adding this final clause can cause the difference in acceptability between (327) and (328) is yet another indication that this apparently syntactic phenomenon does not admit of a purely syntactic explanation.

Therefore, as we might have predicted from what we know about coherent Occasions, it is not the case that extraction must occur across-the-board when an Occasion relation is operative. In particular, clauses which do not mention the extracted topic, but nonetheless contribute coherently to an Occasion centered around that topic, are immune to this constraint. Thus, the judgments on such examples bear more directly on whether the constraints on possible topics in coherent Occasions are met than than on constraints governing extraction from coordinate clauses.

I close this section with a discussion of a remaining fact relevant to those clauses in an Occasion which do mention the topic. In particular, extraction is obligatory for those clauses which mention the topic for a purpose that is central to the Occasion. It is perhaps unsurprising that this would be true of cases in which a mention of the topic could have been extracted but was not, such as when a pronoun is left behind in a clause in which a gap could have existed, as in (329a-b).

(329) a. ?? This is the cake that Harry went to the store, bought, loaded it in his car, drove home, and unloaded.

b. ?? This is the cake that Harry went to the store, bought, loaded in his car, drove home, and unloaded it.

We saw this same effect with examples (300a-b) when we considered Resemblance relations; all of these examples have a resumptive pronoun feel to them. However, this constraint also applies in situations in which a gap could not exist, as in (330a-b).

(330) a. ?? This is the cake that Harry went to the store, bought, loaded in his car, drove home, unloaded, and ate its icing.

b. ?? This is the cake that Harry went to the store, bought, loaded in his car, drove home, unloaded, and ate the icing off.

The use of a possessive pronoun in the last conjunct of (330a), and the use of a noun phrase that is referentially dependent on the cake (*the icing*) in (330b), are enough to maintain the current discourse topic *the cake*. For instance, these examples readily pass the 'Speaking of X' test as shown in (331a-b); the 'Let me tell you about X' discourse topic test also succeeds.

(331) a. Speaking of the cake, Harry went to the store, bought it, loaded it in his car, drove home, unloaded it, and ate its icing.

b. Speaking of the cake, Harry went to the store, bought it, loaded it in his car, drove home, unloaded it, and ate the icing off.

Nonetheless, the fact that the topic cannot be extracted from the final clauses of (330a-b) renders these cases infelicitous.

Again, this fact may be attributable to the central role that these events play within the Occasion. Compare (330a-b) with (332).

(332) This is the cake that Harry went to the store, bought, loaded in his car, drove home, unloaded, tore open its box, and devoured in thirty seconds.

Sentence (332) seems to be at least marginally better than examples (330a-b), if not impeccable. In this case, the clause *tore open its box* acts more as what Deane called a *preparatory action* than the corresponding conjuncts in examples (330a-b).

### 5.5.4 Summary

To summarize so far, not only is there no need to stipulate a constraint barring extraction from coordinate structures in the theory of grammar, but such a stipulation makes a number of wrong predictions. The CSC data can instead be explained by the same constraint on topichood that is seen in cases of extraction from noncoordinated clauses. Crucially, what constitutes a topic for a set of conjoined clauses varies with respect to the type of coherence relation that is operative between them.

Proponents of syntactic approaches have often implicitly considered the Parallel examples to be the 'core' cases of extraction from coordinate clauses (or alternatively, those which involve 'true conjunction'), dismissing the remaining cases as different phenomena parading in the guise of conjunction. While it is possible that the Parallel cases are more common empirically, my categorization of coherence relations makes it clear that there is neither anything more 'core' about such cases nor do

they in any sense involve a 'truer' form of conjunction than the other examples. They are simply characterized by the Parallel relation, one of several possible relations between clauses.

## 5.6 Impact on Theories of Autonomous Syntax

Several researchers have noted (Lakoff 1986, Lakoff 1991, Deane 1991) that the foregoing data presents a potential challenge to theories which posit that rules of syntax are autonomous of other modules of language processing. While one's stance might depend on exactly how this claim is formulated, it would appear that, at a minimum, our concept of what makes for an ungrammatical sentence would need to be revised if the claim that syntax is autonomous is to be maintained.

Consider again some of the data I have discussed in this chapter:

(333) a. * What book did John buy and read the magazine?
b. What did Harry go to the store and buy?
c. How much can you drink and still stay sober?

Lakoff points out that if one allows an autonomous syntactic module to generate (333b-c), then it must also be able to generate (333a), leaving the task of filtering out (333a) to semantic or pragmatic constraints. The converse – keeping the CSC and adding a semantic or pragmatic condition to allow (333b-c) – is not an option, since such conditions could not turn an ungrammatical sentence into a grammatical one.

It may therefore be possible to retain a suitably articulated hypothesis that syntax is autonomous and still account for these data, as long as the CSC is neither included within nor is a by-product of the system of grammar rules. Because this view forces us to the conclusion that sentences like (333a) are perfectly grammatical, however, it brings to light a potentially worrisome situation regarding the manner in which theorists rely on their judgments. Previous researchers have certainly considered sentences like (333a) to be ungrammatical, and it is at least questionable whether this judgment differs qualitatively from other ungrammaticality judgments upon which researchers commonly construct their syntactic theories. This data should force us to reassess whether the intuitions we have about ungrammaticality really represent *syntactic* wellformedness, and if they do not, what one might use as a basis for determining what sentences are unacceptable for purely syntactic reasons.

These data are also striking in that they demonstrate the need to intertwine theories of discourse structure and coherence with those for sentence structure and coherence. A processing model in which entire sentences are analyzed syntactically before any discourse-level reasoning takes place is not likely to prove well-suited to explaining this data. The

situation gets even more complicated by the fact that, as Lakoff notes, examples in which multiple clauses are conjoined can involve more than one scenario (i.e., coherence) relationship. For instance, example (334), repeated from (266b), involves both Type A and B scenarios.

(334) How many courses can you take for credit, still remain sane, and get all A's in?

When viewed as a discourse, the structure relating these clauses is hierarchical. The first two conjuncts participate in a Violated-Expectation relationship, and then they as a unit (with the first clause being dominant) participate in an Occasion relationship with the third conjunct. Labelling the three conjuncts as A, B, and C respectively, the structure is as depicted in Figure 13. (The third argument in each relation represents the central information that is percolated to higher-level nodes in the discourse structure for the purpose of establishing coherence at that level.) Analyzed in this light, the fronted element is both a sentence and discourse topic. In fact, this might suggest that the fact that extraction works within a sentence is a tangential issue: The constraint on extractability of elements corresponds to an ability to serve as a topic *at that level of discourse structure.* In the case of sentence (334), it therefore corresponds to the ability of *how many courses* to serve as a topic at the top node of the tree in Figure 13.

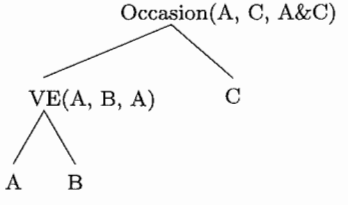


FIGURE 13 'Discourse' Structure for Example (334)

I will resist the temptation to carry this reasoning further, since I would undoubtedly evoke far more questions than I could possibly hope to answer. Needless to say, teasing apart these and related issues lying at the syntax-discourse interface will require significant further research.

## 5.7 An Attempt at Salvaging a Syntactic Solution: Postal's Analysis

Despite the wealth of evidence supporting the claim that an adequate explanation of the CSC facts requires an appeal to semantic or pragmatic

notions, Postal (1998) has recently presented a purely syntactic analysis of these data. In fact, he calls the CSC "an important universal truth about natural languages", adding:

> ...my impression is that the CSC is widely regarded as the most problem-free syntactic constraint ever discovered. As Gazdar (1982, 175) states, "Numerous island constraints other than the CSC have been proposed in recent years. Unfortunately, few if any of them are as resilient to counterexamples as the CSC is." (p. 52)

His attempt to show that the counterexamples to the CSC can be explained solely through syntactic means requires him to enrich the theory of syntax with new operators. As Maslow might have warned us would happen, those enrichments cause his analysis to falter.

To elaborate, Postal posits the existence of "invisible (and thus not phonetically realized) resumptive pronouns" (henceforth, IRPs), which, importantly, are distinct from traces. This leads to a distinction between two types of extraction: A-extraction, which leaves behind a trace, and B-extraction, which leaves behind an IRP. Postal claims that B-extractions are not permitted in those scenarios in which a weak definite pronoun is also not allowed; weak definite pronouns are considered to be the phonetically realized correlates of IRPs. He then attempts to demonstrate that Lakoff's counterexamples to the CSC are either instances of B-extraction, and thus are not true counterexamples, or that they are examples that do not involve 'true conjunction'.

In his review article response to Postal's book, Levine (2001) provides an extensive array of evidence showing that the distinction between traces and IRPs, and thus between A-extraction and B-extraction, cannot be maintained. Rather than repeat his arguments here, I simply refer the reader to his well-argued paper for the details. Notably, however, Levine's careful examination of the data brings him to conclude:

> In short, it appears that Lakoff was right after all, and the CSC cannot be maintained as a syntactic restriction.

Certain arguments that Postal uses to defend his analysis warrant further discussion here, however. First, with respect to Type B scenarios, Postal makes what is by now a familiar move by claiming that such structures are not, despite appearances, true coordinate structures. He instead considers B-conjuncts to be adjuncts. As I have argued, this claim would need a substantial amount of evidence to be taken seriously; this evidence would have to be purely syntactic in nature and not depend on the presumption that the CSC is a valid syntactic constraint.

However, the evidence offered is fairly scant; the main example being a particular English construction, *double neg*, discussed by Lawler (1974) and exemplified in (335).

(335) Can linguists study negation? Not and stay sane, they can't.

Postal, following Lawler, claims that example (335) is a potential counterexample to the *Conjunct Constraint* part of the CSC. Postal concludes from this that either the Conjunct Constraint is false, or Type B scenarios do not involve true conjunction. Since dispensing with the Conjunct Constraint would "solve a quite restricted problem ... at the cost of creating many much graver ones", then the only conclusion we can draw is that Type B scenarios cannot involve true conjunction.

However, Postal does not address certain facts that must be explained for his case to go through. In particular, Lawler himself offers reasons to think that the double neg construction is more idiosyncratic, being severely limited in its distribution. For instance, a version of example (335) with positive modality is unacceptable, as shown in (336).

(336) Can linguists study negation? * And (still) stay sane, they can.

Second, Postal's explanation offers no reason why the extraction cannot be made out of the second conjunct, as in (337).

(337) Can linguists study negation? * Not stay sane and (still), they can't.

The relation between the clauses in these examples is rooted in the same type of Cause-Effect inference seen in (335) and Goldsmith's violations of the CSC. If none of these cases are instances of 'true conjunction', then there is no explanation for why they are unacceptable. The more natural conclusion is that there are other factors at play in this construction, and thus it would be quite premature to use it as a basis for arguing that *and* is not operating as a conjunction. Furthermore, recall that examples (262a-d) showed that *and* fails the frontability test for subordination. To reiterate the point with respect to these examples, unlike adjuncts and subordinators, the *and still* conjunct cannot be fronted, suggesting that it is every bit the conjunction that it appears to be.

(338) a. Linguists can study negation, staying sane as they proceed. (adjunct)

b. Staying sane as they proceed, linguists can study negation. (fronted adjunct)

c. Linguists can study negation, although they won't stay sane. (subordinator)

d. Although they won't stay sane, linguists can study negation. (fronted subordinator)
e. Linguists can study negation and still stay sane.
f. * And still stay sane, linguists can study negation.

See Levine (2001) for additional arguments against the tenability of the claim that these are not cases involving true conjunction.

Another move that Postal makes in his analysis is an attempt to account for the facts in terms of the properties of particular conjuncts themselves, rather than in terms of the operative relationship between them. He says:

> Lakoff's...categories might seem to treat the relevant VP structures as indivisible wholes. But it would probably be more accurate to pick out particular conjuncts as having crucial properties.

This comment is a little puzzling, since one of the central points of Lakoff's study was to show that it is exactly the relationship between the conjuncts that matters. This is particularly clear if we consider example (318) again, repeated below as (339).

(339) a. John read the book and saw the movie. (Result or Parallel)
b. This is the book that John read and saw the movie. (Result only)

Recall that (339a) has both Parallel and Result readings; Postal would presumably consider only the Parallel reading to involve 'true conjunction'. For his arguments to be salvageable, he would have to demonstrate that these two relationships between the conjuncts of (339a) can be attributed to the conjuncts themselves having distinct syntactic properties in each case. He neither offers any arguments that would support such a claim, nor is it easy to imagine how he could.

This data is only mysterious if one assumes *a priori* that there must be a purely syntactic explanation for all of the CSC facts.[21] When considering the coherence relationship between the clauses, we see that the different interpretations that can be inferred for (339a) can also be inferred when the clauses lie across sentence boundaries, as in (340).

(340) John read the book. He saw the movie.

There is no more reason to believe that these two readings correspond to a difference in the syntax of the individual clauses in (339a) than there

[21] Indeed, one cannot help but to believe that proponents of purely syntactic explanations may be reasoning from such an *a priori* commitment, rather than arriving at such an explanation through a process of open-minded scientific inquiry. See Lakoff (1991) for discussion.

is to believe that the same is true for (340). Insofar as any of Postal's purportedly independent syntactic tests identify a difference between the two meanings of (339a), this difference would be more likely to serve as evidence that the basis for his tests can also be explained in terms of discourse-level phenomena than would it support his thesis that the CSC is a valid constraint of grammar.

## 5.8 Remaining Questions about Coherence

Despite my criticisms of Postal's syntactic account, the wealth of data he cites is a testament to the fact that a pragmatic theory has far to go to explain all of the judgments for the CSC data. In many cases, for instance, Levine's (2001) counterexamples to his analysis were generated by making minor adjustments to Postal's examples which nonetheless reversed the resulting acceptability judgments. The question, then, is what factors account for these differences.

In this section, I discuss a few facts of this sort that have been discussed in the CSC literature. I argue that in many cases the differences in acceptability between closely related sentence pairs do not result from a shift in whether a conjoined sentence is a case of 'true conjunction', but that these differences can be attributed instead to a change in the type of coherence relation that hearers infer between the clauses. While the reasons why subtle changes can substantially alter the inferred relationship may remain mysterious in some cases, the correspondence between acceptability judgments and coherence relations is further evidence that the coherence relation is playing a key role. As such, the onus of explaining this data ultimately falls on theories of coherence and its interaction with other factors; an inability for current theories to explain all of these facts in no way argues for salvaging the CSC as a grammatical constraint.

For starters, recall that Ross noted the distinction between (341a) and (341b), which gives rise to the difference in acceptability between (341c) and (341d).

(341) a. I went to the store and bought some whiskey.
b. I went to the store and Mike bought some whiskey.
c. Here's the whiskey which I went to the store and bought.
d. * Here's the whiskey which I went to the store and Mike bought.

Ross reasons that this difference is attributable to a distinction between extracting from conjoined VPs and Ss. However, notice also that the inferred coherence relation has switched from Occasion to Parallel – sentence (341b) is not read as a natural sequence of events as sentence

(341a) is. This difference results from the fact that the actions of going to a store and of buying something form a natural sequence when the same agent performs them, but not when different ones do. Thus, the change in acceptability between (341c) and (341d) corresponds in the expected way with the change in coherence relation inferred.

This is not the whole story, however. As Schmerling noted, even if the agents are the same, the Parallel interpretation is favored as long as the agent is rementioned as the subject of the second clause. Recall examples (253a-b), repeated below as (342a-b).

(342) a. I went to the store and I bought some whiskey.

b. * This is the whiskey which I went to the store and I bought.

Here, the (otherwise unnecessary) remention of the subject in the second clause appears to serve to "break up" the two events, perhaps with the explicit purpose of blocking the inference that they are a tightly connected sequence of events. No matter what the ultimate explanation for this effect is, the change of coherence relation inferred is consistent with the claim that coherence is ultimately responsible for this difference in extractability, and not a purely syntactic distinction between conjoining VPs or Ss.

A similar explanation applies to examples (239a-b), also discussed by Ross, repeated as (343a-b).

(343) a. I went to the movies and didn't pick up the shirts.

b. * The shirts which I went to the movies and didn't pick up will cost us a lot of money.

Again, the negation does not allow this passage to be construed as a natural sequence of events (i.e., Occasion), and thus extraction from only one conjunct is not allowed. Similarly, Ross showed how a change in tense can affect the ability to extract in examples (240a-b), repeated below as (344a-b).

(344) a. I went to the store and have bought some excellent whiskey.

b. * Here's the excellent whiskey which I went to the store and have bought.

Again, this choice of tense causes one to understand the events as unrelated rather than as a natural sequence. The Occasion interpretation is therefore eliminated, and extraction is unacceptable. The fact that all of these variations on example (341c) change the inferred coherence relation from Occasion to Parallel is evidence that coherence is ultimately responsible for these differences.

Having said this, it is worth reiterating a point made by Lakoff about

our level of understanding concerning all of the factors that are relevant to these examples. He concedes:

> I do not want to give the impression that scenarios are completely well-understood... In particular, the relationship between scenarios and their syntactic realizations has by no means been worked out. (p. 155)

As an example, he makes the interesting observation that his B-scenarios appear to "be better at structuring hypothetical rather than realized situations and hence to prefer modals". To illustrate, (345a) seems to be more acceptable than (345b).

(345) a. How much can he drink and still stay sober?
b. ? How much did he drink and still stay sober?

Recall that Goldsmith made a similar point, citing a preference for the first clause to express an 'activity-type' rather than a specific action, as shown by (261a-b), repeated below as (346a-b).

(346) a. How many courses can we expect our graduate students to teach and still finish their dissertations on time?
b. # How many courses did Mrs. Sykes teach last year and still finish her dissertation on time?

It is not clear why we should see these differences. After all, the unextracted versions of (345a-b) and (346a-b), shown in (347a-b) and (348a-b), are both perfectly felicitous assuming a Violated Expectation interpretation.

(347) a. John can drink two six-packs and still stay sober.
b. John drank two six-packs and still stayed sober.

(348) a. Our graduate students can teach two courses and still finish their dissertations on time.
b. Mrs. Sykes taught two courses last year and still finished her dissertation on time.

This difference shows up again in another minimal pair Lakoff discusses, given in (349a-b).

(349) a. # How big a meal did he eat and feel satisfied?
b. How small a meal can you eat and feel satisfied?

Lakoff categorizes sentence (349a) as a Type A scenario (i.e., Occasion) and sentence (349b) as a Type B scenario (i.e., Violated Expectation). This categorization predicts the difference in judgments, since Type A scenarios normally require extraction from the final conjunct. Upon

closer inspection, however, it is not entirely clear why (349a) should not be categorized as a Type C scenario (i.e., Result), since the effect of feeling satisfied can be seen as being caused by the eating.

As described earlier, Na and Huck addressed this issue by distinguishing between 'weak' cases of cause-effect and true Type C scenarios, the former of which only allow extraction that is across-the-board. In such cases, they claim that *and* is perhaps best paraphrased as 'and now' rather than 'and as a result'. Sentence (349a) is presumably of this sort. However, Johannessen (1998) offers a context which makes this sentence more acceptable. In particular, she claims that (349a)

> ...is felicitous to ask in a situation in which you have just heard about a giant who for the first time has eaten enough food to feel satisfied, and you wonder how big that meal was. (p. 233)

The fact that this contextual support, which appears to have the effect of raising the centrality of the size of the meal, improves the acceptability of (349a) might suggest that the topichood constraint is also playing a role. Thus, a mixture of potentially subtle constraints seems to be at play in this data, which likely involves modality, the role of topichood, and the constraints imposed by coherence relations.

Sorting out all of these details will require further study. Despite these remaining issues, however, I believe that the evidence indicating that coherence is in part responsible for these facts is strong. As such, our inability to offer a concrete explanation for subtle differences in the data may be best attributed to our lack of understanding about the fine points of coherence establishment and its interaction with other factors, and should not lead us to a conclusion that the central thesis maintained here is not correct.

## 5.9 Conclusion

As in previous studies, the data and analysis discussed herein suggest that there is no purely syntactic Coordinate Structure Constraint operative in natural language grammar, and furthermore that it is highly unlikely that any purely syntactic explanation will be able to account for the data involving extraction from coordinate structures. Instead, any adequate explanation needs to take into consideration the type of coherence relation that is operative between the clauses over which extraction applies. In none of the three situations delineated by Hume's categories is extraction from a coordinate structure barred entirely; instead, there appear only to be weaker constraints at play that differ with respect to, and that are to some extent predictable from, these categories. The

so-called 'core' data supporting the CSC and the across-the-board exception thus arise from independently motivated factors that apply in only a particular subset of the possible scenarios that involve extraction from coordinate structures. It is important to understand that the remaining scenarios are not in any way 'peripheral' nor do they involve conjunction that is not 'true'; the relations operative in these cases just happen not to be Parallel.

# 6

# Coherence and Pronoun Interpretation

The manner in which people interpret pronouns[22] has been extensively studied in both computational linguistics and psycholinguistics. Much of this work has centered on proposing and arguing about so-called 'preferences' that people employ to identify pronominal referents. To date, however, no clear consensus has emerged regarding which preferences actually exist nor how exactly they are utilized in a method for identifying the referents of pronouns.

One reason for this is that, as with each of the other phenomena I have discussed in this book, the data are in apparent conflict. A consideration of examples (350a-c) offers a brief (and perhaps oversimplistic) glimpse into the situation.

(350) a. John kicked Bill. Mary told him to go home. [= John ]
b. Bill was kicked by John. Mary told him to go home. [= Bill ]
c. John kicked Bill. Mary punched him. [= Bill ]

These examples were used in a poll of native informants reported on by Kameyama (1996). For sentence (350a), Kameyama found that the majority of her informants preferred *John* as the referent of *him*, whereas for (350b), a majority preferred *Bill*. Note that the propositional content of (350a) and (350b) are essentially the same; the main difference is the voice used in the first clause, which in turn corresponds to a difference in whether *John* or *Bill* appears in subject position. Such data have inspired researchers to posit *grammatical role hierarchy* preferences, in which entities evoked from subject position are considered to be more

[22]In this chapter I am concerned with third person personal pronouns such as *he*, *she*, *it*, *they*, and so forth; I do not include demonstrative pronouns such as *this* and *that*. Furthermore, I am only concerned with usages of pronouns in which they receive their standard, unaccented intonation.

salient than those evoked from object position, which in turn are considered to be more salient than those evoked from other grammatical positions.

In the case of sentence (350c), however, Kameyama found that most of her informants preferred *him* to refer to *Bill*, contradicting the preference for referents evoked from subject position that sentences (350a-b) supposedly established. Data like (350c) have caused researchers to posit *grammatical role parallelism* preferences, which in this case amounts to a preference for the pronoun in object position to be resolved to a referent evoked from the object position in the previous sentence. However, a preference for grammatical role parallelism does not explain why the preferred referent for the pronoun in (350a) is *John* instead of *Bill*. Indeed, the difference between the preferred referents in (350a) and (350c) defies the syntactic similarity between these examples, suggesting that non-syntactic factors may be at work.

Grammatical role hierarchy and parallelism preferences are just two of the types of preferences that have been posited in the literature; others have been based on recency, pronominal chains, thematic roles, and semantic plausibility factors. Because the data often give contradictory indications as to the relative importance of each, some researchers have gone a step further and proposed more complex preference-combining systems or competition-based models (discussed in Section 6.6). The complexity of this type of approach is then compounded by their use within a procedure which performs a 'search' for a missing referent, which in some cases involves a combinatorially extensive manipulation and filtering process over possible configurations of coreference assignments.

All this leads to what we might take to be 'the big question' concerning pronoun interpretation: If resolving a pronoun is really as complicated as the literature would seem to imply, why would anyone ever use one? After all, in choosing to use a pronoun, a speaker would be electing to use a potentially highly ambiguous expression that may require a computationally intensive effort on the part of the hearer to resolve, rather than a less ambiguous or even unambiguous one that would presumably not (such as, for example, a proper name). However, as I briefly discussed in Chapter 3, the evidence suggests that discourse interpretation is *facilitated* by the use of pronouns, not hindered by it. It would seem that an adequate account of pronominal reference should reflect this fact.

In this chapter, I again show that although the data appear contradictory within the purview of current theories, it ceases to be so when the inference processes underlying establishment of coherence relations

are taken into account. I will describe three central types of analysis that have been posited in the literature, coupled with three types of preferences that these approaches have been designed to incorporate, and demonstrate that they correspond directly to my neoHumian trichotomy of coherence relations. I then show how these data can be accounted for through the interaction between the linguistic properties of pronouns and the properties of the inference processes underlying the establishment of the different types of relations. By describing the manner in which pronoun interpretation fits within a broader and independently motivated theory of discourse processing and comprehension, the theory can explain why the aforementioned preferences appear to exist (thus rendering them epiphenomenal), rather than having to posit them as primitive heuristics. Furthermore, the theory yields an explanation for why different preferences appear to be operative in different contextual circumstances, eliminating the need to rely on complicated preference-combination or competition-based models of salience and reference. The results cast doubt on theories of pronoun reference that rely on processes that perform an extensive search for a missing referent, and establish the untenability of models that perform discourse state update on a sentence-by-sentence or clause-by-clause basis. While there are still aspects of the theoretical infrastructure on which the analysis relies that remain to be worked out in full – such as the details of an on-line processing model of coherence establishment (see Chapter 8), and a more sophisticated conception of how different predicates associate varying levels of salience to their arguments – I proceed under the assumption that even preliminaries to an adequate theory are more useful than working out the specific details of a more superficial theory that is demonstrably inadequate.

## 6.1 A Coherence-Driven Approach

Over twenty years ago, Hobbs (1979) presented what one might consider to be the most elegant analysis of pronoun resolution offered to date. In his "coherence-driven" theory, pronominal reference resolution is not an independent process at all, but instead results as a by-product of more general reasoning about the most likely interpretation of an utterance, including the establishment of coherence relations. Pronouns are modeled as free variables in logical representations which become bound during these inference processes; potential referents of pronouns are therefore those which result in valid proofs of coherence.

A typical type of example used to support a coherence-driven theory is given in passages (351a) and (351b), adapted from an example from

Winograd (1972).

(351) The city council denied the demonstrators a permit because...

a. ...they *feared* violence.

b. ...they *advocated* violence.

Hearers appear to have little difficulty resolving the pronoun *they* in each case, despite the fact that it refers to *the city council* in sentence (351a) and *the demonstrators* in sentence (351b). In contrast to examples like (350a-b) which provide evidence for subject referents over object referents, different assignments result for examples (351a-b) despite the fact that they have identical syntactic configurations. In fact, the only difference is the verb used in the second clause, suggesting that semantics and world knowledge are the key factors in determining the correct referents.

In Hobbs's account, the correct assignment for the pronoun in each case falls out as a side-effect of the process of establishing the Explanation relation, as signalled by *because*.

> **Explanation**: Infer $P$ from the assertion of $S_1$ and $Q$ from the assertion of $S_2$, where normally $Q \rightarrow P$.

Oversimplifying considerably, I will code the world knowledge necessary to establish Explanation for (351) within a single axiom, given in (352). (See Hobbs et al. (1993, p. 111) for a more detailed analysis of a similar example.)

(352) $fear(X,V) \wedge advocate(Y,V) \wedge enable\_to\_cause(Z,Y,V) \supset deny(X,Y,Z)$

This axiom says that if some $X$ fears some $V$, some $Y$ advocates that same $V$, and some $Z$ would enable $Y$ to bring about $V$, then $X$ may deny $Y$ of $Z$. To make this more concrete, the possible instantiation of this rule that is relevant for example (351) would say that if the city council fears violence, the demonstrators advocate violence, and a permit would enable the demonstrators to bring about violence, then this might cause the city council to deny the demonstrators a permit.

The first sentence in (351) can be represented with the predication given in (353).

(353) $deny(city\_council, demonstrators, permit)$

This representation matches the consequent of axiom (352), triggering a process of abductive inference that can be used to establish Explanation. At this point, $X$ will become bound to *city_council*, $Y$ to *demonstrators*, and $Z$ to *permit*.

Each of the follow-ons (351a-b) provides information that can be used to help 'prove' the predications in the antecedent of the axiom,

thereby establishing a connection between the clauses. Clause (351a) can be represented with predicate (354), in which the unbound variable $T$ represents the pronoun *they*.

(354) $fear(T, violence)$

When this predicate is used to match the antecedent of axiom (352), the variables $T$ and $X$ are necessarily unified. Since $X$ is already bound to *city_council*, the variable $T$ representing *they* also receives this binding, and the pronoun is therefore resolved.

Likewise, clause (351b) can be represented as predicate (355).

(355) $advocate(T, violence)$

This predicate also matches a predicate within the antecedent of axiom (352), but in this case, the variables $T$ and $Y$ are unified. Since $Y$ is already bound to *demonstrators*, the representation of *they* also receives this binding.

Thus, the correct referent for the pronoun is identified as a by-product of establishing Explanation in each case. The crucial information determining the choice of referent is semantic in nature, rooted in the establishment of the relationship between the predication containing the pronoun and the predication containing the potential referents. The fact that coreference came "for free" captures the effortlessness with which people appear to be able to interpret pronouns, offering a potential explanation for how the choice to use of pronoun can actually facilitate, rather than hinder, the process of discourse comprehension.

It would certainly be nice if this was the end of the story. Unfortunately, not all data are as supportive of the tenability of a purely coherence-driven approach.

## 6.2 Attention-Driven Approaches

Contrasting with coherence-driven theories are what I call "attention-driven" theories. Instead of considering pronoun resolution to be a side-effect of more general interpretation mechanisms, attention-driven theories treat pronominal reference as an independent process with its own mechanisms for resolution (Sidner 1983, Kameyama 1986, Brennan et al. 1987, Grosz et al. 1995, inter alia). Just as one could characterize coherence-driven theories as being rooted in people's general cognitive ability to view their world as coherent, one could characterize attention-driven theories as rooted in their cognitive capacity to dynamically (re)focus their attention "on the subset of knowledge relevant to a particular situation" (Grosz 1977). With respect to the kind of focusing necessary to comprehend a discourse, pronominalization can be

considered to be a signal to the hearer that its referent is highly activated and attended to in the hearer's mental state.[23] To model this fact, attention-driven theories contain an explicit component of discourse representation that tracks a level of focus or salience of potential referents, against which pronoun interpretation is performed.

Salience ranking in attention-based theories is usually determined primarily on the basis of surface syntactic information. Recall that we saw the effect of syntactic information in examples (350a-b), in which the preferred referent of the pronoun *him* corresponded to the entity that was mentioned in the subject position of the previous clause. Whereas passages (351a-b) seem to suggest that semantics is the primary determinant of how a pronoun gets interpreted (these examples keep the relevant syntactic relationships constant), passages (350a-b) suggest that grammatical role is the primary determinant (these examples keep semantics roughly constant). The latter set of examples therefore indicates that there is more to pronoun resolution than reasoning with the semantic content of the sentences.

I will use the centering theory of Grosz et al. (1995, henceforth GJW) as an exemplar of attention-driven theories. Centering, an outgrowth of earlier focus-based work on pronoun interpretation, is largely motivated by two related facts about language that are not explained by purely content-based models of reference and coherence, such as that of Hobbs (1979). The first of these is that the coherence of a discourse does not depend only on semantic content but also on the type of referring expressions used. GJW illustrate this point with passage (356), which is meant to be interpreted as part of a longer segment that is currently centered on John.

[23] A rather obvious problem shared by this characterization and a range of others that have been offered is the existence of *cataphora*, in which a pronoun precedes its antecedent:

(i) In his office, Norman read a chapter about pronouns.

Here, the pronoun *his* can be felicitously used to refer to Norman, despite the fact that the latter has not even been mentioned, let alone is salient, at the time the pronoun is uttered. In some circumstances, cataphora can even operate across coordinated clauses; the following data are from Reinhart (1983).

(ii) She has the whole city at her disposal and Rosa just sits at home.

(iii) He hasn't contacted me, but I'm sure John is back.

Reinhart, citing a talk by Mittwoch (1979), claims that it is easy to get such cataphoric reference if the first conjunct is 'pragmatically subordinated' to the second, which is presumably the case in (ii-iii). I believe that such examples shed an instructive light on the manner in which discourse processing is carried out, but will have nothing further to say about this here. Suffice it to say that the implicit assumptions underlying many attention-driven theories of pronominal reference in discourse require some proviso to account for cataphora.

(356) a. He has been acting quite odd. [He = John]
b. He called up Mike yesterday.
c. John wanted to meet him quite urgently.

The third sentence in this passage is quite odd, despite the fact that the pronoun *him* in (356c) is readily interpretable as referring to Mike. Intuitively, the oddness of this sentence stems from the choice of referring expressions used, in particular, the fact that the entity that is more central to the discourse (*John*) is not referred to with a pronoun whereas the less central element (*Mike*) is. Passage (356) can be compared to the similar passage in (357).

(357) a. He has been acting quite odd. [He = John]
b. He called up Mike yesterday.
c. He wanted to meet him quite urgently.

The final sentence in this version sounds perfectly acceptable, although the only difference between it and (356c) is that the mention of John is also pronominalized. These passages demonstrate that a speaker's choice of referential form affects the coherence of a discourse, a fact that is not captured in a purely coherence-driven theory.

The second motivating factor for centering is the existence of garden-path effects in pronoun interpretation, in which a pronoun appears to be interpreted before adequate semantic information has become available. GJW present passage (358) as an example.

(358) a. Terry really goofs sometimes.
b. Yesterday was a beautiful day and he was excited about trying out his new sailboat.
c. He wanted Tony to join him on a sailing expedition.
d. He called him at 6AM.
e. He was sick and furious at being woken up so early.

The passage is perfectly acceptable until sentence (358e), which causes the hearer to be misled. Whereas semantic plausibility considerations indicate that the intended referent for *He* is Tony, hearers tend to initially assign Terry as its referent, creating a garden path effect. Such examples provide further evidence that more is involved in pronoun interpretation than simply reasoning about semantic plausibility. In fact, they suggest that hearers assign referents to pronouns at least in part based on other factors, before interpreting the remainder of the sentence.

In GJW's centering theory, each utterance in a discourse has exactly one backward-looking center (denoted $C_b$) and a partially-ordered set of forward-looking centers $(C_f^1, ..., C_f^n)$. The notation $C_b(U_n)$ is used

to refer to the $C_b$ of utterance $n$, and $C_f(U_n)$ to refer to the $C_f$ list of utterance $n$. Following Brennan et al. (1987), I refer to the highest-ranked forward-looking center $C_f^1$ of utterance $n$ as the *preferred center*, or $C_p(U_n)$. Roughly speaking, $C_f(U_n)$ contains all entities that are referred to in utterance $n$; amongst this list is $C_b(U_n)$. $C_b(U_{n+1})$ is the most highly ranked element in $C_f(U_n)$ that is realized in $U_{n+1}$. The rules specifying how entities are ranked on the $C_f$ list are not fully developed, but factors that have been suggested to affect ranking include surface order, grammatical role, and pronominalization. Three intersentential relationships between a pair of sentences $n$ and $n+1$ are defined:

1. Center Continuation: $C_b(U_{n+1}) = C_b(U_n) = C_p(U_{n+1})$.
2. Center Retaining: $C_b(U_{n+1}) = C_b(U_n)$, but $C_b(U_{n+1}) \neq C_p(U_{n+1})$.
3. Center Shifting: $C_b(U_{n+1}) \neq C_b(U_n)$.

Using these definitions, GJW posit the following two rules, which impose constraints on center realization and movement respectively.

**Rule 1:** If any element of $C_f(U_n)$ is realized by a pronoun in $U_{n+1}$ then the $C_b(U_{n+1})$ must be realized by a pronoun also.

**Rule 2:** *Sequences* of continuation are preferred over *sequences* of retaining; and sequences of retaining are to be preferred over sequences of shifting. In particular, a pair of continuations across $U_n$ and across $U_{n+1}$, represented as $\text{Cont}(U_n,U_{n+1})$ and $\text{Cont}(U_{n+1},U_{n+2})$ respectively, is preferred over a pair of retentions, $\text{Ret}(U_n,U_{n+1})$ and $\text{Ret}(U_{n+1},U_{n+2})$. The case is analogous for a pair of retentions and a pair of shifts.

Rule 1 is posited to capture the oddness of passage (356) as compared to passage (357) discussed above; in (356c), the $C_b$ (John) is not pronominalized whereas a non-$C_b$ (Mike) is. The examples GJW give to illustrate Rule 2 are shown in passages (359) and (360).

(359) a. John went to his favorite music store to buy a piano.
b. He had frequented the store for many years.
c. He was excited that he could finally buy a piano.
d. He arrived just as the store was closing for the day.

(360) a. John went to his favorite music store to buy a piano.
b. It was a store John had frequented for many years.
c. He was excited that he could finally buy a piano.
d. It was closing just as John arrived.

Like passages (356) and (357), passages (359) and (360) express similar propositional content, yet they are not equally coherent. Whereas passage (359) consists of a sequence of Continue relations centered on John,

passage (360) consists of movements between Continuing and Retaining, which gives the effect that the passage flips back-and-forth between being about John and being about his favorite music store.

These rules are presented from a language generation perspective, and are not meant to constitute in and of themselves a theory of pronoun interpretation. Indeed, the predictions these rules make about the preferred referents of pronouns are fairly limited. Rule 2 makes no mention of pronominalization at all, and while Rule 1 does, it makes no prediction about the preferred referents of the pronouns in sentence (358d), nor does it predict the garden path effect in sentence (358e). In each case the Rule 1 is satisfied assuming either possible assignment of referents to the pronouns.

Brennan et al. (1987, henceforth BFP), however, describe an algorithm for pronoun interpretation based on centering principles (also utilized in Walker et al. (1994, henceforth WIC)) in which Rule 2 is also used for making predictions for pronominal reference. For ranking $C_f(U_n)$, they specify a hierarchy of grammatical roles in the following order: subject, object, indirect object, other subcategorized functions, and adjuncts. They augment the transition hierarchy by replacing the Shift transition with two transitions, termed Smooth-Shift and Rough-Shift, which are differentiated on the basis of whether or not $C_b(U_{n+1})$ is also $C_p(U_{n+1})$.[24]

3a. Smooth-Shift: $C_b(U_{n+1}) = C_p(U_{n+1})$, $C_b(U_{n+1}) \neq C_b(U_n)$.
3b. Rough-Shift: $C_b(U_{n+1}) \neq C_p(U_{n+1})$, $C_b(U_{n+1}) \neq C_b(U_n)$.

They redefine Rule 2 as follows:

> Rule 2: Transition states are ordered. CONTINUE is preferred to RETAIN is preferred to SMOOTH-SHIFT is preferred to ROUGH-SHIFT.

The resulting transition definitions are summarized in Table 4.

| | $C_b(U_{n+1}) = C_b(U_n)$ or unbound $C_b(U_n)$ | $C_b(U_{n+1}) \neq C_b(U_n)$ |
|---|---|---|
| $C_b(U_{n+1}) = C_p(U_{n+1})$ | Continue | Smooth-Shift |
| $C_b(U_{n+1}) \neq C_p(U_{n+1})$ | Retain | Rough-Shift |

TABLE 4 Transitions in the BFP Algorithm

Given these definitions, their algorithm (as described in WIC) is defined as follows.

[24]The terms Smooth-Shift and Rough-Shift were introduced in WIC.

1. GENERATE possible $C_b$-$C_f$ combinations
2. FILTER by constraints, e.g., contra-indexing, sortal predicates, centering rules and constraints
3. RANK by transition orderings

The pronominal referents that get assigned are those which yield the most preferred relation in Rule 2, assuming Rule 1 and other coreference constraints (gender, number, syntactic, semantic type of predicate arguments) are not violated.

This strategy correctly predicts that *He* and *him* in sentence (358d) refer to Terry and Tony respectively. This assignment results in a Continue relation (Terry would be both the $C_b$ and $C_p$ of (358d)), whereas the Tony/Terry assignment results in a less-preferred Retain relation (Terry would still be the $C_b$, but Tony would be the $C_p$). Their rules also account for the oddness of sentence (358e), since assigning *he* to Tony results in a Smooth-Shift, whereas assigning *he* to Terry results in a Continue.

Thus, the BFP algorithm successfully accounts for the pronominal reference preferences for one of the central motivating examples of centering theory. I should note, however, that the algorithm does not always give results that conform with what one would expect from a purely attention-driven algorithm. Detailed discussion of this issue would take us too far afield for my current purposes; the reader is referred to Kehler (1997) for examples and further discussion. Briefly, the problem is that, as with Hobbs's coherence-driven theory, the preferred assignment for a pronoun in the BFP algorithm cannot necessarily be determined until the entire sentence has been processed. The reason stems from two properties of the algorithm: that determining the transition type between a pair of utterances $U_n$ and $U_{n+1}$ requires the identification of $C_b(U_{n+1})$, and a noun phrase (pronominal or not) can occur at any point in the utterance that will alter the assignment of $C_b(U_{n+1})$. The attention-driven property of the algorithm is therefore lost, compromising its ability to model the effects that result from people's tendency to resolve pronouns at or soon after the time at which they are encountered. I conclude from this that centering-based algorithms should not rely on Rule 2 to determine the correct referents of pronouns.

Setting aside this weakness, it remains the case that an appropriately constructed attention-driven analysis can model the types of effects that were demonstrated in passage (358) in a truly incremental manner. It is much less clear how a purely coherence-based theory could be made to account for the effects resulting from this incrementality.

## 6.3 Parallelism

Up to this point, I have described two types of approach to pronoun interpretation: coherence-driven theories and attention-driven theories. We have seen how the data that support each analysis contradicts the other. I now discuss a third type of example, one which motivates what one might call a parallelism-driven approach. Sidner (1983) illustrates the point with (361).

(361) a. The green Whitierleaf is most commonly found near the wild rose.

b. The wild violet is found near it too.

There is an extremely strong preference to interpret the pronoun *it* in (361b) as the entity specified by *the wild rose*. It is not clear why this should be so, however, in light of the approaches discussed so far. According to theories that consider entities evoked from subject position to be more salient than those evoked from other grammatical positions, the entity specified by *the green Whitierleaf* is more salient than the entity specified by *the wild rose*. Furthermore, there is no semantic reason why *the green Whitierleaf* cannot be the referent of *it*. Intuitively, there appears to be a parallelism effect at play in (361a-b), like the one we saw in (350c), that is not present in examples such as (350a) and (351b). Sidner says:

> Focussing cannot account for the detection of parallel structure, not only because the computation of such structure is poorly understood, but also because focussing chooses different defaults for co-specification than those required for parallelism. (p. 236)

Citing Sidner's claim, Kameyama (1986) discusses example (362) in addition to (361).

(362) a. Carl is talking to Tom in the Lab.

b. Terry wants to talk to him too.

Kameyama suggests that centering theory be augmented with a *property-sharing* constraint which "requires that two pronominal elements realizing the same *Cb* in adjacent utterances share a certain common grammatical property." As stated, this rule falls short of accounting for (361) or (362), since it only applies to sequences of pronominal elements. She then considers extending the rule to cases of "explicitly signalled parallelism" even in cases in which the first element is not a pronominal. The use of the adverbial *too* is such a signal, which thus accounts for the parallelism effect in (361) and (362), although this extension still leaves

example (350c) unaccounted for.

Kameyama backs off from extending her rule to all cases of grammatical role parallelism because of cases like (363), which, like (350a), witnesses a preference for the object position pronoun to refer to the subject of the preceding clause.

(363) a. Carl is talking to Tom in the Lab.

b. Terry was just looking for him.

Thus, both Sidner and Kameyama saw the importance of incorporating parallelism preferences in their analyses, but also noticed the inherent difficulty in reconciling a parallelism-driven approach with an attention-driven one.

## 6.4 A Pattern in the Data

At this point, we are left with three types of preference in pronoun interpretation, based respectively on semantics and coherence, grammatical role hierarchies, and grammatical role parallelism. Each preference is associated with a set of data that supports it, as well as a style of analysis that is designed to account for it. Unfortunately, the data that support each approach appear to be problematic for the others, and thus taken together give contradictory indications when viewed from the purview of existing theories of pronoun interpretation.

In the next section, I argue that this is just what one might have expected. Upon closer inspection, we see that the type of data used to support each type of theory is qualitatively different. In particular, a now familiar pattern can be found with respect to the type of coherence relation that is operative in each set of examples.

First, recall that the preference for subject referents demonstrated by passages (350a-b), repeated below as (364a-b), and the garden path effects in passages like (358), repeated below as (365), are used to support attention-driven approaches.

(364) a. John hit Bill. Mary told him to go home [ him = John ].

b. Bill was hit by John. Mary told him to go home [ him = Bill ].

(365) a. Terry really goofs sometimes.

b. Yesterday was a beautiful day and he was excited about trying out his new sailboat.

c. He wanted Tony to join him on a sailing expedition.

d. He called him at 6AM.

e. He was sick and furious at being woken up so early.

The sentences in these examples are related by the Contiguity relation Occasion. These examples are problematic for coherence-driven accounts because the latter do not take into account the syntactic differences manifest in passages (364a-b), nor do they explain why one would get a garden-path effect in example (365). Likewise, examples like (364a) are problematic for parallelism-based accounts, since the preference for entities evoked from a parallel grammatical position is not respected.

Second, recall that passages such as (361), (362), and (350c), repeated below as (366), (367), and (368), are used to support parallelism-driven approaches.

(366) The green Whitierleaf is most commonly found near the wild rose. The wild violet is found near it too.

(367) Carl is talking to Tom in the Lab. Terry wants to talk to him too.

(368) John kicked Bill. Mary punched him.

In these examples a Resemblance relation is operative, in particular, Parallel. Such examples pose a problem for attention-driven approaches, since a preference for entities evoked from subject position is not enforced. Furthermore, there is no semantic reason for preferring one possible referent over the other; coherence could be established with either referent.

Finally, recall that minimal pairs such as passages (351a-b), repeated below as (369a-b), are used to support a purely coherence-driven approach to resolution.

(369) The city council denied the demonstrators a permit because

a. they *feared* violence.

b. they *advocated* violence.

These examples fall into the class of discourses related by Cause-Effect, in this case, Explanation. The fact that these two passages yield different preferences for the referent of the subject pronoun is problematic for both attention-driven and parallelism-driven analyses, since the passages are equivalent with respect to the syntactic conditions that are relevant to those approaches. Indeed, it is hard to imagine how these differences can be accounted for without appealing to the semantics of the verb of the second clause, and the role that that semantics plays in the establishment of the Explanation relation.

In sum, the three types of examples used to provide evidence for three corresponding approaches to pronoun resolution appear to be mutually contradictory when viewed from the purview of current theories, but in fact actually pattern with my neoHumian categorization of co-

herence relations. This pattern suggests that an adequate analysis of pronoun interpretation will have to account for the inference processes underlying the establishment of these three classes of relation. Given the pervasiveness of the coherence establishment process in discourse interpretation, one might have been surprised if this had turned out not to be the case.

## 6.5 Coherence and Coreference Revisited

My analysis of pronoun interpretation follows those I have presented for the other phenomena discussed in this book, particularly in being based on the interaction of two aspects of interpretation: (1) the linguistic properties of the linguistic form in question, and (2) the properties of the process of establishing coherence for my three types of relation.

The first component of the theory is therefore an analysis of the linguistic properties of pronouns. Here I follow attention-driven theories (and hence differ from coherence-driven theories) by modeling pronouns as linguistic devices in their own right, ones that encode signals to the hearer about the degree of salience the referent holds within the current discourse state. In particular, pronouns encode the signal that this level of salience is high. Theories of what has been alternately referred to as the *accessibility*, *cognitive status* or *information status* of referents, in fact, use this as the primary characteristic that distinguishes pronouns from other choices of referential form that a speaker could choose to employ, for example, demonstratives, definite lexical noun phrases, or proper names (Ariel 1990, Gundel et al. 1993, Lambrecht 1994, van Hoek 1997, inter alia). This property licenses a hearer to interpret a pronoun against the operative discourse state at or soon after the time at which it is encountered.

As I previously discussed, it is this property that creates several of the effects that support an attention-driven theory. First, it can lead to garden-path effects when the expected referent at the time the pronoun is encountered is different from the one supported by subsequently encountered semantic information. Likewise, conflicting signals about the cognitive status of referents – as illustrated by examples like (356c), in which the mention of a less salient element is pronominalized whereas that of a more central one is not – can create confusion on the part of the hearer concerning the configuration of the operative discourse state. As contended by centering theory, therefore, there is a robust set of evidence showing that the choice to pronominalize a referent leads to inferences that impact discourse comprehension, creating effects that would be difficult to model in a theory that merely represents pronouns

as unbound variables within logical expressions.

Just because pronouns are first-class linguistic devices in their own right, however, does not mean that coherence establishment is irrelevant to their interpretation. I claim that an analysis of the pronoun interpretation facts must take into account the effect that coherence establishment has on the operative discourse state. I will start by following an essential tenet of attention-driven approaches, specifically that grammatical constructions tend to mark the occupants of certain grammatical positions as salient (typically the subject position). In the sections that follow, however, I will also show how the process of coherence establishment is capable of redirecting the focus of attention as needed to support the inferencing process, having the effect of altering the relative salience of discourse entities on a relatively small time scale. I will therefore need to move considerably beyond the type of clause-by-clause discourse update mechanism inherent in centering and other attention-driven theories, as such an approach will prove to be far too coarse-grained to properly account for the behavior of pronominal reference. The fact that different types of pronoun interpretation preferences arise in different contextual circumstances can then be shown to result from properties of the different mechanisms required to establish the three types of coherence relations.

### 6.5.1 Pronoun Interpretation during the Establishment of Resemblance

As I described in Chapter 2, the inference processes that underlie the establishment of Resemblance relations begins by identifying sets of parallel entities and relations as arguments to the coherence relation, and then attempts to identify points of similarity and contrast among each set. When an element in one of the (non-initial) clauses of a passage is a pronoun, the coherence establishment process will pair it with its preceding parallel element before attempting to identify maximal similarity between the two. This in turn makes the parallel element highly salient at that point in the processing, and the inference mechanism simply establishes coreference as a way of maximizing similarity.

I step through this process in greater detail by considering example (370), which was discussed briefly in Chapters 1 and 3.

(370) Margaret Thatcher admires Hillary Clinton, and George W. Bush absolutely worships her.

This example, like others previously discussed, demonstrates the strength of parallelism over other considerations when a Resemblance relation is operative. Given our world knowledge about the political figures men-

tioned, it is entirely plausible that Bush worships Thatcher (both are conservative) and highly unlikely that he worships Clinton (who is liberal), and thus common sense knowledge supports selecting Thatcher as the referent. Likewise, Thatcher occupies the subject position whereas Clinton occupies the object position, and thus grammatical role hierarchy preferences would also select Thatcher as the likely referent. Nonetheless, assuming the pronoun remains unaccented, all informants I have surveyed have interpreted *her* to refer to Clinton.[25]

Interestingly, the force of parallelism is so strong that examples like (370) appear to not even generate the type of garden-path effects we saw for examples like (365). One might expect that at some point during or after the processing of (370), hearers would backtrack and ultimately determine that Thatcher was the intended referent of *her* based on semantic plausibility factors. None of my informants, however, showed any evidence of having done this. In general, hearers are much more likely to question the speaker's (dubious) claim that Bush worships Clinton rather than revise their interpretation for the pronoun, despite the availability of a more plausible referent in the subject position of the previous sentence.

As indicated earlier, this fact can be explained by considering the discourse state at the point at which the pronoun is processed during the establishment of the Parallel relation.

**Parallel:** Infer $p(a_1, a_2, ...)$ from the assertion of $S_1$ and $p(b_1, b_2, ...)$ from the assertion of $S_2$, where for some property vector $q$, $q_i(a_i)$ and $q_i(b_i)$ for all $i$.

The semantics for the two clauses (ignoring for simplicity the adverbial *absolutely* in the second clause) can be represented by the predications shown in (371a-b).

(371) a. $admires(Thatcher, Clinton)$
b. $worships(Bush, her)$

During the establishment of Parallel, Bush ($b_1$) is identified as being parallel to Thatcher ($a_1$), and *her* ($b_2$) as being parallel to Clinton ($a_2$). These pairs will then be proven similar with respect to some inferred set of properties $q_i$. Therefore, at the time that the pronoun is encountered, the inference mechanism is focused solely on the parallel element with which it is attempting to prove this similarity (in this case Clinton), making this element the only salient referent in the discourse state at that moment. Establishing coreference between them then yields the

[25] Contrastive accent on the pronoun can have the effect of shifting this preference, of course. See Kameyama (1999, inter alia) for discussion.

maximal possible degree of similarity.

This explanation maintains a critical property of Hobbs's analysis, in that there is no need to posit a distinct pronoun-specific process that performs a 'search' for a missing referent. The referent is simply the entity that the coherence establishment process causes to be highly salient at the time the pronoun is encountered. In this light, we can see how it could be easier to process an (underspecified and potentially ambiguous) pronoun as opposed to less ambiguous form such as a proper name. The use of the pronoun allowed for the rather effortless and immediate establishment of similarity via coreference, whereas the choice of a fuller description may have generated an implicature (per the type of effect seen in example (356), for instance) that may have required additional steps within the proof to address.

This property of the analysis also predicts an even more striking fact that is problematic for all search theories of pronoun interpretation of which I am aware, specifically that the effects of parallelism resist a mismatch of grammatical features (e.g., gender) between a referent and its parallel element (Oehrle 1981, Smyth 1994). For example, consider the modification to sentence (370) given in (372).

(372) Margaret Thatcher admires Ronald Reagan, and George W. Bush absolutely worships her.

All of my informants have confirmed that this example is infelicitous, assuming again that the pronoun remains deaccented. It is not at all clear why this would be the case given a search approach to pronoun interpretation. Not only is Thatcher (i) evoked from the subject position of the previous sentence and (ii) results in a completely plausible interpretation were the pronoun to receive that assignment, but Thatcher is also *the only possible referent which meets the gender restriction of the pronoun.* The pronominal reference is therefore completely unambiguous, yet its use is infelicitous. Of course, a simple search-based algorithm would happily identify Thatcher as the referent, regardless of the preferences it employs.

This infelicity is again explained by the fact that during the establishment of the Parallel relation, *Reagan* is the parallel element of *her*, and is thus *the only* entity which is highly salient at that point during interpretation. Of course, the establishment of coreference is ultimately blocked due to the gender mismatch. The reference is infelicitous since there is no possibility of identifying Thatcher as the referent instead, because she does not satisfy the criterion of being in focus. Unlike the situation for examples like (365), in which a wrong choice between two entities that are both salient enough to be referred to pronominally can

lead to backtracking, in this case no backtracking can change the fact that Thatcher is not salient at the appropriate time during the interpretation of passage (372).

This analysis in terms of coherence-level parallelism is superior to approaches based on grammatical role parallelism in several respects. First, in cases in which a Resemblance relation is operative, it results in essentially the same predictions as grammatical role parallelism heuristics for the types of examples that Sidner and Kameyama discussed, without the need to specify a separate heuristic. Second, and perhaps more importantly, we will see in the next two sections that this account makes the correct predictions for data in which the two approaches diverge, in particular, data in which there is grammatical role parallelism but no Resemblance relation operative. The inference processes used to establish Resemblance relations are not relevant to such cases, and thus no preference for parallelism is expected to apply in my account. These cases include (364a), in which a pronoun in object position preferentially refers to the entity evoked from the subject position of the previous sentence instead of the one evoked from the object position. In this case an Occasion relation is operative. The lack of grammatical role parallelism in Cause-Effect cases such as (369b) will likewise be explained.

This is not to say that one will never see evidence for grammatical role parallelism when relations other than Resemblance are operative. My claim is simply that the appearance of a grammatical role parallelism preference in such cases is epiphenomenal. I illustrate one such case by considering example (373).

(373) a. Carl visited Tom in the lab.

b. He asked him about next week's lecture.

On the most natural reading of (373), *He* refers to Carl and *him* to Tom, and thus grammatical role parallelism exists with respect to both pronouns even though an Occasion relation is operative. However, these assignments also result from the grammatical role hierarchy preferences that I will associate with Occasion in Section 6.5.3, in conjunction with intrasentential coreference constraints. That is, a preference for entities evoked from subject position selects *Carl* as the referent for *He*, and since intrasentential coreference constraints prohibit *him* from coreferring with *He*, Tom is selected as the referent for *him*. Therefore, such examples do not differentiate grammatical role hierarchy and grammatical role parallelism preferences.

I might speculate that it was this fact that led BFP (1987, p. 157) and WIC (1994, p. 223) to claim that structural parallelism constraints of the sort Kameyama proposed are epiphenomenal to the ordering of

the $C_f$ and preference for Continue transitions in their centering-based algorithm. While this is true for cases like (373), their centering algorithm does not account for the parallelism in examples like (370). Here again, assuming that Thatcher is the $C_b$ of the first sentence (or the $C_b$ is undefined if the sentence is discourse-initial), the Centering algorithm will predict that *her* refers to Thatcher. (See also Kehler (1997) for a more elaborated example.) Since there is no mechanism that could alter the discourse state based on the operative coherence relation, the algorithm cannot account for cases in which grammatical role parallelism is found at the same time as those in which it is not.

The examples I have considered thus far include cases in which grammatical role parallelism exists but coherence-level parallelism does not. There are also cases in which the reverse is true. In fact, I noted in Chapter 2 that while the identification of parallel elements for Resemblance relations is aided by the existence of surface syntactic parallelism, ultimately the establishment of these relations is driven by relationships among purely semantic entities and eventualities. One might therefore wonder what preferences for pronoun resolution are found in cases in which there is a Parallel relation but no syntactic parallelism. Consider examples (374) and (375).

(374) Margaret Thatcher admires Hillary Clinton, and she is worshipped by George W. Bush.

(375) Margaret Thatcher is admired by Hillary Clinton, and she worships George W. Bush.

As expected, these sentences are intuitively harder to process than their syntactically parallel counterparts. Yet, the majority of my informants have reported that they nonetheless prefer the interpretation predicted by coherence-level parallelism rather than syntactic parallelism and semantic plausibility – that is, they interpret *she* to refer to Clinton. The voice alternation is identifiable immediately after the subject pronoun in each case, and thus the fact that these two levels of parallelism are not congruent is also identifiable at that time. While this incongruence may hinder the process of identifying Resemblance, the pronominal reference assignments ultimately correspond to what is predicted by a Resemblance analysis.

Finally, the examples I have cited for the Resemblance category up to this point have all been instances of Parallel relations. Of course, my analysis predicts that the other Resemblance relations will show the same strong preference for coherence-level parallelism, since the same type of inference process underlies their establishment. The evidence supports this prediction, as demonstrated in (376a-d).

(376) a. Bill Bradley supported Al Gore, but John McCain opposed him. (Contrast; him=Gore)

b. Bill Bradley supported Al Gore. In fact, many well-known democrats supported him in 2000. (Generalization; him=Gore)

c. Democratic state governors supported Gore and Lieberman in 2000. For instance, Gray Davis supported them. (Exemplification; them=Gore and Lieberman)

d. A retired senator from New Jersey worked hard for Gore's campaign. Bill Bradley, 54, provided strong arguments for his central campaign themes throughout the fall. (Elaboration; his=Gore)

In sum, the process of establishing Resemblance has the side effect of enforcing a strong preference for a pronoun to refer to its parallel element. This preference overrides any preferences for an entity evoked from a particular grammatical role, for an assignment that would result in the most plausible interpretation, and even the effects of grammatical constraints such as gender.

### 6.5.2 Pronoun Interpretation during the Establishment of Cause-Effect

As in the case of Resemblance relations, pronouns can also be seen to assist in the establishment of coherence within Cause-Effect relationships. The inference processes used to establish Cause-Effect relations differ markedly from those used for Resemblance, however. In particular, coherence is established by drawing a path of inference between the semantics of the clauses being related, rather than being rooted in the identification of similar and contrasting properties of parallel subclausal arguments.

In the earlier discussion of Hobbs's coherence-driven approach, we already saw how pronoun interpretation can result from the unification of arguments within axioms used to establish a Cause-Effect relationship. My analysis for these cases is essentially the same as Hobbs's, except that I must take into account the fact that pronouns require that their referents be in focus. Let us step through the analysis of example (351) once again, repeated below as (377).

(377) The city council denied the demonstrators a permit because

a. they *feared* violence.

b. they *advocated* violence.

In my illustration of Hobbs's approach, I assumed that we have the axiom shown in (378).

(378) $fear(X, V) \wedge advocate(Y, V) \wedge enable\_to\_cause(Z, Y, V) \supset deny(X, Y, Z)$

During the process of establishing Explanation in each case, the representation of the first clause, repeated below in (379), instantiates the consequent of axiom (378).

(379) $deny(city\_council, demonstrators, permit)$

This instantiation process activates axiom (378), along with any others sharing a similar consequent, as being potentially relevant to the establishment of Explanation. (For my purposes here, I will take the liberty of focusing only on the axiom that will ultimately be needed to establish coherence; in reality, many such possibilities must be pursued in parallel.) The matching of the consequent binds the variables $X$, $Y$, and $Z$ in the antecedent predications, as shown in (380).

(380) $fear(city\_council, V) \wedge advocate(demonstrators, V) \wedge enable\_to\_cause(permit, demonstrators, V)$

At this point we are ready to process each of the two follow-ons in (377a-b). In (377a), the inference procedure looks to establish a connection between the *fear* predication in the second clause and the corresponding predication in the antecedent of the axiom. This requires that the occupants of their respective argument positions be merged. At the time the representation for *they* is considered, therefore, the occupant of the corresponding argument position is placed in focus – in this case *city_council* – in manner similar to that for the Resemblance examples. At this point, coreference is assumed and Explanation is established.

The process is analogous for the follow-on in example (377b). In this case, the *advocate* predication in the second clause is matched against its corresponding predication in (380). Again, at the time the representation for *they* is considered, the occupant of the corresponding argument position is placed in focus, which in this case is *demonstrators*. Therefore, as dictated by a coherence-driven approach, in each case the pronoun receives the binding that allows the Explanation relation to be established.

This success notwithstanding, the fact that the referent for the pronoun must be in focus at or soon after the time at which it is encountered imposes additional constraints that are not directly modeled by a purely coherence-driven approach. A crucial property of passages (377a-b) is that the information in the second clause that establishes a tie to the appropriate predication in the antecedent of axiom (378) is contained in the main verb, encountered immediately after the pronoun. This property allows the coherence establishment process to redirect attention to

the appropriate argument of this predication in time for the pronoun to be interpreted.

My analysis therefore predicts that garden path effects will still result when a Cause-Effect relation is operative, in cases in which the information necessary to alter the current focus does not come until well after the pronoun is encountered. A purely coherence-based model that operates on sentence-level interpretations would presumably not predict such garden path effects, since there is no process outside of coherence establishment that would cause the pronoun to be resolved before adequate semantic information has become available. In my analysis, the need to resolve a pronoun at or soon after the time at which it is encountered results from the linguistic properties of a pronoun itself.

Such garden path effects are readily found; consider the adapted version of (377b) given in (381).

(381) The city council denied the demonstrators a permit because they decided that the best way to draw attention to issues is to advocate violence.

As in (377b), the pronoun *they* in (381) is intended to refer to *the demonstrators*. Unlike (377b), however, the information that comes soon after the verb does not constrain the interpretation of the pronoun with respect to any relevant axiomatic knowledge. Instead, the crucial information comes well after the pronoun in encountered, and a garden path effect results. Even though the connective *because* indicates that an Explanation relation is operative, there is no information available at or near the time the pronoun is encountered that would cause the inference procedure to redirect the focus of attention to *the demonstrators*. Therefore, continuing under the assumption that there is an initial attentional bias toward the subject *the city council* until something redirects this focus, the hearer interprets it as the intended referent.

Likewise, the possibility of a garden-path appears to increase, albeit perhaps mildly, when the Cause-Effect relationship is not signaled by a connective. Compare the variations of (377a-b) given in (382a-b), which lack the connective *because*.

(382) The city council denied the demonstrators a permit.

a. They *feared* violence.

b. They *advocated* violence.

Although judgments vary, some informants find the pronoun in sentence (382b) to be more jarring than the one in sentence (377b), even though an Explanation relation is ultimately established for each. (See McKoon et al. (1993), for a description of psycholinguistic experiments that es-

tablished this difference for pairs of examples similar to sentences (377b) and (382b).) Although the effect is perhaps not terribly strong, the qualitative difference between (382b) and (377b) suggests that the explicit marking of the Explanation relationship in passages (377a-b) affects the discourse state in a way that facilitates the interpretation of the pronoun. I would attribute this effect to the fact that the establishment of several different possible coherence relations is being pursued at the point during processing at which the pronoun is encountered in (382b), and that the axioms used to establish Explanation (such as (378)) might not be as salient and readily accessible as they are when this relation is signalled explicitly.

Hobbs himself acknowledges the possibility of garden-path effects, concluding that they "strongly suggest that some psychological reality underlies the heuristic" that favors entities evoked from subject position over those evoked from object position. He notes that upon hearing example (383),

(383) John can open Bill's safe. He ...

one is likely to assume that *John* is the referent of *He*. Hobbs attributes this effect to the fact that for most coherence relations each clause has the same agent, and since agents are generally expressed as subjects, the hearer is inclined to make an initial guess that the previous subject is the referent of a succeeding pronoun. Hobbs acknowledges, however, that this point does not fully explain the heuristic. In fact, while he suggests that the insensitivity of his analysis to the voice used in a clause containing a referent argues in favor of the analysis, we have seen that this very distinction can cause a difference in how pronouns are interpreted (recall, for instance, passages (350a-b) in the introduction).

Although I have concentrated on an example in which the operative coherence relation is Explanation, it is easy to construct similar cases involving other Cause-Effect relations that behave analogously. For instance, we can adapt passages (377a-b) so that a Denial of Preventer relationship holds, as in (384a-b).

(384) The city council granted the demonstrators a permit even though

a. they *feared* violence.

b. they *advocated* violence.

The pronouns are readily interpreted in each case, even though the interpretations differ.

Likewise, in Chapter 3 I briefly discussed how the clauses in example (81), repeated below as example (385), can be interpreted as participating in either a Parallel or Result relation.

(385) Colin Powell defied Dick Cheney, and George W. Bush punished him.

As predicted by the analysis, the preferred referent of the pronoun depends directly on which of these coherence relations is inferred. Assuming the Parallel relation, the pronoun must be interpreted to refer its parallel element, in this case Cheney, as predicted by the mechanism discussed in Section 6.5.1. On the other hand, Powell is the preferred referent for the pronoun if a Cause-Effect relation is inferred, in accordance with our world knowledge about the relationship between defying and punishing. In each case the syntactic conditions remain constant; the preferred referent changes only with respect to the coherence relation that is inferred.

In sum, the inference processes underlying the establishment of Cause-Effect can cause a redirection of focus to a referent that will lead to the establishment of coherence. However, any semantic information necessary to trigger this refocusing must be available at or soon after the time that the pronoun is encountered, or else the interpretation of the pronoun will precede this change of focus.

### 6.5.3 Pronoun Interpretation during the Establishment of Contiguity

Finally, we have the case of Contiguity, which consists of the single relation Occasion. Recall from Chapter 2 that the Occasion relation allows one to express a coherent sequence of events centered around a common system of entities. Coherence arises from the inferences required to interpret the initial state of each expressed eventuality as the final state of the preceding one.

**Occasion (i):** Infer a change of state for a system of entities from $S_1$, inferring the final state for this system from $S_2$.

**Occasion (ii):** Infer a change of state for a system of entities from $S_2$, inferring the initial state for this system from $S_1$.

As I indicated in Chapter 2, the establishment of Occasion requires access to the knowledge people gain from experience in typical scenarios and the granularity with which they perceive events and change resulting from them. Much remains unknown, however, about how exactly to represent and reason with such information, or at least how to do so at the level of specificity necessary to distinguish coherent and incoherent Occasions. Without a better understanding of how these inference mechanisms establish coherence, it is difficult to determine what predictions would result for pronoun interpretation in an independently motivated manner. We can, however, use what we do know to begin to sketch out

what such an analysis might look like. In so doing, we will see several ways in which the manner in which of discourse state is represented in current attention-driven theories needs to be extended.

According to its definition, the inference processes underlying the establishment of Occasion attempt to connect two eventualities through an intermediate state. Therefore, when a subject pronoun is encountered, for instance, we would expect first that the referent must be salient within the hearer's mental model of this state. This constraint alone is not enough, however, since we know that it can be perfectly felicitous to pronominalize mentions of more than one entity in an utterance, and thus multiple entities may satisfy the salience criterion when considering a single pronoun. We therefore need to identify the factors that determine, within a hearer's mental model of a system entities and relationships between them, the relative degrees of salience of these entities and relationships.

It is worth clarifying that salience is a property of a conversational participant's conception of an entity with respect to his or her representation of the operative discourse state, and not an inherent property of an entity in the world. As such, the salience of an entity cannot be predicted solely from semantic information. Instead, as attention-driven theories contend, there is ample evidence that merely *how* information is communicated – that is, how it is linguistically "packaged" – affects, among other things, the relative degrees of salience accorded to different entities. There are two information packaging decisions that speakers make which are of interest to us here: choice of syntactic form and choice of referential form.

First, natural languages offer a speaker many syntactic means for expressing a proposition; in English these include active voice constructions, passives, topicalized sentences, sentences with preposed constituents, clefts, and so forth. This very fact suggests that the choice among them has consequences for comprehension, and one way in which it does results from how different constructions indicate the salience of the entities that are mentioned. In some cases, the differences in salience correspond to what entity is placed in subject position, such as the distinction between active and passive voice shown in (386a-b).

(386) a. John hit Bill.

b. Bill was hit by John.

Although (386a-b) communicate the same propositional content, there are information-structural differences between the two. In some intuitive (albeit hard-to-quantify) sense, (386a) seems to emphasize what John did more than what happened to Bill, whereas the reverse is true in

(386b). As such, it is plausible that a hearer's mental model resulting from interpreting either of (386a-b) will differ with respect to the relative degrees of salience attributed to John and Bill.

Indeed, it has been frequently noted in the information packaging literature that subject position is the canonical place from which to mention an entity that serves as a discourse topic (Chafe 1976, Lambrecht 1994, inter alia). Insofar as this claim can be supported on independent grounds, it would explain why we see a tendency for pronouns to prefer entities evoked from subject position in Occasion. However, it is unlikely that the relative salience of entities accorded by different syntactic constructions can be reduced to preferences for entities evoked from particular grammatical roles. One demonstration of this can be seen in certain constructions which indicate a transfer of possession or change of mental state. Stevenson et al. (1994), for instance, suggest that the thematic role that an entity occupies impacts salience in such constructions, based on examples like (387a-b) and (388a-b).

(387) a. John seized the comic from Bill. He...
b. John passed the comic to Bill. He...

(388) a. Ken admired Geoff. He...
b. Ken impressed Geoff. He...

In a set of psycholinguistic experiments, Stevenson et al. found that hearers are more likely to resolve *he* to John in passage (387a), whereas they are more likely to resolve *he* to Bill in (387b). This result for (387b) belies the fact *Bill* is embedded within a sentence-final prepositional phrase, a position normally considered to be much less salient than the subject. The property that these examples share is that the preferred referent occupies the Goal thematic role of its respective predication, whereas the dispreferred entity occupies the Source role. In light of the definition of Occasion, it is not particularly surprising that the Goal of such events is more salient with respect to their final states than the Source, since the Goal is the recipient of the object that was transferred. Likewise, Stevenson et al. found that hearers are more likely to resolve *he* to Geoff in passage (388a), whereas they are more likely to resolve *he* to Ken in (388b). Again, in each case the preferred entities occupy different grammatical roles but the same thematic roles; in this case the entity that occupies the Stimulus role is preferred to the one that occupies the Experiencer role. The fact that these preferences do not respect grammatical role preferences shows that attention-driven theories must account for, in some manifestation, both grammatical and thematic roles.

The second aspect of information packaging relevant to pronoun interpretation is choice of referential form. Just as natural languages offer many syntactic means for expressing an idea, they allow many ways to specify a referent, including proper names, lexical noun phrases, demonstratives, pronouns, and so forth. In the same way that mentioning an entity in subject position has often been considered to indicate discourse topic, pronominalization has often been considered to indicate a continuation of the current topic. The fact that pronouns often occur in subject position is consistent with this proposal.

This might lead us to ask what happens when a mention to the current discourse topic is pronominalized, but a different entity occupies subject position. Consider the following example, discussed by BFP (1987):

(389) a. Brennan drives an Alfa Romeo.
b. She drives too fast.
c. Friedman races her on weekends.
d. She goes to Laguna Seca.

The question is whether *she* in (389d) refers to Friedman or Brennan; a preference for pronouns to refer to entities evoked from subject position suggests Friedman, whereas a preference for pronouns to refer to the current topic suggests Brennan. The BFP algorithm is forced to make a choice, and selects Brennan because the Continue relation that results outranks the Smooth-Shift that results if the pronoun is resolved to Friedman. Other attention-driven algorithms that simply select the most highly-ranked entity with respect to a grammatical role hierarchy will instead pick Friedman. However, I and several informants find this example to be ambiguous. Both Brennan and Friedman are central to the story, and are thus salient enough to license pronominal reference. The ambiguity stems from whether or not the introduction of Friedman in the subject position of (389c) is meant to signal a shift of discourse topic. Thus, conflicting signals may lead to ambiguities regarding the configuration of the discourse state that is operative when a pronoun is encountered.

Up to this point, we have seen that a variety of factors can contribute to the salience of entities within a hearer's mental model of a discourse. It is likely that tracking salience requires a representation of these entities and the relationships between them that is considerably richer than the grammatical role hierarchies commonly used in attention-driven approaches. I now consider another respect in which this is true. We saw in the discussion of CSC violations in Chapter 5 that one can have a sequence of clauses related by Occasion that are coherently centered

around a topic, even if that topic is not mentioned in every clause. This is particularly true when the discourse contains 'scene-setting' clauses which serve to situate the complex of entities and relationships being described. My analysis would therefore predict that such clauses would not impede the ability to pronominalize mentions of previously evoked entities that remain central to the passage. Consider, for example, the variation of passage (365) given in (390).

(390) a. Terry set out for an outdoor excursion on Sunday.
b. It was a beautiful day, hovering around 83 degrees.
c. He was excited about trying out his new sailboat.
d. He wanted Tony to join him on a sailing expedition.
e. The marina is actually very close to Tony's house.
f. He called him at 6AM.
g. He was sick and furious at being woken up so early.

Recall that the centering algorithm computes the centers of a sentence – and as a result, the assignments for pronouns in that sentence – only on the basis of entities evoked in that sentence and the previous one. There are two utterances in passage (390) that are problematic for this constraint. First, sentence (390c) contains a felicitous pronominal reference to Terry, even though he is last mentioned two utterances prior. Second, and even more problematically, the subject pronoun in sentence (390f) preferentially refers to Terry. The problem is that sentence (390e), also a scene-setting clause, mentions Tony. Tony is thus the Cb of this utterance, since he is the only entity mentioned in (390e) that is also mentioned in (390d). This in turn predicts that the subject pronoun in (390f) will refer to Tony, since this assignment results in a Continue. (Again, Terry is outside of the one-sentence window that centering considers.) In actuality, this mention of Tony does not play a salient enough role to change the fact that the Occasion remains centered around Terry at that point in the discourse. Thus, a more sophisticated model of attentional state is necessary, one which maintains an evolving representation of salience in a developing Occasion in a manner capable of capturing the above facts. Based on our understanding of Occasions, the ability to pronominalize the mention of an entity should correlate with the degree to which the (local) discourse remains actively centered around that entity, and not based on a fixed, limited sentential window.

There is little doubt that remedying these deficiencies is a tall order. In many respects, I have only provided preliminaries to a theory of how the coherence of a passage related by Occasion is established and how pronoun interpretation interacts with this process. I nonetheless hope

that I have provided evidence that an adequate account establishing coherence in Occasions is a prerequisite to providing an adequate account of pronoun interpretation in these contexts. In particular, I do not believe that a satisfactory account of discourse state (let alone pronominal reference) can be given that relies only on grammatical roles and other relations computable from superficial cues. Instead, a more elaborated analysis of how sentences make certain aspects of the eventualities they describe salient (cf. Langacker (1991)), and how this is used in establishing connections between sequences of such eventualities in a discourse, is necessary before these fundamental questions about pronoun resolution can be addressed.

### 6.5.4 Summary

To summarize to this point, I hope to have shown that the seemingly contradictory behavior of pronominal reference can be explained by accounting for the interaction of two aspects of discourse interpretation: (1) the linguistic properties of pronouns, and (2) the manner in which these properties interact with the process of establishing coherence. In this analysis, so-called 'preferences' are not first-class resolution heuristics, but instead epiphenomena resulting from the properties of deeper inference processes used to establish different types of coherence relations. As a result, the analysis has the additional advantage of being able to explain why different preferences appear to be operative in different contextual circumstances.

A key property of the analysis is that, like Hobbs's account, it maintains a very simple story for the properties of pronouns. This fact shifts more of the burden of accounting for the data onto the problem of modeling discourse state evolution and update, a move which I believe more adequately reflects the correct division of labor between discourse state maintenance and reference resolution. In particular, such a model must account for the manner in which discourse state changes on a rapid time scale as driven by the inference processes that underlie the establishment of coherence. It is important to acknowledge that such an analysis is not necessitated merely by the need to account for pronoun interpretation; indeed, there is overwhelming evidence to support the existence of these coherence establishment processes that is independent of the behavior of pronouns. By providing an analysis of these coherence establishment processes, I move away from the need to posit complex 'search' procedures as part of pronoun interpretation, from which we can begin to explain how the use of pronouns could actually facilitate, rather than hinder, the process of discourse comprehension.

It is particularly telling that two very basic aspects of cognition come

into play in this account: (i) our ability to interpret our world as displaying coherence, which requires that distinct entities and eventualities are construed as connected, and (ii) our ability to focus our attention on subcomponents of a situation while backgrounding the rest, and to subsequently shift this focus as necessary. Viewed in this light, an attempt to consider the behavior of pronominal reference without an understanding of these larger questions about cognition sets the domain of inquiry far too narrowly.

Needless to say, pinning down the exact details of how all of the relevant factors for pronoun interpretation work together requires an extremely sophisticated model of discourse processing and coherence establishment, one that admits of a left-to-right, online comprehension process in which syntactic, semantic, and discourse processing are all incrementally integrated. While I am aware of no such theory that is formalized to the extent to which we can extract predictions about pronoun resolution, I believe that such a model will ultimately be necessary to account for all the facts regarding pronoun resolution.

## 6.6 Previous Attempts to Resolve Preference Interactions

As I have already noted, the divergence of effects displayed by the pronominal reference data have proven to be recalcitrant for computational linguistic and psycholinguistic theories alike. In some cases, these effects have generated debates about which preferences do and do not exist. In other cases, they have caused researchers to propose 'mixed' approaches, in which several different preferences are posited to apply simultaneously with rules governing their interaction. While space precludes me from offering a comprehensive survey of the pronoun literature relevant to this point, I briefly discuss one such debate as an exemplar, in which the relationship between the preference for referents evoked from subject position and the preference for referents evoked from a parallel grammatical role is at issue.

Crawley et al. (1990) carried out a set of experiments to test the extent to which human subjects apply these two preferences, which they term the *subject assignment strategy* and *parallel function strategy* respectively. Unlike previous studies which conflated these preferences by considering only subject-to-subject reference effects, Crawley et al. studied pronouns in object position to see if they tended to be assigned to the subject or object of the immediately preceding sentence. Passages (391–393) are examples from their stimuli.

(391) Brenda and Harriet were starring in the local musical. Bill was in

it too and none of them were very sure of their lines or the dance steps. Brenda copied Harriet and Bill watched her.

(392) Claire and Beverley usually went to town together on Saturdays. They sometimes stopped to see their old headmaster on the way back. Claire envied Beverley and the headmaster warned her about it.

(393) Shirley and Carol were organizing the Christmas pantomime. They had persuaded Martin to help and asked him to get in touch. Shirley wrote to Carol about a meeting and Martin phoned her.

They found that in two task environments – a question answering task which revealed how the human subjects interpreted the pronoun, and a referent naming task in which the subjects identified the referent of the pronoun directly – the human subjects resolved pronouns to the subject of the previous sentence more often than the object. They concluded that these results "clearly support the use of a subject assignment strategy rather than a parallel function strategy" (p. 256).

Smyth (1994), however, subsequently argued against Crawley et al.'s result, taking issue with the stimuli they used. He claims that the sentences used were not fully parallel, and that the preference for parallel function (PF) "applies optimally to fully parallel structures, and that SA (=the preference for subject assignment) is the default when PF fails". His definition of "fully parallel" requires that the syntactic structures and verb types across the clauses be identical. He cites only four of the discourse-final sentences from Crawley et al.'s corpus that he considers to meet these criteria, listed in (394a-d).

(394) a. John pushed Sammy and Evelyn kicked him.

b. Sarah visited Cathy at home and Charles phoned her at work.

c. Robert bullied Peter and Melanie attacked him.

d. Cheryl spoke to Monica about the next meeting and Steven questioned her about it.

Smyth then ran experiments using a set of stimuli that met his criteria for parallelism, and found that "subjects overwhelmingly followed PF for both subject and nonsubject pronouns". Based on his findings, Smyth posits an "Extended Feature Match Hypothesis" in which "the fundamental assumption is that pronoun assignment is a search process based on feature matching". He says:

> I conclude from these observations that pronoun interpretation in conjoined sentences involves an obligatory search for a morphologically compatible antecedent which meets the

> binding theory (Chomsky, 1981) criteria for coreference and which, in addition, has the same grammatical role as the pronoun. If a match is found, the the parallel interpretation is obligatory, unless the pronoun is stressed, in which case it is selectively blocked. If no match is found, resolution is less certain, but will most often result in SA, although if the pronoun or the first clause verb is stressed, alternative strategies govern the selection of an antecedent.
>
> On this view, SA is a default strategy for sentences in which the degree of nonparallelism exceeds some limit; PF is a specific outcome of the more general principle that the probability of parallel resolution depends on the number of features shared by the pronoun and the candidate antecedents. Retaining SA in the model allows us to account for an otherwise mysterious asymmetry between subject and nonsubject pronouns. (pp. 204–205)

This rather complex and indeterminate characterization of the behavior of pronominal reference exemplifies the type of analysis that I have been arguing against in this chapter; for one it is hard to see how pronouns could have the effect of facilitating discourse comprehension under this view. The mixed behavior of pronoun resolution found in these studies appeared to necessitate such an explanation, however, and subsequent works have in fact posited other preference-combination analyses. For instance, Stevenson et al. (1993) considered three interpretation strategies – a subject assignment strategy, a parallel grammatical role strategy, and parallel order-of-mention strategy – and concluded that all three may constrain the assignment of ambiguous pronouns. In particular, they posit that each of these strategies produces a candidate referent, with a 'competition mechanism' (McWhinney et al. 1984) selecting the final choice among these.

Similarly, Stevenson et al. (1995) carried out a set of question answering experiments to determine whether the subject assignment and parallel function strategies jointly contribute to the interpretation of pronouns. Building on the Crawley et al. and Smyth experiments, they included cases of subject and non-subject pronouns in their stimuli, constructing three sets of examples. The first set contained a subject position pronoun with two possible referents, one in a parallel subject position and one in a nonsubject position. In such cases, the two strategies at issue both indicate a preference for the subject referent. The second set contained a nonsubject pronoun with two possible referents, one in subject position and one in a nonsubject position different from the one

occupied by the nonsubject pronoun. In these cases, the subject assignment strategy selects the subject referent, whereas their formulation of the parallel function strategy does not apply. The third set contained a nonsubject pronoun with two possible referents, one in subject position and one in a grammatically parallel non-subject position. In such cases, the two resolution strategies disagree as to the correct referent.

Stevenson et al. found that subjects more often resolved the pronoun to the subject position entity when both strategies indicated that preference, as compared to when only the subject assignment strategy applied. Furthermore, they found that nonsubject assignments were actually preferred when the two strategies disagreed. They conclude from this that the subject assignment and parallel function heuristics operate jointly, and that this in turn "implies a model of discourse processing in which a number of constraints compete in the interpretation of noun phrases".

This debate has therefore generated a fairly substantial body of work (of which I have only discussed a sample), with each side concluding that both strategies interact in a way that is still not entirely clear. Based on an informal analysis of the stimuli used in these experiments in light of the account of pronoun interpretation I have presented in this chapter, I strongly suspect that a significant contributor to the confusion surrounding these preferences is the type of coherence relation that the subjects inferred for each example. In particular, I would predict that there is a strong correlation between the appearance of a subject assignment strategy and the inference of Occasion, and likewise between the appearance of a parallel function strategy and the inference of Parallel. The fact that subjects were very likely to have inferred Occasion for some examples and Parallel for others within the same subset of stimuli may have introduced noise that, if eliminated by controlling for the coherence relation inferred, would have resulted in much sharper distinctions in the final results.

While I have no way to recover this information from a post-hoc analysis of their data, the predictions of my analysis do accord with many of the facts noted in these works. For instance, it is probably not a coincidence that the examples from Crawley's data that Smyth found acceptable also tend to favor the Parallel interpretation; of examples (394a-d) I find that only (394d) also has Occasion as a likely interpretation. These can be compared with four examples used by Crawley et al. that Smyth cites as objectionable:

(395) a. Patricia gave Martha a present and Nicholas smiled at her.

b. Mary helped Julie change the wheel and Peter talked to her.

c. Shirley wrote to Carol about a meeting and Martin phoned her.
d. The waitress followed the woman and the manager gave her a menu.

While Smyth's objection stems from the fact that the pronouns are not in a grammatical position that is strictly parallel to their potential non-subject antecedents, I also find that the Occasion relation is more likely to be inferred in these cases. More importantly, insofar as each of these examples can receive either reading, I find that the referent preferentially assigned to the pronoun differs with respect to the relation that is inferred in just the manner that is predicted by my analysis. In each case an Occasion interpretation implies a subject assignment, whereas a Parallel interpretation implies a parallel function assignment.

My account therefore suggests that the need for preference combination analyses will be obviated once the operative coherence relation is taken into account, which I believe is a promising direction for further psycholinguistic experimentation. If shown to be true, it would demonstrate that the emphasis on superficial structural cues that characterizes many previous psycholinguistic studies is largely misplaced; the relationship between such cues and pronoun interpretation preferences would be shown to be mediated by the coherence relation inferred between the clauses. That is, while it is undoubtedly the case that structural cues help to trigger certain ways of establishing coherence between sentences (for instance, the way that full structural parallelism and similarities in verb class may help trigger the recognition of the Parallel relation), the manner in which pronouns are interpreted would be shown to ultimately derive from the coherence relation inferred, and not as a direct consequence of these superficial cues.

Interestingly, there is a discussion in Smyth's paper that suggests that he in fact recognized the role that coherence plays in pronoun interpretation. He addresses passages in which Cause-Effect relations are operative, such as (396a), which he compares to (396b).

(396) a. Phil tickled Stanley, and (so) Liz poked him. (*him* = Phil)
b. Phil tickled Stanley, and Liz poked him. (*him* = Stanley)

As Smyth points out, the causal reading indicated in sentence (396a) makes the nonsubject assignment less felicitous than subject assignment, whereas in sentence (394b) the nonsubject assignment is preferred under the parallel interpretation. He posits that "the semantic structure of a conjunction ... imposes constraints on the discourse model that the listener constructs to interpret the sentence", following with "in some cases, a conjunction can introduce a pragmatic bias which is incom-

patible with a PF interpretation". He lists five possible relations that *and* is compatible with: *Unspecified, Simultaneity, Sequence, Simultaneous Causality*, and *Causal Sequence*, arguing that "sentences with *and* are therefore vague without pragmatic or contextual support" (p. 209). Despite this apparent acknowledgment of the effect of coherence, however, Smyth maintains the characterization of pronoun interpretation as a feature-matching process. In his analysis, PF is the limiting case in which all the relevant features match. He does not consider the possibility that his PF strategy corresponds only with the Parallel meaning of *and*, and that the conflicting preferences arise in just those cases in which a different coherence relation holds. (This possibility does receive a brief discussion in Chambers and Smyth (1998).)

Psycholinguists are not the only researchers who have attempted to account for the pronominal reference data with a preference-combining approach. In a computational treatment, Kameyama (1996) offers an analysis which integrates the effects of four now-familiar preferences, paraphrased below.

1. Subject Antecedent Preference: prefer a subject referent over a non-subject one.
2. Pronominal Chain Preference: prefer a referent that was also pronominalized.
3. Grammatical Parallelism Preference: prefer a referent that occupies a similar grammatical role.
4. Commonsense Preference: prefer a referent that leads to a plausible interpretation.

Kameyama offers the following examples in support each of these preferences; the reference assignments shown in brackets indicate the most preferred referents found in a survey that she performed. Examples (350a) and (350b), discussed previously and repeated below as (397) and (398), provide support for the *subject antecedent* preference.

(397) John hit Bill. Mary told him to go home [ him = John ].

(398) Bill was hit by John. Mary told him to go home [ him = Bill ].

Again, while the propositional content in both of these cases is the same, the preferred referent alternates with respect to which entity is placed in subject position.

Examples (399) and (400) are cases in which the *pronominal chain* preference applies.

(399) Babar went to a bakery. He greeted the baker. He pointed to a blueberry pie. [ He = Babar ]

(400) Babar went to a bakery. The baker greeted him. He pointed to a blueberry pie. [ He = the baker ]

In passage (399), both the subject antecedent preference and the pronominal chain preference predict the correct referent. In passage (400), the two preferences make different predictions; in this case the subject antecedent preference won out (although three out of thirteen respondents chose *Babar* as the referent).

Example (401) is of the now familiar sort that supports the *grammatical parallelism* preference.

(401) John hit Bill. Mary hit him too. [ him = Bill ]

Finally, example (402) supports the *commonsense* preference; in this case semantic information appears to override grammatically-based preferences.

(402) John hit Bill. He was severely injured. [ him = Bill ]

There are clear parallels between these four preferences and my three types of coherence relation. Kameyama's *subject antecedent* and *pronominal chain* preferences correspond to two preferences in attention-driven theories such as centering: a preference to refer to the entity which occupies the most salient grammatical role (the $C_p$), and a preference to refer to the current topic of the discourse (the $C_b$). (Recall the previous discussion of the ambiguity created by these two options with respect to example (389).) The clauses in these examples are related by the Contiguity relation Occasion. Kameyama's *grammatical parallelism* preference is supported by example (401), in which the Resemblance relation Parallel is operative. Lastly, her *commonsense preference* is supported by example (402), in which a Cause-Effect relationship holds. Note that in this case the operative Explanation relationship is not even explicitly marked.

As Kameyama notes, these preferences often contradict. To resolve the conflicts, she provides a list of *preference interactions* that specify which preferences have precedence over others. As I have established, however, the overriding preference depends on the nature of the coherence relationship between the clauses; therefore we would not want to apply the grammatical parallelism preference when a Cause-Effect relationship is operative, for example. A preference-combining scheme which applies all preferences in all contexts will not ultimately succeed in accounting for all of the data, and a much cleaner system results when coherence relationships are taken into account.

## 6.7 Conclusion

To summarize this chapter, the literature on pronoun resolution bears a similarity to those for the other phenomena I have discussed in this book, one that once again brings Maslow's Maxim to mind. Many instances of previous work have operated on an implicit assumption that a single pronoun interpretation strategy can explain pronominal reference behavior in all contextual circumstances. In light of the empirical problems we have seen this assumption cause, other researchers have posited that several strategies are employed that 'compete' in some potentially unwieldy and perhaps unpredictable fashion. As a result, we see a continuing tradeoff between empirical coverage and theoretical elegancy in these proposals.

I have shown how different sets of data used to support competing analyses of pronoun resolution correlate with my independently-motivated distinction among Hume's three classes of coherence relation. As such, a theoretically elegant yet more empirically adequate account results from demonstrating how relatively basic properties of pronouns interact with the processes underlying the establishment of these coherence relations. The analysis is compatible with the fact that pronouns usually have the effect of facilitating, rather than hindering, the discourse comprehension process, a fact that is difficult to account for in both search-based and preference-combining approaches. It also shows that so-called 'preferences' should not be viewed as fundamental resolution strategies, but are instead epiphenomenal by-products of the characteristics of deeper inference processes. As a result, the analysis likewise accounts for why different preferences appear to be operative in different contextual circumstances. While this theory would benefit from further psycholinguistic experimentation studying the correlation between pronoun interpretation and coherence relations, I hope to have brought to light the importance of controlling for coherence relations in such studies.

# 7

# Coherence and Tense Interpretation

In the previous chapter, we compared and contrasted two fundamental types of approach to modeling pronominal reference. We saw how attention-driven theories treat pronouns as first-class linguistic devices with specific anaphoric properties, whereas pronoun interpretation within coherence-driven theories results as a by-product of the process of coherence establishment. Ultimately, my theory incorporated aspects of both types of analysis.

In this chapter, I analyze the problem of inferring temporal relations from tense in discourse. We will see that one can distinguish between two basic types of approach in the tense literature also. I describe an analysis in each category as an exemplar, and then provide a basic set of data that reveals problems with each. I then explain this data using an analysis in which the constraints imposed by tense interact with the constraints imposed by coherence relations in a relatively transparent manner.

Although it makes crucial use of the constraints imposed during the establishment of coherence, my theory of tense interpretation is different from those for the other phenomena addressed in this book in that the constraints do not correlate in any fundamental way with my neoHumian trichotomy of coherence relations. This results from the fact that only the temporal constraints imposed by coherence relations, and not the types of inference processes underlying their establishment, are relevant to the analysis.

It must be made clear that what follows is not intended by any stretch to be a comprehensive analysis of tense and temporal relations. The problem is vast, and requires that one address a variety of related phenomena, including aspectual form and class, modality, time and temporal intervals, quantification and scoping, event structure, temporal connectives and adverbials, and discourse structure, among others.

My discussion will instead focus primarily on only the simple past and past perfect tenses, the meaning of which will be cast in terms of relations over temporal points. Furthermore, for now I only consider cases in which these tenses coexist with (Vendlerian) accomplishments and achievements, in passages with linear discourse structures and without interaction with other temporal modifiers. I will touch on a subset of these remaining issues in Section 7.5, which includes brief discussions of extensions to future tenses, to examples with stative predicates, and to inferring temporal relations over larger discourse structures. In the meantime, focusing on a constrained subcomponent of the problem will allow me to concentrate on my central point, which is that a theory of tense and an account of the temporal relations imposed by coherence relations are both necessary to explain the temporal relations that hearers infer between eventualities in a discourse.

## 7.1 Two Approaches to Tense Interpretation

The identification of temporal relations in discourse has been a popular topic in theoretical linguistics (Partee 1984, Comrie 1985, Dowty 1986, Hinrichs 1986, Nerbonne 1986, Kamp and Reyle 1993, inter alia) as well as in a computational linguistics setting (Passonneau 1988, Webber 1988, Kameyama et al. 1993, Lascarides and Asher 1993, inter alia). Several researchers (Partee 1984, Hinrichs 1986, Nerbonne 1986, Webber 1988) have sought to explain the temporal relations induced by tense by treating it as anaphoric, drawing on Reichenbach's separation between event, speech, and reference times (Reichenbach 1947). On the other hand, Lascarides and Asher (1993) take the view that temporal relations are resolved purely as a by-product of reasoning about the coherence relations that hold between utterances, and in doing so, argue that treating simple and complex tenses as anaphoric is unnecessary. The Temporal Centering framework (Kameyama et al. 1993) integrates aspects of both approaches, but patterns with the first in treating tense as anaphoric. Kamp and Reyle (1993, p. 528) briefly consider the need for an account that utilizes coherence relations, but they do not pursue the idea in detail.

To limit the scope of this discussion and make it more concrete, I focus on two analyses: the tense-as-anaphora approach of Hinrichs (1986), and the coherence-driven approach of Lascarides and Asher (1993, henceforth, L&A). I illustrate them using examples (403a-b), taken from L&A.

(403) a. Max slipped. He spilt a bucket of water.
b. Max slipped. He had spilt a bucket of water.

Passage (403a), which contains two clauses in the simple past, is typically

understood as an Occasion. Therefore, the events are understood to have occurred in the order in which they are presented in the text, that is, with the spilling occurring after the slipping. On the other hand, the natural interpretation of the clauses in passage (403b) is one in which they participate in an Explanation relation. In this case, the events are understood to have occurred in the order opposite to order of presentation, that is, with the spilling occurring before the slipping.

### 7.1.1 A Tense-As-Anaphora Approach

Approaches that treat tense as anaphoric generally utilize some form of Reichenbach's (1947) analysis of tense. Reichenbach distinguishes between three times associated with an event denoted by an utterance: the *point of speech* ($S$), which is the time that the utterance was made, the *point of the event* ($E$), which is the time that the event is taken to have occurred, and the *point of reference* ($R$), which provides a temporal perspective point from which to interpret the point of the event. Different tenses express different sets of relationships between these times. Reichenbach's relations for certain tenses are shown in Table 5.

| Tense | Relations | Example |
|---|---|---|
| Present | $E =_t R =_t S$ | see |
| Past | $E =_t R <_t S$ | saw |
| Future | $S <_t R =_t E$ | will see |
| Past Perfect | $E <_t R <_t S$ | had seen |
| Future Perfect | $S <_t E <_t R$ | will have seen |
| Future Perfect | $S =_t E <_t R$ | will have seen |
| Future Perfect | $E <_t S <_t R$ | will have seen |

TABLE 5 Temporal Relations in Reichenbach's System

Reichenbach's analysis can be utilized in treating tense as anaphoric by taking the reference time $R$ of an event to cospecify some previously evoked event time $E$. For instance, in example (403b), the reference time of the spilling can be taken to cospecify the event time of the slipping. Since events introduced with the past perfect have their event time ordered before their reference time as indicated in Table 5, the second event in passage (403b) will be understood to occur before the first. That is, with the times as marked in (404a), the temporal relations inferred are as indicated in (404b).

(404) a. Max slipped ($E_1$). He had ($R_2$) spilt ($E_2$) a bucket of water.

b. $R_2 =_t E_1 \ \wedge \ E_2 <_t R_2 \ \Rightarrow \ E_2 <_t E_1$

With this interpretation, the Reichenbachian account makes the correct predictions for (403b). However, as it stands, the account does not predict the forward progression of time for sequences in the simple past such as example (403a), since it will predict that the event time associated with the second clause is the same as the event time of the first.

(405) a. Max slipped ($E_1$). He spilt ($E_2$, $R_2$) a bucket of water.
b. $R_2 =_t E_1 \ \wedge \ E_2 =_t R_2 \ \Rightarrow \ E_2 =_t E_1$

Hinrichs (1986) proposes an account that is motivated by examples like (403a). He says:

(406) It turns out that if two sentences in the past tense *both* contain events that can be identified as either an accomplishment or an achievement, then the events are understood as happening in succession. (p. 68)

In Hinrichs's system, the anaphoric function of the past tense results from the fact that a new event is always linked to a currently operative reference point in the discourse. Furthermore, accomplishments and achievements introduce a new reference point that is temporally ordered after, and ultimately replaces, the current reference point. Thus, a subsequent event evoked from a past tense clause will be linked to this new reference point, "ensuring that two consecutive accomplishments or achievements in a discourse are always ordered in a temporal sequence." (Note that, unlike our utilization of the reference time in Reichenbach's system, the reference point in Hinrichs's system operates more as a referent than an anaphor.) This process for example (403a) results in the relations shown in (407).

(407) $E_1 <_t R_1 \ \wedge \ E_2 =_t R_1 \ \Rightarrow \ E_1 <_t E_2$

Hinrichs does not discuss the past perfect, so it is unclear how the introduction of a new reference point from accomplishments and achievements will affect the predictions for examples such as (403b). For my purposes here, however, I will give the analysis the benefit of the doubt and assume that the predictions of the standard Reichenbachian approach apply.

### 7.1.2 A Coherence-Driven Approach

In contrast to tense-as-anaphora analyses, L&A (1993) present an account in which no tenses are treated as anaphoric. Instead, temporal relations between events are recovered purely as a by-product of establishing coherence relations. The simple past and past perfect tenses do not differ in their referential properties, but instead only with respect to their "discourse roles".

Because the simple past tense is not anaphoric in L&A's account, the tenses in examples (403a-b) order the events only with respect to the speech time, and thus not with respect to each other. The temporal ordering between the events in passage (403a) instead follows directly from inferring a Narration relation (i.e., what I have called an Occasion relation) between the two clauses. They provide the two rules for Narration given below; the first being a default rule (which is therefore defeasible), and the second an indefeasible axiom.

**Narration** If the clause $\beta$ currently being processed is to be attached by a discourse relation to the clause $\alpha$ that is part of the text processed so far, then normally, *Narration*$(\alpha, \beta)$ holds.

**Axiom on Narration** If *Narration*$(\alpha, \beta)$ holds, and $\alpha$ and $\beta$ describe the eventualities $e_1$ and $e_2$ respectively, then $e_1$ occurs before $e_2$

In the case of example (403a), a Narration relation is inferred between the first and second sentences in the absence of any information to the contrary. As a result, the events are understood as occurring in the order in which they were introduced into the discourse.

The situation becomes more complicated in their account of the past perfect. The past perfect is treated as sententially equivalent to the simple past, and thus the difference between the two forms is accounted for by other means. First, they postulate the *Connections When Changing Tense* Law given in (408).

(408) **Connections When Changing Tense** (CCT)
$\Box(\langle \tau, \alpha, \beta \rangle \wedge sp(\alpha) \wedge pp(\beta) \rightarrow C_{pp}(\alpha, \beta))$

Here, $\langle \tau, \alpha, \beta \rangle$ means that some discourse relation holds between propositions $\alpha$ and $\beta$, $sp(\alpha)$ means that $\alpha$ is in the simple past, $pp(\beta)$ means that $\beta$ is in the past perfect, and $C_{pp}(\alpha, \beta)$ means "that $\alpha$ and $\beta$ are connected by the kind of discourse relation allowed between simple pasts and pluperfects"; in other words, those relations compatible with the backward movement of time. This last part is summarized in the rule given in (409).

(409) $C_{pp}(\alpha, \beta) \rightarrow \text{Elaboration}(\alpha, \beta) \vee \text{Explanation}(\beta, \alpha) \vee \text{Parallel}(\alpha, \beta) \vee \text{Contrast}(\alpha, \beta)$

For the particular example given in passage (403b), L&A also introduce a *Slipping Law*, shown in (410).[26]

(410) **Slipping Law**
$\langle \tau, \alpha, \beta \rangle \wedge C_{pp}(\alpha, \beta) \wedge \text{Info}(\alpha, \beta) > \text{Explanation}(\beta, \alpha)$

[26]The $>$ operator is used for default implication.

They gloss this law as follows:

> ... if spilling the water and slipping are connected so that either the spilling explains, elaborates, parallels or contrasts the slipping, then normally the spilling explains the slipping. (p. 472)

The predicate Info$(\alpha, \beta)$ is glossed by "$\alpha$ describes Max slipping and $\beta$ describes Max spilling a bucket of water." Note that this law makes crucial reference to the $C_{pp}$ predicate, so it only applies in cases in which the first clause is in the simple past and the second is in the past perfect.

Given these rules, interpreting passage (403b) causes the CCT Law to be satisfied, which then yields $C_{pp}(\alpha, \beta)$. This in turn causes the Slipping Law to be satisfied, which yields Explanation$(\beta, \alpha)$. Inherent in the definition of *Explanation* is the fact that the first argument[27] precedes the second since causes precede effects; therefore the correct temporal inferences result.

To summarize to this point, we have seen two quite different types of approach to identifying the temporal relations that result from interpreting tense, one that treats all tenses as anaphoric, and another that does not treat any tenses as anaphoric. Both result in accurate predictions for examples (403a-b).

### 7.1.3 Problems with the Analyses

Although both analyses ultimately derive the correct temporal relations for passages (403a-b), in each case these predictions came at the cost of requiring additions to their basic machinery that were in some sense unnatural. In particular, Hinrichs's analysis required that events introduce additional reference times into the discourse model to account for the forward movement of time in narratives, whereas L&A's approach required the inclusion of a predicate which constrains linguistic form (i.e., $C_{pp}$) in their otherwise purely semantic axioms (e.g., the *Slipping Law*) in order to get the results for the past perfect to work.

In this section, I extend the set of examples I have discussed thus far to include passages that prove to be problematic for these aspects of the proposals. The examples are shown in (411a-d) and (412a-d); passages (411a) and (412a) are repeated from passage (403a), and passage (411b) is repeated from passage (403b). Each example is shown with the coherence relation that is operative under the most natural interpretation.

[27]Note that in L&A's system, the representation of the second clause, $\beta$, is the first argument to the Explanation relation and the representation of the first clause, $\alpha$, is the second.

(411) a. Max slipped. He spilt a bucket of water. (Occasion)
b. Max slipped. He had spilt a bucket of water. (Explanation)
c. Max slipped because he spilt a bucket of water. (Explanation)
d. Max slipped because he had spilt a bucket of water. (Explanation)

(412) a. Max slipped. He spilt a bucket of water. (Occasion)
b. Max spilt a bucket of water. He tripped on his shoelace. (Explanation)
c. Max spilt a bucket of water. He spilt it all over the rug. (Elaboration)
d. Max spilt a bucket of water. John dropped a jar of cookies. (Parallel)

Hinrichs's treatment runs into problems on both sets of examples. With respect to (411a-d), his explanation of the forward progression of time in example (411a) fails to explain why sentence (411c) is as felicitous as sentence (411d). In particular, one would instead expect a clash of temporal relations for sentence (411c), because the simple pasts will induce the forward progression of time whereas the inferred Explanation relation, in this case marked explicitly with *because*, implies the reverse temporal ordering. One would therefore expect that only sentence (411d) is felicitous, since in this case the temporal relations induced by the tenses agree with those required by the coherence relation. In reality, sentences (411c) and (411d) are both felicitous, having essentially the same reading.

Passages (412a-d) are also problematic for Hinrichs's account because the generalization upon which the account is based – cited in (406) – is incorrect. In reality, any temporal relation can hold between accomplishments and achievements described by adjacent clauses in the simple past. While his account correctly predicts the forward movement of time between the events in (412a), it also makes the same prediction for examples (412b-d). In actuality, passage (412b) has a reading in which the reverse temporal order holds, in passage (412c) the event times are inferred to be the same, and no order is necessarily implied in (412d).

Passage (411c) is also problematic for L&A's approach. Recall from Section 2.3 that while connectives like *because* may constrain the set of possible coherence relations that can be realized, they do not in and of themselves create coherence. It is still necessary that the coherence relation be established with respect to one's knowledge and beliefs about the world. For instance, passage (413) is incoherent because a typical hearer

will presumably have no knowledge capable of establishing a causal relationship between someone slipping and the capital of Wisconsin, in the same way that the Slipping Law does for slipping and spilling.

(413) # Max slipped because Madison is the capital of Wisconsin.

As is the case for passage (411b), therefore, the establishment of Explanation in (411c) must utilize the causal knowledge expressed in the Slipping Law. However, in their attempt to account for (411b) without treating the past perfect as anaphoric, recall that L&A formulated this law to *require* that the second clause be in the past perfect, using the $C_{pp}$ predicate. This law therefore cannot be used to establish the coherence of passage (411c), in which the second clause is in the simple past. Thus, without another ad-hoc insertion of linguistic form constraints into the axiom (such as disjoining $C_{pp}$ with a predicate meaning "the word *because* conjoins the clauses", a move which would neither carry explanatory value nor mitigate the potential need for additional modifications for each new counterexample), the coherence of example (411c) cannot be established.

I claim that the cases involving the past perfect for which L&A's axiomatic machinery yields the correct predictions, cases not covered such as example (411c), and examples involving other compound tenses can all be handled more naturally by treating such tenses as anaphoric. On the other hand, I will argue that L&A's account provides the correct analysis of passages (412a-d), which proved problematic for Hinrichs's tense-as-anaphora approach. In the next section, I propose an analysis that captures aspects of both types of approach, and show that it accounts for these data without recourse to unmotivated rules or principles.

## 7.2 A New Account

Once again, my analysis of tense and temporal relations follows those I have presented for the other phenomena discussed in this book, in that it captures the interaction between the properties of the linguistic form in question and the properties of the process of establishing coherence. In this case, the analysis emerges from the interaction between the temporal constraints imposed by tense and those imposed by different coherence relations. I discuss these two types of constraint in turn.

### 7.2.1 Temporal Constraints Imposed by Tense

For the contribution of tense to temporal relation determination, I depart from the assumptions of Reichenbach and turn instead to an elegant and intuitive formulation due to Comrie (1981, 1985). Comrie summa-

rizes his theory with the following schema:

$$E \underset{magn}{(relative\ R)^n} \underset{magn}{(relative\ S)}$$

The predicate *relative* stands for one of the relations *before*, *after*, or *simul(taneous)*. The *magn* relation is included because some languages have different forms for expressing near-term and long-term temporal relations, such as a recent past and remote past distinction. As in Reichenbach's system, $E$, $R$, and $S$ indicate event time, reference time, and speech time, respectively; however, these concepts play different roles in Comrie's system.

I will start by removing a couple of complexities in the above formula that will not impact my discussion here. First, whereas the relation to speech time is optional in the formula because it is not used for what Comrie calls the *relative* tenses (e.g., relative past, present, and future), English tenses do express such a relationship. Second, I will ignore the *magn* relations since English does not have separate forms for near-term and long-term temporal relations. For my present discussion, I am only concerned with what Comrie calls the English *absolute* tenses (simple past, present, future) and the *absolute-relative* tenses (e.g., past perfect, future perfect), and so for these Comrie's theory can be summarized with the following formula:

$$E\ (relative\ R)^n\ relative\ S$$

As this formula suggests, there can be zero, one, or more than one reference time $R$. For the absolute tenses there is none, the event time is simply ordered with respect to the speech time. For tenses like past perfect and future perfect, there is one reference time. Comrie suggests that the conditional perfect (e.g., *the fields would have been burnt to stubble*) encodes two reference times, although I will not pursue such examples further.

The temporal relations for certain tenses are shown in Table 6, in which $<_t$ encodes Comrie's relation *before*, $>_t$ encodes *after*, and $=_t$ encodes *simul*. Note that unlike Reichenbach's system, the ordering between all times associated with a tense is not always completely specified. In particular, whereas the need to order speech and event times with respect to each other required Reichenbach to specify three different meanings for the future perfect (see Table 5), Comrie's system includes only one definition that leaves them unordered.

The use of Comrie's system in treating tense as anaphoric is relatively straightforward. An event description introduces a new event time $E$, ordered with respect to the relations indicated in Table 6. Events in-

| Tense | Relations | Example |
|---|---|---|
| Present | E $=_t$ S | I see |
| Past | E $<_t$ S | I saw |
| Future | E $>_t$ S | I will see |
| Past Perfect | E $<_t$ R $<_t$ S | I had seen |
| Future Perfect | E $<_t$ R $>_t$ S | I will have seen |

TABLE 6 Temporal Relations in Comrie's Tense System

troduced using simple tenses, such as those for passage (411a), repeated below as (414), are therefore ordered only with respect to the speech time.

(414) Max slipped. He spilt a bucket of water.

That is, we simply have the following two relations:

$$E_1 <_t S$$

$$E_2 <_t S$$

No reference times nor anaphora are involved, and thus no relationship between the two event times results from tense meaning.

On the other hand, the anaphoricity of the absolute-relative tenses results from the need to resolve the reference times $R_i$ to contextually available times, on par with the manner in which the reference time was utilized in our previous discussion of Reichenbach's system. These times therefore impose constraints through their relations among each other and the speech and event times. The correct relations for passage (411b) are recovered in the manner previously shown for Reichenbach's system, since Reichenbach's and Comrie's relations for the past perfect are identical. To reiterate, consider again passage (411b), repeated below as (415).

(415) Max slipped. He had spilt a bucket of water.

As before, the first sentence introduces a new event time for the slipping, which is ordered with respect to the speech time.

$$E_1 <_t S$$

The second sentence uses the past perfect, and thus indicates the existence of an established reference time $R_2$ that is prior to the current speech time, in this case, $E_1$. The newly created event time $E_2$ is then ordered with respect to that time, as dictated by the relation specified by the past perfect.

$$E_2 <_t R_2 =_t E_1 <_t S$$

Comrie's framework is more attractive than Reichenbach's not only because it dispenses with reference times for simple tenses, but also because it captures the iterative nature of absolute-relative tenses, and in doing so has the capacity to extend to more complex tenses not covered by Reichenbach. (See also Comrie (1981) for other arguments that his account is superior to Reichenbach's.) Simply put, a tense introduces a series of times and relations between them of the following form:

$$\underbrace{E \text{ reln}}_{\text{Absolute Tense}} \quad \underbrace{R_1 \text{ reln}, \ldots, R_n \text{ reln}}_{\text{Relative Tense}} \quad S$$

The newly created event time E is related to the time immediately to its right in this series, which in the case of absolute tenses is the speech time. In the case of absolute-relative tenses, there will be a chain of one or more anaphorically-identified reference times and relations between them, which seems to accord with intuitions concerning the type of context-dependency associated with such forms. Of course, as Comrie points out, a natural language will typically grammaticalize only a small number of the potentially infinite possibilities allowed by the above schema, which is likely a result of the general lack of need for increasingly complex tenses, and the considerable computational burden that would be associated with interpreting them.

### 7.2.2 Temporal Constraints Imposed by Coherence Relations

In my analysis of passage (415), the temporal relations that hearers normally infer – in particular, that the slipping occurred before the speech time, and the spilling occurred before the slipping – were derived solely from the constraints imposed by the tenses used. In general, however, the tenses used may not completely specify the temporal relations between the described events. This is the case for my analysis of passage (414): Whereas the facts that the slipping and the spilling both occurred before the speech time were derived, hearers also infer that the spilling occurred after the slipping. Recall that this fact is what inspired Hinrichs to augment Reichenbach's system with additional reference times.

As we have seen in previous chapters, however, the process of coherence establishment can also impose constraints on interpretation. I therefore join L&A in claiming that the temporal constraints imposed by coherence relations are another source of temporal relations. While in my account the coherence relation inferred must be consistent with any temporal relations that are imposed by the tenses used (and thus these

temporal relations may constrain the set of coherence relations that can be inferred, cf. systems which treat such constraints as defeasible, e.g. Kameyama et al. (1993)), the relations imposed by coherence may also go beyond those imposed by tense. The temporal constraints associated with a selection of coherence relations that are relevant to the examples I have been discussing are listed below.

**Occasion:** The Occasion relation is characterized by a series of eventualities that are connected through a chain of final and initial states. When these eventualities are accomplishments or achievements, this implies that they display forward movement of time, such as in passage (411a). This constraint mirrors the one specified by L&A (1993) for their Narration relation.

$$\text{Occasion}(S_1, S_2) \rightarrow E_1 <_t E_2$$

**Parallel:** The Parallel relation relates utterances that share a common topic. This relation does not impose constraints on the temporal relations between the events beyond those provided by the tenses themselves. For instance, consider passage (411a) again, in this case placed in a context that supports a Parallel relation.

(416) A: What bad things happened to Max today?
B: Max slipped. He spilt a bucket of water.....

Under the intended discourse relation, a temporal ordering among the slipping and spilling events is no longer implied.

$$\text{Parallel}(S_1, S_2) \rightarrow \text{no constraint}$$

**Elaboration:** Utterances standing in the Elaboration relation each describe the same event, therefore imposing the constraint that the event times be the same.

$$\text{Elaboration}(S_1, S_2) \rightarrow E_1 =_t E_2$$

**Result:** The Result relation imposes the same temporal constraints as Occasion, since causes precede effects.

$$\text{Result}(S_1, S_2) \rightarrow E_1 <_t E_2$$

**Explanation:** The Explanation relation denotes a Result relationship with reversed clause ordering, as in sentences (411b-d). Therefore, the second event is required to precede the first.

$$\text{Explanation}(S_1, S_2) \rightarrow E_1 >_t E_2$$

To summarize, the integration between tense meaning and coherence establishment in my proposal is straightforward. First, tense meaning

imposes constraints on temporal relations per Comrie's theory. Coherence establishment may then impose additional constraints as in L&A's system, although the relations inferred must be consistent with the constraints already imposed by tense.

### 7.2.3 Predictions

I now analyze what the account predicts for examples (411a-d) and (412a-d), repeated below as (417a-d) and (418a-d).

(417) a. Max slipped. He spilt a bucket of water. (Occasion)
b. Max slipped. He had spilt a bucket of water. (Explanation)
c. Max slipped because he spilt a bucket of water. (Explanation)
d. Max slipped because he had spilt a bucket of water. (Explanation)

(418) a. Max slipped. He spilt a bucket of water. (Occasion)
b. Max spilt a bucket of water. He tripped on his shoelace. (Explanation)
c. Max spilt a bucket of water. He spilt it all over the rug. (Elaboration)
d. Max spilt a bucket of water. John dropped a jar of cookies. (Parallel)

| Example | Tense | Coherence | Result |
|---|---|---|---|
| 417a | none | $E_1 <_t E_2$ | $E_1 <_t E_2$ |
| 417b | $E_1 >_t E_2$ | $E_1 >_t E_2$ | $E_1 >_t E_2$ |
| 417c | none | $E_1 >_t E_2$ | $E_1 >_t E_2$ |
| 417d | $E_1 >_t E_2$ | $E_1 >_t E_2$ | $E_1 >_t E_2$ |

TABLE 7 Contributions of Tense and Coherence for Examples (417a-d)

The contributions of tense and coherence for (417a-d) are summarized in Table 7. The treatment of passage (417a) is analogous to L&A's; the simple past tenses leave the event times unordered, but the Occasion coherence relation imposes an ordering with respect to the forward progression of time. Thus, this ordering is obtained without the need to specify an additional reference time as in Hinrichs's system. On the other hand, as I previously illustrated, the constraints imposed by the tenses in passage (417b) order the spilling prior to the slipping. This necessitates inferring a coherence relation that is consistent with this temporal order (in this case, Explanation), accounting for the difference between the coherence relations inferred for (417a) and (417b). Thus, the correct

ordering is obtained without the need for an additional predicate governing linguistic form (such as the $C_{pp}$ predicate), nor the requirement to intermix such constraints on linguistic form within our world knowledge axioms.

Passage (417c) is similar to (417a), except that the conjunction *because* cues the Explanation relation rather than Occasion, resulting in the reverse ordering. Passage (417c) is therefore unproblematic in my system, unlike L&A, again because I have no need to constrain the world knowledge axioms required for inferring Explanation to apply only when the second clause is in the past perfect. We also see no clash of temporal relations like we did in Hinrichs's approach, since the simple tenses do not themselves impose a forward temporal ordering. Finally, in passage (417d) the tense orders the times in backward progression as in passage (417b), and the Explanation relation cued by *because* is consistent with that ordering.

The predictions of my analysis for examples (418a-d) are shown in Table 8. Since these examples all involve only simple past tenses, my analysis treats them in the same manner as L&A do, in which the coherence relations are solely responsible for ordering the events. Both analyses therefore avoid the problems with Hinrichs's approach, since these tenses do not impose any ordering between the events themselves.

| Example | Tense | Coherence | Result |
|---|---|---|---|
| 418a | none | $E_1 <_t E_2$ | $E_1 <_t E_2$ |
| 418b | none | $E_1 >_t E_2$ | $E_1 >_t E_2$ |
| 418c | none | $E_1 =_t E_2$ | $E_1 =_t E_2$ |
| 418d | none | none | none |

TABLE 8 Contributions of Tense and Coherence for Examples (418a-d)

To sum, my analysis shows how the temporal constraints imposed by a theory of tense interact with those imposed by the process of establishing coherence relations to make the correct predictions. Each component of the analysis is independently motivated and requires no additional rules or principles.

I conclude this section by pointing out that my analysis leaves open the question of why examples such as (418b) appear to be readily understood as an Explanation without being cued by either the past perfect or the connective *because*, in the face of similar examples such as (418a) which, while in principle allowing an Explanation interpretation, are typically understood as an Occasion in absence of such a cue. L&A's

analysis will actually derive this difference, as long as the $C_{pp}$ predicate is included in the Slipping Law for (418a) but omitted in whatever axiom is necessary to establish Explanation for (418b). Obviously, this approach begs the question of how one determines when the $C_{pp}$ predicate should be included in an axiom on independently-motivated (and hence non-circular) grounds. In any case, I believe the burden for explaining this difference is properly placed onto theories of coherence establishment, and thus should be kept independent of the theory of tense. One can readily find passages in which many relations are possible but one is clearly preferred, and thus the fact that passages like (418a) and (418b) might preferentially receive different interpretations, while perhaps mysterious, is nothing out of the ordinary. Inserting constraints on linguistic form in world knowledge axioms does little to shed light on the factors that are ultimately responsible for this difference, and as we have seen, such insertions typically encumber the use of such axioms in other contexts in which they are necessary.

## 7.3 Comparison with Webber's Analysis

In Section 7.1.3, I pointed out several problems with Hinrichs's tense-as-anaphora approach, showing that his mechanism for recovering the forward movement of time between events in an Occasion generates the wrong predictions for examples in which one of several other possible relations is operative. Webber (1988), on the other hand, presents a more sophisticated tense-as-anaphora account which handles a greater range of examples than does Hinrichs's approach. In this section, I compare the predictions of my analysis with hers.

In Webber's framework, a simple past tense is anaphoric, but can specify one of three times associated with a previously evoked event: the time of the event itself, the *preparatory* phase, and the *consequent* phase. (These terms come from an ontology of events proposed by Moens and Steedman (1988).) Example (419) is used to illustrate the first of these three possibilities.

(419) John played the piano. Mary played the kazoo.

Webber claims that example (419) is understood such that the two events happen at the same time, as predicted by interpreting the past tense in the second sentence to specify the event time evoked by the first sentence. She elaborates by stating "whether this is further interpreted as two simultaneous events or a single event of their playing a duet depends on context and, perhaps, world knowledge as well". However, I do not find example (419) to necessarily imply that the two playings are contemporaneous, and suggest that this is also an inference derived from

context. For instance, if these sentences are spoken in response to the question *What instruments did John and Mary each play today?*, passage (419) implies only that the playings each happened some time that day. In my framework, the past tenses in passage (419) both evoke new event times into the discourse model, constrained only to precede the speech time. The sentences are related by the Parallel relation, which imposes no further constraints on the times evoked.

Example (420) illustrates the case in which a past tense specifies the *consequent phase* of a previously evoked event.

(420) a. John went into the florist shop.
b. He picked out three red roses, two white ones and one pale pink.

Because the consequent phase of an event is ordered after the event itself, such reference induces the forward movement of time. In my framework, this fact results from understanding this text as an Occasion.

Finally, example (421) illustrates the case in which a past tense specifies the *preparatory phase* of a previously evoked event.

(421) a. John bought Mary some flowers.
b. He picked out three red roses, two white ones and one pale pink.

Here, the picking event is understood as an initial step of what, when completed, can be described as a buying event, which is captured by linking the tense of the former to the preparatory phase of the latter. That is, the second sentence begins an Elaboration of the buying event; for instance, this passage might be completed with sentence (421c).

(421) c. He walked up to the register and paid for them.

Sentences (421b) and (421c) form an Occasion, which as a unit are related to the first sentence by Elaboration. Because such examples involve coherence between discourse segments (rather than only individual sentences), they require that the temporal constraints imposed by coherence relations be extended to account for intervals. I will return to such cases in Section 7.5.3.

Given the three possibilities in Webber's analysis, it is not clear how cases like (417c) and (418b) can be handled. In these examples, the second event is understood to occur before the first, even though the simple past is used. This interpretation is not compatible with any of the three alternatives that Webber provides. It is closest to the possibility of specifying the preparatory phase of the first event, but this analysis cannot be maintained; the analysis would then fail to distinguish between the case in which the event happened strictly earlier than

the evoked event (as in examples (417c) and (418b)), and the case in which the event was part of (and thus temporally overlapping with) a more complex event that had been evoked (as in example (421)). Thus, in addition to overcommitting to a contemporaneous interpretation for example (419), Webber's analysis appears not to allow for the strictly backward movement of time in examples (417c) and (418b).

One question that remains bears on the relationship between coherence, anaphora, and event structure in the interpretation of these examples. In particular, the cases which Webber's system handles correctly necessarily rely on a mechanism for choosing the correct referent among the three possibilities. I would argue that it is in fact the semantic relationships between the sentences that hearers infer – the coherence relationships – that determine the correct choice. For instance, the fact that example (420) involves reference to the consequent phase in her analysis, whereas example (419) involves reference to the event itself, results from theory-external factors responsible for inferring Occasion for (420) and Parallel for (419). As we have seen, the correct temporal relations are recoverable solely from these relations even if the simple past is treated as nonanaphoric, and thus the constraints imposed by Webber's analysis are unnecessary.

On the other hand, Webber (p.c.) points out that her approach and the one presented here are perhaps more closely related than might initially be apparent. This is due to the fact that the inference processes underlying the establishment of different coherence relations appear to focus in on different aspects of the structure of the events being related, in much the same way that the anaphoric property of tense in her analysis does. For instance, a clause which elaborates a culminated process (i.e., an accomplishment) naturally ties into the preparatory part of its event nucleus, whereas a clause related by Occasion will naturally tie into the consequent state of the previous event. Therefore, while a (partially) coherence-driven approach may eliminate the need to treat simple tenses as anaphoric, the role of event structure in establishing coherence – and as a result, in the inference of temporal relations in discourse – remains of central importance, as Webber claims.

## 7.4 Comparison with Coherence-Driven Approaches

In L&A's account, recall the simple past and past perfect tenses are both treated as non-anaphoric, in contrast to tense-as-anaphora approaches which treat them both as anaphoric. Temporal relations are instead recovered purely as a by-product of coherence establishment. While my account agrees with their treatment of the simple past, it disagrees with

their treatment of the past perfect. Here, I provide a more detailed discussion of the differences between the two approaches.

In arguing that the past perfect should not be treated as anaphoric, L&A note the incoherence of example (422).

(422) ? Max poured a cup of coffee. He had entered the room.

They state:

> Theories that analyse the distinction between the simple past and pluperfect purely in terms of different relations between reference times and event times, rather than in terms of event-connections, fail to explain why [(417b)] is acceptable but [(422)] is awkward. (p. 470)

Example (422) indeed shows that coherence relations need to be utilized to account for temporal relations, but it does not bear on the issue of whether the past perfect is anaphoric. The incoherence of example (422) is predicted by both their and my accounts by virtue of the fact that there is no coherence relation that corresponds to Occasion with reverse temporal ordering, for instance, in a manner analogous to how Explanation corresponds to Cause-Effect with reverse temporal ordering. Recall that L&A specify a special rule (the *Connections When Changing Tense (CCT)* Law) that stipulates that a sentence containing the simple past followed by a sentence containing the past perfect can be related only by a subset of the otherwise possible coherence relations, which were shown in axiom (409). However, this subset contains just those relations that are predicted to be possible in my account by treating the past perfect as anaphoric; they are the ones that do not constrain the temporal order of the events against displaying backward progression of time. Therefore, I see no advantages to adopting such a rule over treating the past perfect as anaphoric. Furthermore, they do not comment on what other laws have to be stipulated to account for the facts concerning the wide variety of other possible tense combinations.

Second, to explain why the Explanation relation can be inferred for passage (417b) but not for passage (417a), L&A stipulate that their causal *Slipping Law* (stating that spilling can cause slipping) requires that the CCT Law be satisfied. As I have indicated, this constraint is imposed only to require that the second clause contain the past perfect instead of the simple past, for instance, to differentiate passages (417a) and (417b). Of course, by treating the past perfect as anaphoric, I had no need for such a stipulation. Furthermore, this stipulation contradicts the fact that the simple past is perfectly coherent when the Explanation relationship is cued overtly with the inclusion of the word *because*, as it

is in sentence (417c). (As previously noted, in their framework discourse relations should be inferable whether or not they are cued by an overt conjunction.) Likewise, as I alluded to in Section 7.2.3, they do not adequately explain why CCT must be satisfied for this causal law and not for those supporting similar examples for which they successfully infer an unsignalled Explanation relation, such as in their discussion of example (423).

(423) Max fell. John pushed him.

Third, the L&A account does not explain why the past perfect cannot stand alone nor discourses generally be opened with it. For instance, consider the oddness of stating sentence (424) in isolation.

(424) Max had spilt a bucket of water.

Intuitively, this sentence is infelicitous when used discourse-initially because of a dependency on a contextually-salient time that has not been previously introduced.[28] This fact is not captured by the L&A account because sentences containing the past perfect are treated as sententially equivalent to those containing the simple past. On the other hand, sentences in the simple past are perfectly felicitous when standing alone or opening a discourse, introducing an asymmetry in accounts treating the simple past as anaphoric to a previously evoked time. All of these facts are explained by the account given here.

## 7.5 Extensions to the Analysis

As I indicated in the introduction, my analysis of tense is by no means comprehensive; in many respects it merely scratches the surface of the phenomena to be explained. For instance, language offers many other ways in which to express temporal information (adverbials, temporal connectives) which I have not addressed here. Furthermore, as previously mentioned, I will ultimately have to integrate models of aspect and event structure into the analysis in a more fundamental way.

In the sections that follow, I briefly address three of the areas that my analysis leaves open for further study. Specifically, they include extending the analysis to tenses beyond the simple past and past perfect, extending it to discourses with stative predicates, and extending it to accommodate the temporal constraints that need to be inferred over events in larger discourse structures. I consider these in turn.

[28] One does find the past perfect used in this manner as a scene-setting device in certain genres. However, it is this contextual dependency that is being flouted to achieve the desired rhetorical effect. The effect is analogous to the use of a pronoun in a story before its referent has been introduced.

### 7.5.1 Application to Future Tenses

The works that I have discussed have concentrated largely on the simple past and past perfect. It is not always clear how these approaches and others in the literature would extend to other tenses. For example, it is not apparent if and how Hinrichs's rules governing temporal progression for successive past tenses would apply to future and more complex tenses. Likewise, as I have already argued, L&A's account would presumably need to be augmented with additional rules that are analogous to those which they specifically designed to apply to the simple past and past perfect.

The relatively transparent manner in which my analysis integrates Comrie's theory of tense with the constraints imposed by coherence relations allows us to readily identify a broad set of predictions it makes about the temporal relations resulting from other possible tense combinations. While I have yet to fully examine the many possibilities available, I will make an initial foray into investigating the simple future and future perfect tenses. Examples (425a-d) are similar to (417a-d) except that they employ these tenses.

(425) a. Max will slip. He will spill a bucket of water.
b. Max will slip. He will have spilt a bucket of water.
c. Max will slip, because he will spill a bucket of water.
d. Max will slip, because he will have spilt a bucket of water.

Both clauses in passage (425a) are in the simple future, whereas in passage (425b) the first clause is in the simple future and the second is in the future perfect. Passages (425c) and (425d) are analogous to passages (425a) and (425b) except that the coherence relation has been explicitly cued with *because*.

Of course, under normal communicative circumstances a sequence of events is not described using the future tense. There are several reasons for this, not the least of which is that normally people do not know what events are going to transpire in the future. To assess judgments for these passages, it might be helpful to consider a scenario in which they would be natural, such as one in which a conversational participant is explaining to another what is going to transpire with respect to a prescripted event, such as a play or professional wrestling match.

The predictions of the analysis are summarized in Table 9. With respect to the relative ordering of the event times, these predictions are exactly the same as for (417a-d), which appears to be the correct result. Since the first clause of each passage in (425a-d) is in the simple future, an event time is evoked that is ordered subsequent to the speech time per Comrie's theory.

| Example | Tense | Coherence | Result |
|---|---|---|---|
| 425a | none | $E_1 <_t E_2$ | $E_1 <_t E_2$ |
| 425b | $E_1 >_t E_2$ | $E_1 >_t E_2$ | $E_1 >_t E_2$ |
| 425c | none | $E_1 >_t E_2$ | $E_1 >_t E_2$ |
| 425d | $E_1 >_t E_2$ | $E_1 >_t E_2$ | $E_1 >_t E_2$ |

TABLE 9 Contributions of Tense and Coherence for Examples (425a-d)

$$S <_t E_1$$

The second clause in passages (425a) and (425c) is also in the simple past, generating the same relation.

$$S <_t E_2$$

On the other hand, the second clause in passages (425b) and (425d) is in the future perfect, and thus, per Comrie's theory, the reference time is equated with a contextually available temporal referent ($E_1$) and the event time is ordered prior to it. $E_1$ remains ordered subsequent to the speech time as above.

$$S <_t E_1; \quad E_2 <_t R_2 = E_1$$

Having established the relations imposed by tense, next comes the effects of coherence establishment. As with (417a), a natural interpretation of passage (425a) is Occasion (at least assuming the type of context I described), which orders the times in forward sequence. Of course, it is possible to infer other relations for (425a), such as Explanation or Parallel, depending on the context. All of these possibilities are perfectly consistent with the fact that the simple future tenses do not order the event times with respect to each other. As I indicated at the end of Section 7.2.3, identifying which of several possible relations gets inferred in a given contextual circumstance is the job of a suitable theory of coherence establishment, and not of a theory of tense.

The remainder of the cases are all analogous to their counterparts in examples (417b-d). Specifically, the future perfect in (425b) orders the event times in reverse temporal progression, requiring a coherence relation that is compatible with that ordering (in this case, Explanation). The simple futures in (425c) do not impose an ordering, which allows the establishment of Explanation, as cued by *because*, to put them in reverse temporal order. Finally, in (425d), the future perfect and connective both signal reverse temporal order.

There remains one loose end that requires explanation, however,

which concerns the fact that the spilling events in passages (425b) and (425d), as in passages (425a) and (425c), are both typically understood to occur in the future, that is, subsequent to the speech time. While this relation is generated for passages (425a) and (425c), Comrie's meaning of the future perfect does not, strictly speaking, capture this fact for passages (425b) and (425d). In particular, the requirements that the event time precede the reference time and the speech time precede the reference time do not imply an ordering between the event time and the speech time.[29]

Comrie addresses this issue, arguing convincingly that this relationship is actually the result of an implicature and as such should not be considered as part of the meaning of the future perfect. Consider example (426).

(426) John will have finished his manuscript by tomorrow.

Sentence (426) can be felicitously used in a circumstance in which the described event has already taken place, most notably when the speaker is unaware of this fact. The feeling that this statement is odd if the speaker knows that the event has occurred is not a result of the sentence not being true, but instead due to a violation of Grice's Maxim of Quantity, *Make your contribution as informative as is required.* The maxim therefore dictates that the simple past should be used in this situation. The claim that this relationship results from an implicature suggests that it should be cancelable, which appears to be the case. In particular, a hearer could respond to (426) by saying *Yes, in fact, he has already finished it* without expressing a contradiction.

### 7.5.2 Stative Predicates

As I discussed in Section 7.1.1, Hinrichs's theory predicts forward movement of time only for sequences of accomplishments and achievements. The situation for statives is different; he continues the passage I previously cited in (406) with the following:

> If, however, one of the two sentences contains an activity or state, then the events can be viewed as either happening in succession or as overlapping each other in time. If *both* sentences contain activities or states, then the events overlap each other. (p. 68)

[29]Recall from Table 5 that this is also true in Reichenbach's system, in which the future perfect is associated with three different combinations of event, reference, and speech time, corresponding to the three possible ways of ordering the speech and event times.

He allows for two possibilities when only one sentence describes an activity or state due to the existence of examples like (427) and (428).

(427) He went the the window and pulled aside the soft drapes. It was a casement window and both panels were cranked out to let in the night air. The apartment was on the second floor. The window itself was a scant five feet above the roof.

(428) Jameson entered the room, shut the door carefully, and switched off the light. It was pitch dark around him, because the Venetian blinds were closed.

The first two events in passage (427) are most naturally understood to occur in temporal succession, but the states described by the remaining sentences are understood to overlap temporally with those events. This is not always the case, however, as demonstrated by passage (428). Although the fourth clause of (428) is stative, the state it describes is most naturally understood as occurring after the sequence of events described in the first three clauses. Hinrichs handles these cases by proposing that the current reference time is included in the temporal interval of a state. Since the updated reference time introduced by an accomplishment or achievement is ordered after the time of the event itself, this effectively leaves open the question of whether the temporal interval associated with the state overlaps with the event, or is ordered strictly after it.

However, the attentive reader may have noticed that there is another difference between passages (427) and (428), particularly with respect to the coherence relations that are operative between the clauses in question. In particular, whereas an Occasion relation is operative in passage (427), the third and fourth clauses of passage (428) participate in a Result relation. Thus, a natural treatment in my framework would simply specify that states are associated with temporal intervals in the same way that I have (perhaps oversimplistically) associated temporal points with events, and leave it up to the coherence establishment process (which, as always, will rely heavily on world knowledge) to determine the relationship between this interval and prior eventualities.

This analysis requires that I extend the temporal constraints imposed by coherence relations from temporal points to intervals. The constraints imposed by the Result and Occasion relations, while essentially the same with respect to temporal points, differ when scaled to intervals. The constraints on the Result relation, being based in cause and effect, do not change when intervals are involved: The duration of a state that is strictly an effect of some other cause must necessarily begin subsequently to the occurrence of that cause. In contrast, the constraints underlying Occasion are not based so much on natural laws, but

instead on the cognitive principles underlying a speaker's attempt to assist a hearer in creating a mental model of the situation being described. Since such descriptions will often require that scenes be set, foreground events be explained against backgrounds, and so forth, temporal overlap between states and events is consistent with the basis for Occasion. This distinction thus shows how Hinrichs's two cases are predicted by a coherence analysis.

However, given that Occasion is consistent with temporal overlap and Result is consistent with forward temporal progression, my analysis would predict the existence of another possibility not mentioned by Hinrichs. In particular, one would expect that Explanation, which imposes the reverse temporal ordering of Result, would be consistent with backward temporal progression when statives are involved. This prediction is in fact borne out; consider (429).

(429) Jameson raised the blind. It was too dark to read.

Passage (429) has a reading in which the state of darkness is understood to precede the raising event (but not extend beyond it, cf. example (427)), since the darkness can be inferred to be what caused Jameson to carry out that action. This relationship is not permitted in Hinrichs's analysis because the darkness does not overlap with the reference point introduced by the raising event (which, again, is ordered after that event). The current analysis not only handles this case, but in some sense predicts its existence from first principles.

### 7.5.3 Discourse Structure

Up to this point, I have focused almost exclusively on examples containing clauses that are related directly by a coherence relation (example (421) was an exception). Consecutive clauses in a discourse will not always participate in such a relation, however, since discourse structure is hierarchical. Coherence relations apply not only to pairs of clauses but more generally to larger discourse segments, and thus my analysis will ultimately have to be extended to account for constraints on temporal relations at that level. I will not pursue such an extension here, but merely acknowledge and briefly discuss some relevant properties of several examples previously cited in the literature.

Example (430), from Webber (1987), shows how the use of the past perfect may signal the start of an embedded discourse segment, rather than relate directly to the preceding sentence.

(430) a. John went over to Mary's house.
b. On the way, he had stopped by the flower shop for some roses.
c. He picked out 5 red ones, 3 white ones, and one pale pink.

Sentences (430b-c) form an Occasion which, as a unit, comprises an Elaboration on the event described in sentence (430a). Whereas Occasion imposes a forward ordering between the events described in sentences (430b-c), Elaboration requires that the series of events described in (430b-c) taken together be contemporaneous with the event described in (430a). As we are now dealing with temporal intervals over which events take place, this example shows that the constraint on temporal simultaneity in Elaborations is both too simplistic and too strong, and must be generalized to one which enforces temporal inclusion. Passage (431), an example cited by Dowty (1986) that he attributes to Kamp, shows that a discourse can have this same internal structure even when only simple past tenses are used; this interpretation is assisted by the use of several temporal adverbials.

(431) Pedro dined at Madame Gilbert's. First there was an hors d'oeuvre. Then the fish. After that the butler brought a glazed chicken. The repast ended with a flaming dessert.

Of course, other relations are possible within segments that elaborate other segments. Passage (432), also discussed by Dowty (1986), is similar to the previous examples except that the final three sentences are related by Parallel.

(432) At the signal, everyone went to work at once. Mary searched the room for any of the items on the list that might be there. John went next door to do the same in Bill's apartment. Susan organized the rest of the players to canvass the block.

The Parallel relation does not impose any temporal constraints between the event times for these three sentences; the fact that these events are understood to have occurred at roughly the same time is due to the implication made to this effect in the first sentence. As these examples show, there are many factors that can contribute to the temporal relations one infers from a discourse, and plenty of work remains to be done to determine what they are and how they interact.

## 7.6 Conclusion

To conclude, I have considered two types of approach to tense interpretation: *tense-as-anaphora* and *coherence-driven* approaches. Each was demonstrated to have problems, using Hinrichs's (1986) and L&A's (1993) analyses as exemplars. These problems resulted from stipulations motivated solely by the need to account for different types of data with a single uniform mechanism.

I also provided and argued for an account that combines aspects

of both types of approach. In this analysis, tense meaning gives rise to certain temporal relations between eventualities, but may not in itself account for all such relations. These relations may in turn be further refined by independently-motivated temporal constraints imposed by coherence relations. The account covers the problematic data while avoiding the need to posit additional unmotivated machinery.

As Maslow once again might have told us, one therefore needs to avoid the temptation to force a solution under the assumption that it must operate within the confines of a single module of language processing. In fact, with respect to theories of tense, Comrie also warns us against this very temptation:

> ...the failure to distinguish between meaning and implicature is one of the main problems in working our an adequate characterisation of tenses...the investigation of the use of a grammatical category in discourse should not be confused with the meaning of that category; instead, the discourse functions should ultimately be accounted for in terms of the interaction of meaning and context. (Comrie 1985, pp. 28-29)

By combining a simpler account of the properties of the linguistic phenomenon at hand with an analysis of how these properties interact with the inference processes underlying coherence establishment, the desired predictions can be seen to result from independently motivated principles.

# 8

# Conclusion

Despite the breadth of ground that I have sought to cover in this book, my main purpose has centered on one overarching goal: to convince the reader that, contrary to the prevailing wisdom, methods for examining the effect of inferentially-based, knowledge-intensive coherence establishment mechanisms are necessary tools to bring to linguistic theory. This is particularly, but perhaps not only, true for the study of those phenomena that operate interclausally. I have argued that previous work on five different linguistic phenomena has been stymied by an implicit assumption that the data can be explained with a uniform set of tools, when in fact this data appears to strongly defy that assumption. As Abraham Maslow might have warned us, in such a situation we should resist the temptation to force our tools beyond their limits, and instead step back and consider finding some new ones that might be more appropriate to the task. I hope to have convinced the reader that one such set of tools is provided by the theoretical concepts used to model the manner in which people establish coherence in discourse.

Of course, I also hope the reader finds merit in the details of my particular theory of discourse coherence, as well as in my linguistic analyses that utilize it. For each of the phenomena addressed, I have shown how data that is problematic for previous approaches can be explained by a cross-modular theory that interfaces a relatively straightforward account of the properties of the linguistic phenomenon in question with the effect of coherence establishment, taking care to demonstrate the independent motivation underlying each component. Indeed, I believe I was able to supersede previous studies in precisely the areas in which those works appear to have stretched their tools beyond the aspects of the problem for which they are best suited.

That coherence establishment would play a crucial role in accounting for this data should not be surprising. As I described in Chapter 2,

there is overwhelming evidence that coherence establishment is central to our language understanding capacity, so one might naturally expect these processes to interact with other aspects of language interpretation. Indeed, it would perhaps have been surprising if it turned out that they did *not* affect the distribution and behavior of these phenomena. Yet, linguistic analyses that have employed coherence establishment mechanisms in a serious manner are few and far between (Hobbs's (1979) account of pronoun interpretation and Lascarides and Asher's (1993) analysis of tense interpretation being two such examples), and analyses that use them in conjunction with (and not instead of) certain constraints imposed by the linguistic phenomena themselves are, to my knowledge, almost nonexistent.

Having said that, I should acknowledge that in many respects this work opens up as many questions as it answers; indeed, I believe that I have only scratched the surface with respect to the interaction between coherence, reference, and the theory of grammar. One reason for this is that the study of coherence establishment is itself an area that is still in its infancy. While I would maintain that the current state of research provides a useful working simplification, our ability to examine the interaction between coherence and other aspects of linguistic processing will ultimately require the development of more realistic theories of the former. Here, I will briefly describe three ways in which I believe this to be the case.

First, more psychologically plausible models of coherence establishment are needed. The large majority of existing work on coherence relations makes no reference at all to the processing mechanisms by which these relations are established. A primary exception is the proposal of Jerry Hobbs and his colleagues (Hobbs et al. 1993), who, as I mentioned in Section 2.3, have articulated a mechanism for applying relation constraints based on the principle of logical abduction. This model assumes that the input to the inference procedure consists of fully analyzed semantic representations for sentences. While this is a worthwhile simplification (one that I used in most of the analyses herein, with the exception of the account of pronoun resolution described in Chapter 6), the situation is considerably more complicated than this. First, there is both intuitive and experimental evidence that interpretation at all levels of processing (e.g., syntax, semantics, discourse) occurs in a left-to-right, word-by-word fashion. Second, the model has no explicit representation of attentional state suitable for modeling information coming in and out of focus. Both of these properties might affect the predictions of a linguistic analysis that relies on such a model; this was in fact the case for my analysis of pronominal reference. For lack of a better analogy, one

might view the Hobbs et al. framework as a 'competence' model that embodies the simplifying assumption that processing is performed on a sentence-per-sentence basis with unlimited resources, whereas what we may eventually need is more the analog of a 'performance' model, in which processing is modeled word-by-word with limited computational resources.

Second, the Hobbs et al. model does not explain a basic fact about natural languages, which is that they afford speakers many different ways with which to express the same idea. For instance, as I have already noted in several places, speakers typically have a broad selection of syntactic forms that they can employ to express a proposition; in English the choices include active voice, passive voice, several types of cleft, topicalization, preposing, and inversion, among others. It is well known that these choices are not arbitrary; each is associated with different sets of constraints on its use in context as well as different effects on the resulting discourse state. If all that mattered for language interpretation was the logical form computed from the syntactic structure – a level of representation at which syntactic distinctions are presumably lost – we would be left with no explanation for why we have this variety of choices nor these differing sets of properties. Likewise, there are information-structuring tendencies in discourse that have been widely documented, such as that old information tends to occur earlier in a sentence than new information. While only a tendency, this pattern suggests that comprehension is assisted if the hearer is able to tie the current sentence to the previous context earlier rather than later. Again, a sentence-by-sentence model does not embody this tendency in any direct way.

Finally, it may well turn out that the manner in which I and others have described coherence – in terms of relations that can hold between clauses and larger discourse segments – is simply too coarse-grained. In any reasonably complex naturally-occurring discourse, one will find many relations that can be inferred among entities and eventualities that reflect parallelisms, contrasts, elaborations, causalities, contiguities, and so forth. The inferences driven by the need to establish interclausal coherence may be but a subset of a larger set of inferences that take place during discourse comprehension of which our theories need to take into account. We have seen how the judgments for various examples can be both fragile and gradient, for which small alterations can have large effects on interpretability. Such effects appear to be recalcitrant only because of our lack of understanding of the relevant factors at a fine enough level of detail.

Of course, what I am suggesting here would be the basis for a very large research program, and I by no means intend to cast aspersions on

previous authors for not having produced such a theory. (The reader will note that I have not produced one either.) My hope instead is that the analyses offered in this book will be seen as having enough potential to spur new interest in this relatively pristine area of investigation. One could foresee a stream of research develop in which the results of applying coherence establishment concepts within theories of linguistic phenomena shed new light back onto theories of coherence, cycling in a manner such that each iteration pins down the issues for both at a finer-grained level of detail than the last. In this vein, I will conclude this book as I started it, with a quote from *The Psychology of Science*; in this case, the quote is from the preface authored by Arthur G. Wirth:

> [Maslow] returns to an insistence that students of Dewey will recognize as a recurrent theme of that thinker: reliable insights into the world require a never-ending interplay of theoretical abstraction and the stuff of experience. We settle needlessly for partial knowledge when we opt for one without the check of the other.

# References


Ariel, Mira. 1990. *Accessing Noun Phrase Antecedents.* Routledge.

Asher, Nicholas. 1993. *Reference to Abstract Objects in Discourse.* SLAP 50, Dordrecht, Kluwer.

Brennan, Susan E., Marilyn W. Friedman, and Carl J. Pollard. 1987. A Centering Approach to Pronouns. In *Proceedings of the 25th Meeting of the Association for Computational Linguistics*, 155–162. Stanford, CA.

Chafe, Wallace L. 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View. In *Subject and Topic*, ed. Charles N. Li. 25–55. New York: Academic Press.

Chambers, Craig C., and Ron Smyth. 1998. Structural Parallelism and Discourse Coherence: A Test of Centering Theory. *Journal of Memory and Language* 39:593–608.

Chao, Wynn. 1987. *On Ellipsis.* Doctoral dissertation, University of Massachusetts at Amherst.

Chomsky, Noam. 1981. *Lectures in Government and Binding.* Foris, Dordrecht.

Comrie, Bernard. 1981. On Reichenbach's Approach to Tense. In *Papers from the Seventeenth Annual Regional Meeting of the Chicago Linguistic Society.* 24–30. Chicago: Chicago Linguistic Society.

Comrie, Bernard. 1985. *Tense.* Cambridge University Press.

Crawley, R. A., R. J. Stevenson, and D. Kleinman. 1990. The Use of Heuristic Strategies in the Interpretation of Pronouns. *Journal of Psycholinguistic Research* 19:245–264.

Dahl, Östen. 1974. How to Open a Sentence: Abstraction in Natural Language. Technical report. University of Göteborg. Logical Grammar Reports, No. 12.

Dalrymple, Mary. 1991. Against Reconstruction in Ellipsis. Technical Report SSL-91-114. Xerox.

Dalrymple, Mary, and Andrew Kehler. 1995. On the Constraints Imposed by *Respectively. Linguistic Inquiry (Squibs and Discussion)* 26(3):531–536.

Dalrymple, Mary, Stuart M. Shieber, and Fernando Pereira. 1991. Ellipsis and Higher-Order Unification. *Linguistics and Philosophy* 14:399–452.

Deane, Paul. 1991. Limits to Attention: A Cognitive Theory of Island Phenomena. *Cognitive Linguistics* 2(1):1–63.

Dowty, David. 1986. The Effects of Aspectual Class on the Temporal Structure of Discourse: Semantics or Pragmatics? *Linguistics and Philosophy* 9:37–61.

Dretske, Fred I. 1972. Contrastive Statements. *The Philosophical Review* 81(4):411–437.

Erteschik-Shir, Nomi, and Shalom Lappin. 1979. Dominance and the Functional Explanation of Island Phenomena. *Theoretical Linguistics* 6:41–86.

Evans, Gareth. 1980. Pronouns. *Linguistic Inquiry* 11:337–362.

Fiengo, Robert, and Robert May. 1994. *Indices and Identity.* Cambridge, MA: MIT Press. Linguistic Inquiry Monograph Twenty-Four.

Francis, W. Nelson. 1964. A Standard Sample of Present-Day English for Use with Digital Computers. Report to the U.S. Office of Education on Cooperative Research. Project No. E-007, Brown University, Providence, RI.

Gardent, Claire. 1997. Sloppy Identity. In *Logical Aspects of Computational Linguistics*, ed. Christian Retoré. 188–207. Springer Verlag. Lecture Notes in Computer Science 1328.

Gawron, Mark, and Andrew Kehler. 2000. Respective Readings and Gaps. Presented at the Annual Meeting of the Linguistic Society of America, Chicago, January.

Gawron, Mark, and Stanley Peters. 1990. *Anaphora and Quantification in Situation Semantics.* CSLI/University of Chicago Press, Stanford University. CSLI Lecture Notes, Number 19.

Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum, and Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar.* Cambridge, MA: Harvard University Press.

Goldsmith, John. 1985. A Principled Exception to the Coordinate Structure Constraint. In *Papers from the Twenty-First Annual Regional Meeting of the Chicago Linguistic Society.* 133–143. Chicago: Chicago Linguistic Society.

Goodall, Grant. 1987. *Parallel Structures in Syntax.* Cambridge: Cambridge University Press.

Gordon, Peter C., Barbara J. Grosz, and Laura A. Gilliom. 1993. Pronouns, Names, and the Centering of Attention in Discourse. *Cognitive Science* 17(3):311–347.

Grosu, Alexander. 1973. On the Nonunitary Nature of the Coordinate Structure Constraint. *Linguistic Inquiry* 4:88–92.

Grosz, Barbara J. 1977. The Representation and Use of Focus in a System for Understanding Dialogs. In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, 67–76. Cambridge, MA.

Grosz, Barbara J., Karen Sparck Jones, and Bonnie Lynn Webber (ed.). 1986. *Readings in Natural Language Processing.* Morgan Kaufmann Publishers, Inc.

Grosz, Barbara J., Aravind K. Joshi, and Scott Weinstein. 1995. Centering: A Framework for Modelling the Local Coherence of Discourse. *Computational Linguistics* 21(2):203–225.

Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69(2):274–307.

Haïk, Isabelle. 1987. Bound Variables that Need to be. *Linguistics and Philosophy* 10:503–530.

Halliday, M.A.K., and Ruqaiya Hasan. 1976. *Cohesion in English.* London: Longman's. English Language Series, Title No. 9.

Hankamer, Jorge. 1971. *Constraints on Deletion in Syntax.* Doctoral dissertation, Yale University.

Hankamer, Jorge. 1973. Why There are Two *than*'s in English. In *Proceedings of the Ninth Regional Meeting of the Chicago Linguistic Society*, 179–191. University of Chicago.

Hankamer, Jorge. 1978. On the Nontransformational Derivation of Some Null VP Anaphors. *Linguistic Inquiry* 9:66–74.

Hankamer, Jorge, and Ivan Sag. 1976. Deep and Surface Anaphora. *Linguistic Inquiry* 7:391–426.

Hardt, Daniel. 1990. A Corpus-Based Survey of VP Ellipsis. Unpublished Manuscript, University of Pennsylvania.

Hardt, Daniel. 1992. VP Ellipsis and Contextual Interpretation. In *Proceedings of the International Conference on Computational Linguistics (COLING-92)*, 303–309. Nantes, July.

Hardt, Daniel. 1993. *Verb Phrase Ellipsis: Form, Meaning, and Processing.* Doctoral dissertation, University of Pennsylvania.

Hardt, Daniel. 1999. Dynamic Interpretation of Verb Phrase Ellipsis. *Linguistics and Philosophy* 22(2):187–221.

Hellan, Lars. 1988. *Anaphora in Norwegian and the Theory of Grammar.* Studies in Generative Grammar 32. Foris, Dordrecht.

Hestvik, Arild. 1993. Strict Reflexives and the Subordination Effect. In *Proceedings of the Stuttgart Workshop on Ellipsis: Arbeitspapiere des Sonderforschungsbereich 340, Bericht Nr. 29-1992, SFB 340*, ed. S. Berman and A. Hestvik. University of Stuttgart, University of Tuebingen, and IBM Germany.

Hestvik, Arild. 1995. Reflexives and Ellipsis. *Natural Language Semantics* 3:211–237.

Hinrichs, Erhard. 1986. Temporal Anaphora in Discourses of English. *Linguistics and Philosophy* 9:63–82.

Hirschberg, Julia, and Gregory Ward. 1991. Accent and Bound Anaphora. *Cognitive Linguistics* 2(2):101–121.

Hobbs, Jerry R. 1979. Coherence and Coreference. *Cognitive Science* 3:67–90.

Hobbs, Jerry R. 1990. *Literature and Cognition.* CSLI Lecture Notes 21.

Hobbs, Jerry R., and Andrew Kehler. 1997. A Theory of Parallelism and the Case of VP Ellipsis. In *Proceedings of the 35th Conference of the Association for Computational Linguistics (ACL-97)*, 394–401. Madrid, Spain, July.

Hobbs, Jerry R., Mark E. Stickel, Douglas E. Appelt, and Paul Martin. 1993. Interpretation as Abduction. *Artificial Intelligence* 63:69–142.

Hovy, Eduard H. 1990. Parsimonious and Profligate Approaches to the Question of Discourse Structure Relations. In *Proceedings of the Fifth International Workshop on Natural Language Generation*, 128–136. Dawson, PA, June.

Hudson, Richard A. 1976. Conjunction Reduction, Gapping, and Right-Node Raising. *Language* 52(3):535–562.

Hume, David. 1748. *An Inquiry Concerning Human Understanding.* The Liberal Arts Press, New York, 1955 edition.

Jackendoff, Ray S. 1971. Gapping and Related Rules. *Linguistic Inquiry* 2:21–35.

Jackendoff, Ray S. 1972. *Semantic Interpretation in Generative Grammar.* MIT Press.

Jayaseelan, K. A. 1990. Incomplete VP Deletion and Gapping. *Linguistic Analysis* 20(1-2):694–81.

Johannessen, Janne Bondi. 1998. *Coordination.* Oxford Studies in Comparative Syntax. Oxford University Press.

Kameyama, Megumi. 1986. A Property-Sharing Constraint in Centering. In *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, 200–206. New York.

Kameyama, Megumi. 1996. Indefeasible Semantics and Defeasible Pragmatics. In *Quantifiers, Deduction, and Context*, ed. M. Kanazawa, C. Pi non, and H. de Swart. 111–138. Stanford, CA: CSLI.

Kameyama, Megumi. 1999. Stressed and Unstressed Pronouns: Complementary Preferences. In *Focus: Linguistic, Cognitive, and Computational Perspectives*, ed. Peter Bosch and Rob van der Sandt. 306–321. Cambridge: Cambridge University Press.

Kameyama, Megumi, Rebecca Passonneau, and Massimo Poesio. 1993. Temporal Centering. In *Proceedings of the 31st Conference of the Association for Computational Linguistics (ACL-93)*, 70–77. Columbus, Ohio, June.

Kamp, Hans. 1981. A Theory of Truth and Semantic Representation. In *Formal Methods in the Study of Language*, ed. Jeroen Groenendijk, Theo Janssen, and Martin Stokhof. 277–321. Amsterdam: Mathematical Centre.

Kamp, Hans, and Uwe Reyle. 1993. *From Discourse to Logic: An Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory.* Dordrecht: Kluwer Academic Publishers.

Kehler, Andrew. 1993a. A Discourse Copying Algorithm for Ellipsis and Anaphora Resolution. In *Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics (EACL-93)*, 203–212. Utrecht, the Netherlands, April.

Kehler, Andrew. 1993b. The Effect of Establishing Coherence in Ellipsis and Anaphora Resolution. In *Proceedings of the 31st Conference of the Association for Computational Linguistics (ACL-93)*, 62–69. Columbus, Ohio, June.

Kehler, Andrew. 1994a. Common Topics and Coherent Situations: Interpreting Ellipsis in the Context of Discourse Inference. In *Proceedings of the 32nd Conference of the Association for Computational Linguistics (ACL-94)*, 50–57. Las Cruces, New Mexico, June.

Kehler, Andrew. 1994b. A Discourse Processing Account of Gapping and Causal Implicature. Manuscript presented at the Annual Meeting of the Linguistic Society of America, Boston, January.

Kehler, Andrew. 1994c. Temporal Relations: Reference or Discourse Coherence? In *Proceedings of the 32nd Conference of the Association for Computational Linguistics (Student Session)*, 319–321. Las Cruces, New Mexico, June.

Kehler, Andrew. 1995. *Interpreting Cohesive Forms in the Context of Discourse Inference.* Doctoral dissertation, Harvard University.

Kehler, Andrew. 1996. Coherence and the Coordinate Structure Constraint. In *Proceedings of the 22nd Annual Meeting of the Berkeley Linguistics Society (BLS-22)*, 220–231. Berkeley, CA, February.

Kehler, Andrew. 1997. Current Theories of Centering for Pronoun Interpretation: A Critical Evaluation. *Computational Linguistics* 23(3):467–475.

Kehler, Andrew. 2000a. Coherence and the Resolution of Ellipsis. *Linguistics and Philosophy* 23(6):533–575.

Kehler, Andrew. 2000b. Discourse. Chapter 18 of *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, by Daniel Jurafsky and James H. Martin (contributing writers: Andrew Kehler, Keith Vander Linden, and Nigel Ward). 669-718. Prentice-Hall, New Jersey.

Kehler, Andrew. 2000c. Resolving Temporal Relations using Tense Meaning and Discourse Interpretation. In *Formalizing the Dynamics of Information*, ed. Martina Faller, Stefan Kaufmann, and Marc Pauly. 189–205. CSLI Publications, Stanford.

Kehler, Andrew, and Gregory Ward. 1995. On the Anaphoric Status of *do so*. Paper presented at the Annual Meeting of the Linguistic Society of America, New Orleans, January.

Kehler, Andrew, and Gregory Ward. 1999. On the Semantics and Pragmatics of 'Identifier So'. In *The Semantics/Pragmatics Interface from Different Points of View (Current Research in the Semantics/Pragmatics Interface Series, Volume I)*, ed. Ken Turner. 233–256. Amsterdam: Elsevier.

Kennedy, Christopher. 1997. VP Deletion and "Nonparasitic Gaps". *Linguistic Inquiry (Squibs and Discussion)* 28(4):697–707.

Kitagawa, Yoshihisa. 1991. Copying Identity. *Natural Language and Linguistic Theory* 9:497–536.

Knott, Alistair, and Robert Dale. 1994. Using Linguistic Phenomena to Motivate a Set of Coherence Relations. *Discourse Processes* 18(1):35–62.

Knott, Alistair, and Chris Mellish. 1996. A Feature-Based Account of the Relations Signalled by Sentence and Clause Connectives. *Language and Speech* 39(2-3):143–183.

Knott, Alistair, and Ted Sanders. 1998. The Classification of Coherence Relations and their Linguistic Markers: An Exploration of Two Languages. *Journal of Pragmatics* 30:135–175.

Kuno, Susumu. 1976a. Gapping: A Functional Analysis. *Linguistic Inquiry* 7:300–318.

Kuno, Susumu. 1976b. Subject, Theme, and the Speaker's Empathy – A Reexamination of the Relativization Phenomena. In *Subject and Topic*, ed. Charles N. Li. 419–444. New York: Academic Press.

Kuno, Susumu. 1987. *Functional Syntax - Anaphora, Discourse and Empathy.* Chicago and London: The University of Chicago Press.

Lakoff, George. 1968. Pronouns and Reference. In *Syntax and Semantics, Volume 7: Notes from the Linguistic Underground*, ed. J. McCawley. 275–335 (1976). New York: Academic Press. Originally distributed by Indiana University Linguistics Club, 1968.

Lakoff, George. 1986. Frame Semantic Control of the Coordinate Structure Constraint. In *CLS-22, Part 2: Papers from the Parasession on Pragmatics and Grammatical Theory at the Twenty-Second Annual Regional Meeting*, 152–167. Chicago. Chicago Linguistic Society.

Lakoff, George. 1991. Cognitive versus Generative Linguistics: How Commitments Influence Results. *Language and Communication* 11(1-2):53–62.

Lakoff, Robin. 1971. If's, And's, and But's About Conjunction. In *Studies in Linguistic Semantics*, ed. Charles J. Fillmore and D. Terence Langendoen. 114–149. New York: Holt, Rinehart, and Winston.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form.* Cambridge: Cambridge University Press.

Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar, Vol. 2: Descriptive Application.* Stanford, CA: Stanford University Press.

Lappin, Shalom. 1993a. Ellipsis Resolution at S-Structure. In *Proceedings of NELS 23*, 255–269. Amherst, MA. GLSA, University of Massachusetts.

Lappin, Shalom. 1993b. The Syntactic Basis of Ellipsis Resolution. In *Proceedings of the Stuttgart Workshop on Ellipsis: Arbeitspapiere des Sonderforschungsbereich 340, Bericht Nr. 29-1992, SFB 340*, ed. S. Berman and A. Hestvik. University of Stuttgart, University of Tuebingen, and IBM Germany.

Lappin, Shalom. 1996. The Interpretation of Ellipsis. In *Handbook of Contemporary Syntactic Theory*, ed. Shalom Lappin. 145–175. Blackwell.

Lappin, Shalom. 1999. An HPSG Account of Antecedent-Contained Ellipsis. In *Fragments: Studies in Ellipsis and Gapping*, ed. Shalom Lappin and Elabbas Benmamoun. 68–97. New York: Oxford University Press.

Lascarides, Alex, and Nicholas Asher. 1993. Temporal Interpretation, Discourse Relations, and Common Sense Entailment. *Linguistics and Philosophy* 16(5):437–493.

Lawler, John. 1974. Ample Negatives. In *Proceedings of the Tenth Regional Meeting of the Chicago Linguistic Society*, 357–377. University of Chicago.

Levin, Nancy, and Ellen Prince. 1982. Gapping and Causal Implicature. Presented at the Annual Meeting of the Linguistic Society of America.

Levin, Nancy, and Ellen Prince. 1986. Gapping and Causal Implicature. *Papers in Linguistics* 19:351–364.

Levine, Robert D. 2001. The Extraction Riddle: Just What Are We Missing? *Journal of Linguistics* 37(1):145–174. Review Article of Postal's *Three Investigations of Extraction*.

Lobeck, Anne. 1999. VP Ellipsis and the Minimalist Program: Some Speculations and Proposals. In *Fragments: Studies in Ellipsis and Gapping*, ed. Shalom Lappin and Elabbas Benmamoun. 98–123. New York: Oxford University Press.

Longacre, Robert E. 1983. *The Grammar of Discourse*. Plenum Press.

Mann, William C., and Sandra A. Thompson. 1987. Rhetorical Structure Theory: A Theory of Text Organization. Technical Report RS-87-190. Information Sciences Institute.

Martin, James R. 1992. *English Text: Systems and Structure*. Amsterdam: Benjamins.

Martin, Jim R. 1983. Conjunction: The Logic of English Text. In *Micro and Macro Connexity of Texts*, ed. János S. Petöfi and Emel Sözer. 1–72. Hamburg: Buske.

Maslow, Abharam H. 1966. *The Psychology of Science: A Reconnaissance*. New York: Harper and Row.

May, Robert. 1985. *Logical Form: Its Structure and Derivation*. Cambridge, MA: MIT Press. Linguistic Inquiry Monograph Twelve.

McKoon, Gail, Steven B. Greene, and Roger Ratcliff. 1993. Discourse Models, Pronoun Resolution, and the Implicit Causality of Verbs. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:266–283.

McWhinney, B., E. Bates, and R. Kliegl. 1984. Cue Validity and Sentence Interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior* 23:127–150.

Mittwoch, Anita. 1979. Backward Anaphora in Utterances Conjoined with *but*. Presented at the Annual Meeting of the Linguistic Society of America.

Moens, Marc, and Mark Steedman. 1988. Temporal Ontology and Temporal Reference. *Computational Linguistics* 14(2):15–28.

Na, Younghee, and Geoffrey Huck. 1992. On Extracting from Asymmetrical Structures. In *The Joy of Grammar*, ed. Diane Brentari, Gary N. Larson, and Lynn A. MacLeod. 251–274. Amsterdam: John Benjamins.

Napoli, Donna Jo. 1983. Comparative Ellipsis: A Phrase Structure Analysis. *Linguistic Inquiry* 14(4):675–694.

Neijt, Anneke. 1981. Gaps and Remnants – Sentence Grammar Aspects of Gapping. *Linguistic Analysis* 8(1):69–93.

Nerbonne, John. 1986. Reference Time and Time in Narration. *Linguistics and Philosophy* 9:83–95.

Oehrle, Richard T. 1981. Common Problems in the Theory of Anaphora and the Theory of Discourse. In *Possibilities and Limitations of Pragmatics*, ed. Herman Parret, Marina Sbisà, and Jef Verschueren. 509–530. Amsterdam: John Benjamins. Studies in Language Companion Series, Volume 7.

Partee, Barbara. 1984. Nominal and Temporal Anaphora. *Linguistics and Philosophy* 7:243–286.

Partee, Barbara. 1989. Deictic and Anaphoric Pieces of Meaning. LSA Conference on Logic and Linguistics, Santa Cruz. July 19-21.

Passonneau, Rebecca. 1988. A Computational Model of the Semantics of Tense and Aspect. *Computational Linguistics* 14(2):44–60.

Pereira, Fernando. 1990. Categorial Semantics and Scoping. *Computational Linguistics* 16(1):1–10.

Polanyi, Livia. 1988. A Formal Model of the Structure of Discourse. *Journal of Pragmatics* 12:601–638.

Pollard, Carl, and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar.* CSLI and University of Chicago Press.

Postal, Paul. 1998. *Three Investigations of Extraction.* MIT Press.

Prince, Ellen. 1986. On the Syntactic Marking of Presupposed Open Propositions. In *Papers from the Parasession on Pragmatics and Grammatical Theory at the 22nd Regional Meeting of the Chicago Linguistic Society*, 208–222. Chicago, IL.

Prüst, Hub. 1992. *On Discourse Structuring, VP Anaphora, and Gapping.* Doctoral dissertation, University of Amsterdam.

Prüst, Hub, Remko Scha, and Martin van den Berg. 1994. Discourse Grammar and Verb Phrase Anaphora. *Linguistics and Philosophy* 17:261–327.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1972. *A Grammar of Contemporary English.* London: Seminar Press.

Redeker, Gisela. 1990. Ideational and Pragmatic Markers of Discourse Structure. *Journal of Pragmatics* 14:305–319.

Reichenbach, Hans. 1947. *Elements of Symbolic Logic.* Macmillan, New York.

Reinhart, Tanya. 1982. Pragmatics and Linguistics: An Analysis of Sentence Topics. *Philosophica* 27:53–94.

Reinhart, Tanya. 1983. *Anaphora and Semantic Interpretation.* University of Chicago Press, Chicago.

Rooth, Mats. 1981. A Comparison of Three Theories of Verb Phrase Ellipsis. University of Massachusetts Occasional Papers in Linguistics, Volume 7.

Rosenbaum, Harvey. 1977. Zapotec Gapping as Counterevidence to Some Universal Proposals. *Linguistic Inquiry* 8(2):379–395.

Rosenthal, Harry F. 1988. Agency Orders UPS to Resume Fireworks Deliveries–After July 4th. Associated Press Newswire, June.

Ross, John R. 1967. *Constraints on Variables in Syntax.* Doctoral dissertation, Massachusetts Institute of Technology.

Ross, John R. 1970. Gapping and the Order of Constituents. In *Progress in Linguistics*, ed. Manfred Bierwisch and Karl Erich Heidolph. 249–259. Mouton, The Hague.

Sag, Ivan. 1976. *Deletion and Logical Form.* Doctoral dissertation, Massachusetts Institute of Technology.

Sag, Ivan, and Jorge Hankamer. 1984. Toward a Theory of Anaphoric Processing. *Linguistics and Philosophy* 7:325–345.

Samet, Jerry, and Roger Schank. 1984. Coherence and Connectivity. *Linguistics and Philosophy* 7(1):57–82.

Sanders, Ted. 1997. Semantic and Pragmatic Sources of Coherence: On the Categorization of Coherence Relations in Context. *Discourse Processes* 24:119–147.

Sanders, Ted J. M., Wilbert P. M. Spooren, and Leo G. M. Noordman. 1992. Toward a Taxonomy of Coherence Relations. *Discourse Processes* 15:1–35.

Sanders, Ted J. M., Wilbert P. M. Spooren, and Leo G. M. Noordman. 1993. Coherence Relations in a Cognitive Theory of Discourse Representation. *Cognitive Linguistics* 4(2):93–133.

Scha, Remko, and Livia Polanyi. 1988. An Augmented Context Free Grammar for Discourse. In *Proceedings of the International Conference on Computational Linguistics (COLING-88)*, 573–577. Budapest, August.

Schachter, Paul. 1977a. Constraints on Coordination. *Language* 53:86–103.

Schachter, Paul. 1977b. Does She or Doesn't She? *Linguistic Inquiry* 8(4):762–767.

Schachter, Paul. 1978. English Propredicates. *Linguistic Analysis* 4(3):187–224.

Schmerling, Susan. 1972. Apparent Counterexamples to the Coordinate Structure Constraint: A Canonical Conspiracy. *Studies in the Linguistic Sciences* 2(1):91–104.

Schmerling, Susan. 1975. Asymmetric Conjunction and Rules of Conversation. In *Syntax and Semantics, Volume 3: Speech Acts*, ed. P. Cole and J.L. Morgan. 211–231. New York: Academic Press.

Shieber, Stuart M., Fernando C. N. Pereira, and Mary Dalrymple. 1996. Interactions of Scope and Ellipsis. *Linguistics and Philosophy* 19(5):527–552.

Sidner, Candace. 1983. Focusing in the Comprehension of Definite Anaphora. In *Computational Models of Discourse*, ed. M. Brady and R. Berwick. 267–330. MIT Press.

Siegel, Muffy E. A. 1984. Gapping and Interpretation. *Linguistic Inquiry* 15:523–530.

Smyth, R. 1994. Grammatical Determinants of Ambiguous Pronoun Resolution. *Journal of Psycholinguistic Research* 23:197–229.

Steedman, Mark J. 1985. Dependency and Coordination in the Grammar of Dutch and English. *Language* 61:523–568.

Steedman, Mark J. 1990. Gapping as Constituent Coordination. *Linguistics and Philosophy* 13(2):207–263.

Stevenson, R. J., R. A. Crawley, and D. Kleinman. 1994. Thematic Roles, Focus, and the Representation of Events. *Language and Cognitive Processes* 9:519–548.

Stevenson, R. J., A. W. R. Nelson, and K. Stenning. 1993. Strategies in Pronoun Comprehension. In *Proceedings of the 15th Annual Meeting of the Cognitive Science Society*, 976–981.

Stevenson, Rosemary J., Alexander W. R. Nelson, and Keith Stenning. 1995. The Role of Parallelism in Strategies of Pronoun Comprehension. *Language and Speech* 38(4):393–418.

Stillings, Justine T. 1975. The Formulation of Gapping in English as Evidence for Variable Types in Syntactic Transformations. *Linguistic Analysis* 1(2):247–273.

Sweetser, Eve E. 1990. *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure.* Cambridge University Press.

Takami, Ken-ichi. 1988. Preposition Stranding: Arguments Against Syntactic Analyses and an Alternative Functional Explanation. *Lingua* 76:299–335.

van Dijk, Tuen A. 1979. Pragmatic Connectives. *Journal of Pragmatics* 3:447–456.

van Hoek, Karen. 1997. *Anaphora and Conceptual Structure.* Chicago: University of Chicago Press.

van Oirsouw, Robert R. 1985. A Linear Approach to Coordinate Deletion. *Linguistics* 23:363–390.

Walker, Marilyn A., Masayo Iida, and Sharon Cote. 1994. Japanese Discourse and the Process of Centering. *Computational Linguistics* 20(2):193–232.

Webber, Bonnie Lynn. 1978. *A Formal Approach to Discourse Anaphora.* Doctoral dissertation, Harvard University. Reprinted in Outstanding Dissertations in Linguistics Series, Garland Publishers, 1979.

Webber, Bonnie Lynn. 1987. The Interpretation of Tense in Discourse. In *Proceedings of the 25th Meeting of the Association for Computational Linguistics*, 147–154. Stanford, CA.

Webber, Bonnie Lynn. 1988. Tense as Discourse Anaphor. *Computational Linguistics* 14(2):61–73.

Williams, Edwin. 1977. Discourse and Logical Form. *Linguistic Inquiry* 8(1):101–139.

Wilson, Deirdre, and Dan Sperber. 1979. Ordered Entailments: An Alternative to Presuppositional Theories. In *Syntax and Semantics XI: Presupposition*, ed. C. K. Oh and D. Dinneen. 299–323. Academic Press.

Winograd, Terry. 1972. *Understanding Natural Language.* Academic Press, New York.

# Index


abduction, 31, 59, 208
antecedent contained ellipsis, 63–66
Appelt, Douglas, 31, 59, 146, 208
apples, 81, 91
Ariel, Mira, 156
Asher, Nicholas, 71, 99, 182, 184–188, 192–194, 197–200, 208

Bates, Elizabeth, 174
beansprouts, 113
binding conditions, 67–69, 71
  condition A, 36, 55, 58, 60, 69–70, 92
  condition B, 55, 58, 60, 92
  condition C, 36, 40, 55, 58, 60, 62, 93
Brennan, Susan, 147, 150–152, 160, 169
broccoli, 34

cake, 129–131
cataphora, 43, 90–91, 148
Centering theory, 148–153, 156, 161, 178
  temporal centering, 182
Chafe, Wallace, 168
Chambers, Craig, 177
Chao, Wynn, 44, 65, 77, 83, 92
chicken, glazed, 205
chili, 114, 127
Chomsky, Noam, 36
coffee, 198
cognitive status, *see* information status
coherence
  and conjunction, 26, 32–34
  core cognitive principles underlying, 4, 14, 23–31
  identifying in data, 31–34
  mechanical methods for establishing, 31
  relations, 3, 14–23
  what is, 2–4
comparatives, 59–61, 87, 93
Comrie, Bernard, 182, 188–191, 200, 202, 206
conjunction
  symmetric versus asymmetric, 83–84, 106, 113–116
contrastive accent, 67–69, 86, 97, 158
cookies, 187, 193
coordinate structure constraint, 6, 101–142
  conjunct and element constraints, 102
Cote, Sharon, 151, 160
Crawley, R., 168, 172

Dahl, Östen, 74
Dale, Robert, 25, 33
Dalrymple, Mary, 35, 39–41, 53, 55, 59, 64, 67, 71–74, 126
Deane, Paul, 111–112, 120, 121, 129, 133

deep and surface anaphora, 76–78
Discourse Representation Theory, 71
discourse structure
  and extraction, 134
  and tense, 204–205
*do so* anaphora, 77–78
Dowty, David, 182, 205
Dretske, Fred, 86

Egg McMuffin, 84
Erteschik-Shir, Nomi, 119
Evans, Gareth, 67
event structure, 195, 197
extraction, 6, 101–142

Fiengo, Robert, 35, 39, 63, 65, 66, 70
Fillmore, Charles, 111
fish, 205
  salmon, 22

gapping, 5, 81–99, 128
  interaction with VP-ellipsis, 95–96
  presuppositional effects of, 86–88
  syntactic and semantic properties of, 85–91
Gardent, Claire, 73, 75
Gawron, Jean Mark, 65, 74, 126
Gazdar, Gerald, 103
Generalized Phrase Structure Grammar, 103
Gilliom, Laura, 45
Goldsmith, John, 6, 107–109, 140
Goodall, Grant, 103
Gordon, Peter, 45
Greenbaum, Sidney, 89
Greene, Steven, 164
Grosu, Alexander, 102
Grosz, Barbara, 45, 147–151
Gundel, Jeanette, 156

Haïk, Isabelle, 35, 37, 65
Halliday, M.A.K., 3, 23, 29, 42
Hankamer, Jorge, 76–78, 83, 87
Hardt, Daniel, 35, 40, 41, 44, 54, 77
Hasan, Ruqaiya, 3, 23, 29, 42
Hedberg, Nancy, 156
Hellan, Lars, 35
Hestvik, Arild, 35, 39, 59–61, 70, 109
Hinrichs, Erhard, 182–184, 186–187, 191, 193, 194, 200, 202, 204
Hirschberg, Julia, 73
Hobbs, Jerry R., 2–4, 14, 15, 19, 22, 23, 25, 30–31, 59, 62, 74, 75, 145–148, 162, 208
Hovy, Eduard, 3
Huck, Geoffrey, 104, 105, 113–116, 127, 128, 141
Hudson, Richard, 82, 83, 93
Hume, David, 3, 15, 34

Iida, Masayo, 151, 160
information status, 156, 168

Jackendoff, Ray, 43, 83, 90, 93
Jayaseelan, K., 83
Johannessen, Janne, 102, 128, 141
Joshi, Aravind, 147–151

Kameyama, Megumi, 50, 143, 147, 153, 158, 160, 177–178, 182, 192
Kamp, Hans, 182, 205
Kennedy, Christopher, 66, 93
Kitagawa, Yoshihisa, 36, 39
Klein, Ewan, 103
Kleinman, D., 168, 172
Kliegl, R., 174
Knott, Alistair, 25, 33
Kuno, Susumu, 83, 85, 87, 97, 103, 117–119

Lakoff, George, 6, 43, 90, 103, 109–111, 126, 133, 137, 139
Lakoff, Robin, 32, 83, 123
Lambrecht, Knud, 156, 168
Langacker, Ronald, 171

Lappin, Shalom, 35, 36, 39, 65–69, 77, 119
Lascarides, Alex, 182, 184–188, 192–194, 197–200, 208
Lawler, John, 136
Leech, Geoffrey, 89
Levin, Nancy, 5, 83–85, 97–99, 128
Levine, Robert, 135, 137, 138
Linguistic Discourse Model, 71
Lobeck, Anne, 44, 77
Longacre, Robert, 3, 23

Mann, William, 3, 23, 24, 113
Martin, James R., 23, 29
Martin, Paul, 31, 59, 146, 208
Maslow's Maxim, 1, 42, 135, 179, 206, 207
Maslow, Abraham, 1, 42, 135, 179, 206, 207, 210
May, Robert, 35, 39, 63, 65, 66, 70
McKoon, Gail, 164
McWhinney, Brian, 174
Mellish, Chris, 33
Mittwoch, Anita, 148
Moens, Mark, 195

Na, Younghee, 104, 105, 113–116, 127, 128, 141
Napoli, Donna Jo, 87
Neijt, Anneke, 83
Nelson, Alexander, 174
Nerbonne, John, 182
nominalizations, 40, 53–55, 57, 62
Noordman, Leo, 23, 24, 26–30, 34

Oehrle, Richard, 159
open proposition, 86–93, 97
oranges, 81, 91

parasitic gaps, 66–67
Partee, Barbara, 72, 182
Passonneau, Rebecca, 182, 192
Pereira, Fernando, 35, 38, 40, 41, 64, 67, 71–74
plums, 25
Poesio, Massimo, 182, 192
Polanyi, Livia, 3
Pollard, Carl, 102, 147, 150–152, 160, 169
Postal, Paul, 125, 135–138
Prüst, Hub, 71, 99
pragmatic control, *see* situationally-evoked referents
Prince, Ellen, 5, 83–86, 97, 99, 128
pronouns, 6, 143–179
  analogy with VP-ellipsis, 43–44, 50
  attention-driven approaches to, 147–152
  coherence-driven approaches to, 145–147
  resolution preferences, 143
    grammatical parallelism, 144, 154, 160, 162, 173, 174, 177
    grammatical role, 143, 148, 154, 158, 168, 169, 171, 172, 174, 177
    thematic role, 168
Pullum, Geoffrey, 103

Quirk, Randolph, 89

Ratcliff, Roger, 164
Redeker, Gisela, 29
Reichenbach, Hans, 182–184, 189, 190, 202
Reinhart, Tanya, 73, 118, 148
Reyle, Uwe, 182
Rhetorical Structure Theory, 24, 113
Rooth, Mats, 66
Rosenbaum, Harvey, 83
Ross, John R., 6, 83, 101, 102, 138

Sag, Ivan, 35, 39, 63, 69, 76–78, 81–83, 85, 86, 89, 93, 95–96, 102, 103
Samet, Jerry, 22, 23
Sanders, Ted, 23, 24, 26–30, 33, 34
Scha, Remko, 71, 99
Schachter, Paul, 44, 77, 103
Schank, Roger, 22, 23

Schmerling, Susan, 83, 105–107, 110, 111, 128, 139
Shieber, Stuart, 35, 40, 41, 64, 67, 71–74
Sidner, Candace, 50, 147, 153, 160
Siegel, Muffy, 83
situationally-evoked referents, 76–78
Smyth, Ron, 159, 173–174, 177
Sperber, Dan, 86, 97
spinach, 2
Spooren, Wilbert, 23, 24, 26–30, 34
statives
  and extraction, 104
  and tense, 202–204
Steedman, Mark, 83, 86, 87, 97, 103, 195
Stenning, Keith, 174
Stevenson, Rosemary, 168, 172, 174
Stickel, Mark, 31, 59, 146, 208
Stillings, Justine, 83
strict and sloppy readings, 35, 58, 72–76
subjacency, 37, 65
Svartvik, Jan, 89
Sweetser, Eve, 29
syntactic reconstruction, 38, 48, 51, 70–71, 91–93, 95

Takami, Ken-ichi, 120
tense, 7, 181–206
  as anaphoric, 183–184, 195–197
  coherence-driven approaches to, 184–186
  Comrie's analysis of, 188–191
  future, 200–202
  Reichenbach's analysis of, 183–184
Thompson, Sandra, 3, 23, 24, 113
topic, 117–132, 141
  and *respectively*, 125–126
traces, 63–67, 135
  and gapping, 89

van den Berg, Martin, 71, 99
van Dijk, Tuen, 29
van Hoek, Karen, 156
van Oirsouw, Robert, 83
vehicle change, 66, 71
verb phrase ellipsis, 4, 35–79
  anaphoric property of, 43–46, 50, 54, 59, 65, 76–77
  syntactic reconstruction and, 38, 48, 51, 70–71

Walker, Marilyn, 147, 150–152, 160, 169
Ward, Gregory, 73, 78, 127
Webber, Bonnie, 69–70, 73, 182, 195–197, 204
Weinstein, Scott, 147–151
Williams, Edwin, 35, 39
Wilson, Deirdre, 86, 97
Winograd, Terry, 6, 146
Wirth, Arthur G., 210

Zacharski, Ron, 156