A Data-Driven Methodology for Motivating a Set of Coherence Relations

Alistair Knott

Ph.D.
University of Edinburgh
1996

By *Consequence*, or TRAYNE of Thoughts, I understand that succession of one Thought to another, which is called (to distinguish it from Discourse in words) *Mentall Discourse*. When a man thinketh on any thing whatsoever, His next Thought after, is not altogether so casuall as it seems to be. Not every Thought to every Thought succeeds indifferently...

The generall use of Speech, is to transferre our Mentall Discourse, into Verbal; or the Trayne of our Thoughts into a Trayne of Words...

Thomas Hobbes: Leviathan *(1651)*

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 of the world we find ourselves in, principles that rest ultimately on some notion of cognitive economy.

Jerry Hobbs: On the Coherence and Structure of Discourse *(1985)*

Abstract

The notion that a text is coherent in virtue of the 'relations' that hold between its component spans currently forms the basis for an active research programme in discourse linguistics. **Coherence relations** feature prominently in many theories of discourse structure, and have recently been used with considerable success in text generation systems. However, while the concept of coherence relations is now common currency for discourse theorists, there remains much confusion about them, and no *standard* set of relations has yet emerged.

The aim of this thesis is to contribute towards the development of a standard set of relations. We begin from an explicitly empirical conception of relations: they are taken to model a collection of **psychological mechanisms** operative during the tasks of reading and writing. This conception is fleshed out with reference to psychological theories of skilled task performance, and to Rosch's notion of the **basic level** of categorisation.

A methodology for investigating these mechanisms is then presented, which takes as its starting point a study of **cue phrases**—the sentence/clause connectives by which they are signalled. Although it is conventional to investigate psychological mechanisms by studying human behaviour, it is argued here that evidence for the constructs modelled by relations can be sought in an analysis of the linguistic resources available for marking them explicitly in text.

The methodology is based on two simple linguistic tests: the **test for cue phrases** and the **test for substitutability**. Both tests are functional in inspiration: the former test identifies a heterogenous class of phrases used for linking one portion of text to another; and the latter test is used to discover when a writer is willing to substitute one of these phrases for another. The tests are designed to capture the judgements of ordinary readers and writers, rather than the theoretical intuitions of specialised discourse analysts.

The test for cue phrases is used to analyse around 200 pages of naturally occurring text, from which a corpus of over 200 cue phrases is assembled. The substitutability test is then used to organise this corpus into a hierarchical **taxonomy**, representing the substitutability relationship between every pair of phrases.

The taxonomy of cue phrases lends itself neatly to a model of relations as **feature-based constructs**. Many cue phrases can be interpreted as signalling just some features of relations, rather than whole relations. Small extracts from the taxonomy can be used systematically to determine the alternative values of single features; complex relation definitions can then be formed by combining the values of many features.

The thesis delivers results on two levels. Firstly, it sets out a methodology for motivating a set of relation definitions, which rests on a systematic analysis of concrete linguistic data, and demands a minimum of theoretical assumptions. Also provided are the relation definitions which result from applying the methodology. The new definitions give an interesting picture of the variation that exists amongst cue phrases, and offer a number of innovative insights into text coherence.

Acknowledgements

I would like to take this opportunity to express my deepest thanks to my supervisors, Chris Mellish, Robert Dale and Steve Isard, for their immense help and encouragement over the last four years. Robert first prompted me to think about the theoretical status of coherence relations; he led me to appreciate the need for a firm foundation for these constructs, as well as overseeing the assembly of the taxonomy of cue phrases. Chris was a huge help in guiding me through the theoretical interpretation of the taxonomy, in all its details. Steve provided a valuable perspective on the thesis as a whole.

I would also like to thank my examiners, Donia Scott, Graeme Ritchie and Alex Lascarides, for their detailed comments and suggestions on the manuscript, and for a very interesting discussion during the viva.

Many other people have been influential in the development of the ideas presented here. A big thank you to all of the following, whose feedback and suggestions were much appreciated: Jean Carletta, Ed Hovy, Elisabeth Maier, Kathy McCoy, Johanna Moore, Megan Moser, Jon Oberlander, Cécile Paris, Orin Percus, Ted Sanders, Karen Spärck Jones, Wilbert Spooren, Manfred Stede, and Keith Vander Linden. Special thanks to Ted, for long discussions in several countries, over many glasses of Belgian beer.

It seems to be easier to start a PhD than to finish one. This thesis would certainly not have been finished without the concerted efforts of a large and ruthless team of people, and I would like to say thank you to all of them for managing to force me into submission. Firstly to Chris, for the diverse rhetorical strategies he has employed to encourage me to finish, ranging from rational arguments and encouragements, through thinly veiled warnings, to direct threats and emotive appeals; and finally for eliciting from me a solemn oath to stop my research and start writing. Next, to the Engineering and Physical Sciences Research Council (EPSRC), for their generous support over three years; and also for their Faustian 'four year deadline', which I have watched approaching with increasing horror but which has finally—I think—had its intended effect on me. And lastly, to the assorted PhD students and denizens of the AI department, who have made my life on E floor most pleasant. Thanks are due to Alberto, Chris, Ed, Elena, Ian, Ion, Janet, Jean, Jeremy, Kim, Nam Seog, Steffan, and Wamberto, who have all been a great influence on me during the past few years. Needless to say, any defects that remain in my character are solely my own responsibility.

I would like to give special thanks to my family. Their constant support, particularly during the past year when I was running on empty, has been very important to me. Thank you Mum, Dad, Beck, Har, Jule. Thanks also to my wider family, in Britain, in New Zealand and in Tonga.

Above all I would like to thank Mele, who has had to put up with a lot of ranting and raving on the subject of discourse structure just recently. Thank you for your love, and cheerfulness, and steadfastness. And especially for the notice pinned above my bed, reminding me of the important fact that No-One Reads PhD Theses.

Declaration

I hereby declare that I composed this thesis entirely myself and that it describes my own research.

Alistair Knott
Edinburgh
December 3, 1996

Contents

List of Figures

Chapter 1

Introduction

1.1 Discourse Coherence: The Phenomenon Under Investigation

Theories of discourse coherence aim to investigate the rules which govern how clauses and sentences can be strung together into a text. While syntactic theories concentrate on the internal structure of sentences, theories of discourse look beyond single sentences to the organisation of larger units of language, such as conversations, books or newspaper articles.

The starting point for any such theory is that what we normally think of as 'connected discourse' is more than just a concatenation of random sentences. This fact is uncontentious, and a few examples will suffice to illustrate it. Consider, for instance, how a reader opening a newspaper might react to the two paragraphs given in Figure 1.1. The first is taken from the editorial of an issue of *The Economist*. The second contains exactly the same clauses as the first, but arranged in a different order. Whereas the former paragraph can easily be understood, the latter is at best odd, and at worst completely unintelligible. We can say, pending more precise definitions, that the former text is **coherent**, whilst the latter is **incoherent**.

1.1.1 Two Types of Incoherence

At least two separate problems can be identified in the second paragraph in Figure 1.1. One has to do with deciding about the referents of anaphoric expressions in the text. For instance, the phrase *these middle-aged rich people* cannot be interpreted—it occurs before any such group of people has been introduced. It is simply impossible for the reader to understand or evaluate any propositions in which such unresolved anaphora appear.

A second class of problems in the incoherent text relate to the reader's inability to understand why two portions of text are placed next to one another in the first place. In these cases, there are no problems with finding interpretations for the two portions of text; it is merely surprising that they have been juxtaposed. For instance, consider

The World in 1993

1993 will start with the world in a pessimistic frame of mind. That gloom should soon dispel itself. A clear economic recovery is under way. Though it will be hesitant at first, it will last the longer for being so. If you are sitting in one of the world's blackspots, this prediction will seem hopelessly optimistic. But next year's wealth won't return to yesteryear's winners; these middle-aged rich people need to look over their shoulders to the younger world that is closing in on them.

The World in 1993

1993 will start with the world in a pessimistic frame of mind. A clear economic recovery is under way. That gloom should soon dispel itself. These middle-aged rich people need to look over their shoulders to the younger world that is closing in on them. But next year's wealth won't return to yesteryear's winners; it will last the longer for being so if you are sitting in one of the world's blackspots. Though it will be hesitant at first, this prediction will seem hopelessly optimistic.

Figure 1.1: Coherent and Incoherent Texts

the first two sentences of the incoherent text:

(1.1) 1993 will start with the world in a pessimistic frame of mind. A clear economic recovery is under way.

Why should the world be pessimistic if an economic recovery is under way? The two statements appear to be contradictory, and we do not expect the writer to countenance both of them.

The two examples above appear incoherent in quite different ways. In fact, it is a common assumption amongst discourse linguists that coherence is not a unitary phenomenon at all, but rather the product of a number of different mechanisms. If this is the case, then it should be possible to undertake an explanation of some *aspects* of discourse coherence without providing a complete account of it. In the present study, this will be the objective.

1.1.2 The Role of Context in Decisions about Coherence

It should be noted immediately that when we rule the second paragraph in Figure 1.1 as incoherent, we are not making a judgement about an isolated piece of text; we are also making certain assumptions about its reader, its writer, and the situation in which it was produced. For instance, if the text was accompanied by a photograph of a group of well-heeled businessmen, then a plausible referent for the phrase *these middle-aged rich people* would be available. Again, if we imagine the text to be aimed at readers who believe that the world is full of anarchists who dread an economic recovery, then the sentences in Example 1.1 become intelligible.

Judgements about coherence, therefore, are not made about strings of sentences, but about complexes comprising several additional components, including a reader, a writer, the stock of world knowledge which they share, and what we might call a 'communicative situation' in which they find themselves. This point has often been made in the past: Halliday and Hasan (1976) were influential in distinguishing between a **text** and the **context** in which it appears, envisaging a strong role for the latter concept in an explanation of the former. While a text is relatively concrete, taking the form either of marks on a page or of a series of spoken sounds, its context is a much more intangible notion, and much harder to capture. Indeed, most of the difficulties which arise in simulating the tasks of text production and interpretation have to do with the adequate representation and use of the reader and writer's world knowledge, and of the situation in which a text is produced.

The importance of context in the study of discourse makes the presentation of example texts somewhat of a problem. It is impossible to specify the context for a text under discussion in all its detail; in what follows, the reader will often have to imagine plausible contexts for the texts being discussed. An effort will be made to keep this task as easy as possible.

It is interesting to discover that inventing texts which are incoherent in *any* context is actually quite difficult. This is particularly true as regards the kind of incoherence which results from juxtaposing apparently unrelated portions of text—if we allow ourselves to imagine unusual contexts, we can often give very bizarre discourses coherent interpretations. For instance, consider Text 1.2:

(1.2) Sally decided to take the history course. The ducks on the lake were not eating the bread.[1]

This text, although odd at first sight, can be perfectly well understood if we imagine Sally to have unusual superstitions about the ducks on the lake.

It seems as though readers often go out of their way to find a coherent interpretation of a text, even when this involves making some odd assumptions about it. In a way, this is to be expected: it cannot be easy for a reader to abandon the assumption that the text conforms at some level to Gricean Maxims (Grice (1975)); but this is often what is required in order to appreciate that a text is incoherent. In addition, readers are not often exposed to incoherent texts, and so it is natural for them to operate on the assumption of coherence.

1.1.3 Degrees of Coherence and Incoherence

A final observation about coherence is that it can be present in texts in varying degrees. Texts should not be thought of as either coherent or incoherent, rather as more or less coherent; the dividing line between coherent and incoherent texts is a hazy one.

Part of this haziness can be traced to the problem mentioned in the previous section, of the role of context in decisions about coherence, and of the tendency of readers to look for plausible contexts even for the most implausible texts. However, even if contexts are fully specified in advance, texts still admit of degrees of coherence.

[1] Linguistic examples are hand-crafted unless otherwise noted. One of the reasons for this is so that examples of incoherent texts can be provided as easily as examples of coherent ones.

There are two reasons for this. The first is that coherence is a *global* measure of text. Thus it is possible that a text is coherent in some places, but incoherent in others. It would of course be preferable to use a more local measure of coherence to avoid this problem. However, we are still working with a pre-theoretical concept at present, and without a theory of coherence, it seems premature to try tracking down problems to specific features of a text.

A second reason for the haziness of the concept of coherence is simply that we can understand some texts more easily than others. If a text is well written, the intended reader should understand it without much difficulty; however, even if it is badly written, a reader may still be able to piece together its meaning. Consider the following example, from 'The Brief English Handbook' (Dornan and Dawe (1984)):

(1.3) Alice Adams is a successful writer, and she lives in San Fransisco, and she has received grants from the Guggenheim foundation and the National Foundation for the Arts.

This text is stylistically awkward, but it still manages to get its message across. Ideally, however, we would like a way of distinguishing between such texts and 'well-written' ones.

1.1.4 A Definition of Coherence

For the purposes of this thesis, an operational definition of coherence is required; preferably, one which does not pre-empt too many theoretical questions. In what follows, therefore, we will think of an incoherent text as one whose 'structure' a high school teacher would be inclined to question or correct if it appeared as part of a student's essay. We will not be concerned with any corrections a teacher might make arising from bad grammar or spelling mistakes, or from errors of fact: an incoherent text is one which avoids such errors, but is nonetheless hard or impossible to understand.

Admittedly, this is a very informal definition to start off with. But an initial description of 'what a theory is about' precisely *should* be informal, so as not to begin doing the work intended for the theory itself. Subsequent more precise conceptions of coherence should ultimately be judged according to how well they reflect an informal conception such as this one.

1.2 First and Second Order Tasks for Text Analysts

Theories of coherence call for texts to be analysed—that is, broken up into parts which are given descriptions in theoretical terms. The theoretical description of an individual text should contribute to an account of why it is or is not coherent.

Two quite separate questions confront a discourse analyst. Firstly, what are the appropriate theoretical terms to be used for describing texts, and how are they to be used? Secondly, what is the right analysis for any particular text—for instance the text on the left in Figure 1.1? The first question calls for a decision amongst competing theories of coherence. The second calls for a decision about how to represent a particular text in terms of one theory or another. We can refer to this latter task as the **first-order** task, and to the task of deciding between theories as the **second-order** task. Both tasks have their problems. Theorists can (and do) disagree about the primitives to be used by a theory of coherence; and also, once a theory has been agreed on, over individual analyses.

This thesis is directed primarily at the second-order task—that of deciding on the aspects of a text which an analysis should identify, rather than of determining *how* these aspects are to be identified for any given text, either by a human or by a machine. The second-order task seems logically primary; after all, it is not possible to analyse a text until a particular representation formalism has been chosen. The first-order task will also be addressed to some extent, however, as is detailed in Section 1.3.2.

1.3 Requirements for a Theory of Discourse Coherence

While the previous sections described *what* is to be investigated in this thesis, the present section examines *how* it is to be investigated. Two requirements for a theory of coherence are set out, one concerning the goals that the theory should have, and one concerning the kinds of intuitions which should be admissible as evidence.

1.3.1 Descriptive versus Explanatory Adequacy

To assess a theory of discourse coherence, we can make use of some Chomskyan terminology. One of the tasks of the theory is clearly to provide a means for distinguishing between coherent and incoherent texts. It should be possible to use the theory to decide whether any given text is coherent or incoherent; if the theory is good, then these decisions will correspond to our own judgements of coherence. A theory which satisfies this requirement can be termed **descriptively adequate**.

A second, more stringent requirement for a theory of coherence is that of **explanatory adequacy**. Chomsky's (1964) original formulation of this notion is with reference to syntactic theories. An explanatorily adequate theory is one which is able not only to distinguish between well-formed and ill-formed texts, but also to explain why it is that well-formed texts are well-formed; in other words, to explain why it is that we have the grammars that we do. The notion can apply just as easily to theories of coherence—in this case, it would demand that a theory is able not only to distinguish between coherent and incoherent texts, but also to explain why it is that readers and writers have these standards for coherence.

It is ambitious even to aim for descriptive adequacy in a theory of discourse. However, this is no reason to completely ignore the criterion of explanatory adequacy. In this thesis, particularly in Chapter 3, the notion of explanatory adequacy will figure quite prominently.

1.3.2 Pre-Theoretical and Post-Theoretical Intuitions

The raw data for a theory of discourse are 'intuitions of coherence'. As Section 1.1.4 specifies, we are examining the kind of intuitions that a high school teacher might have about a range of different texts—if corrections of a certain kind are felt to be needed, a text is classed as incoherent, otherwise, it is classed as coherent.

This type of judgement about a text is what a theory of coherence must explain; it can be termed **pre-theoretical**. It should be contrasted with another type of judgement about text, which we can call **post-theoretical**—such judgements are about 'the way a text should be described'. A post-theoretical judgement might be, for instance, that 'a discourse segment boundary falls between sentences s_1 and s_2', or that 'the focused entity in span S_5 is E_3'. The first-order task of text analysis is normally thought of as involving such judgements.

As readers and writers of text, we might well have our own ideas about what makes a text hang together well, and these might even involve notions of text segments, focus and so on. If asked what the topic of a given sentence is, or how a given discourse should be segmented, we would probably be able to make some suggestions. However, these intuitions should be differentiated sharply from our intuitions of coherence. They cannot be relied upon in the same way: while intuitions of coherence have the status of irrefutable facts to be explained, post-theoretical intuitions are intended as *part* of an explanation of those facts, and are subject to the same standards of assessment as any other explanation.

In fact, our intuitions about text analysis do not seem to be especially reliable. As already noted, disagreements between analysts are very common. And although some measure of consensus among analysts is often claimed for a particular theory, differences between analysts from different theoretical backgrounds are often quite significant. The problem is that it is hard to see how such differences can be resolved. Why should one theorist's intuitions be any more reliable than another's? The theorists are competing for an explanation of the same text; if we are looking for a single theory of text, then there is no question of both theorists' intuitions being equally good.

A second, more fundamental problem with a reliance on post-theoretical intuitions can also be noted. There is evidence in the psychological literature that people's intuitions about how they perform complex and highly learned tasks differ considerably from the way they are actually performed. For instance, Berry and Broadbent (1984) have found that subjects' verbalisations about how they perform a task do not change markedly as their skill increases, although qualitative changes in performance can be observed. The tasks of text generation and interpretation are highly skilled, and thus it could be that our intuitions about them (however much we agree or disagree) are inaccurate.

How then are we to go about analysing texts? In this thesis, a method will be suggested that relies less heavily on post-theoretical intuitions.[2] It is proposed that some of the constructs in a theory of discourse coherence can be linked a priori to a *second* class of pre-theoretical data, independent from straightforward judgments of coherence. The

[2] Of course, the development of the theory still requires post-theoretical intuitions. The problem of using *only* pre-theoretical intuitions is highlighted by Scott and Paris (1995).

new class of data concerns the judgements of ordinary writers about when two phrases of a certain type can be substituted for one another in a text, and will be described in detail in the following chapters. The important thing to note now is that the judgements are ones which a normal writer is frequently called upon to make; they do not require the analyst to think explicitly about theoretical concepts like 'focus' or *text segments* and so diverge from tasks present in the normal writing process. If the theory of coherence is successful, then, it will identify a correlation between the two classes of pre-theoretical intuitions.

This approach to theorising about discourse is much closer to the traditional pattern of empirical theories; it consists in making predictions about the relationship between two independent sets of data to be explained. For another example of this approach, we can again refer to theories of syntax. In a syntactic theory, the original pre-theoretical intuitions to be accounted for are judgements of well-formedness. These are explained in terms of a theory whose central construct is that of a 'constituent'. This construct is in turn grounded in *other* pre-theoretical intuitions, enshrined in the so-called 'tests for constituency'. For instance, it is specified that constituents can typically be moved from one part of a sentence to another, or replaced by an anaphoric expression, without affecting well-formedness. The crucial point is that our intuitions about these manipulations are different from our intuitions of well-formedness themselves. And moreover, they cannot be questioned; they are just another phenomenon to be explained. Essentially, what we have is a theory which makes predictions about a relationship between two independent classes of data. If these predictions are borne out, we have an empirical result.

Of course, the development of the theory itself still requires post-theoretical intuitions. Definitions for theoretical constructs must be proposed, and we must justify the decision to link these constructs to pre-theoretical judgements.

1.4 An Outline of the Thesis

We began this chapter by introducing and expanding on the notion of discourse coherence. We then distinguished the task of creating a theory of discourse coherence from the task of analysing particular texts in terms of such a theory—only the former task will be attempted in this thesis. Finally, we set out the requirements for a theory of coherence; namely that it seeks for an explanatory account of the phenomenon, and that it draws only on the pre-theoretical intuitions of ordinary readers and writers.

A summary of the rest of the thesis will now be given.

Chapter 2: Coherence Relations

The thesis targets a particular class of discourse theories; those which attribute the coherence of a text (at least in part) to the **relations** which hold, at different levels of hierarchy, between its various sub-parts. Many such theories can be found in the literature; in the last few years, the notion of coherence relations has become increasingly popular as they have been successfully adapted for use in computational applications such as text generation systems.

This chapter reviews the many different theories of coherence relations, emphasising the differences between them, and their lack of agreement on a standard set of relations. The proliferation of alternative sets of relations is the central problem to be addressed in the thesis. What is needed is a clear conception of the role of relations, and based on this, a method for justifying one particular set of relations over the others.

Chapter 3: Motivating a Set of Relations

In this chapter two existing approaches for justifying relations are described and compared. One approach associates relations with linguistic conjunctions or **cue phrases**. This permits a subtle classification of relations, but seems to lack explanatory adequacy; it is not clear why relations thus defined should provide a particularly revealing account of text. A second approach is to think of relations as modelling a set of psychological constructs used by readers and writers when they process text. This promises a more explanatory account; but there are problems with the experimental methodologies designed to investigate these constructs, because we have no reliable experimental window on 'the relations people use'.

The main proposal in the thesis is that these two methods can be combined, in such a way as to maximise the advantages of both, while avoiding their main problems. It is argued that cue phrases can be taken as evidence for relations precisely if they are thought of as modelling psychological constructs.

Chapter 4: A Data-Driven Methodology for Determining a Set of Relations

On the basis of the argument in Chapter 3, a step-by-step methodology is proposed for motivating a set of relations, making use of the pre-theoretical intuitions of readers and writers. The first step is to gather a corpus of cue phrases. A pre-theoretical **test for cue phrases** is described, which is used to gather a corpus of some 200 phrases from several hundred pages of naturally occurring text. These phrases are then organised into a taxonomy, using a second pre-theoretical **test for substitutability**, which taps writer's intuitions about whether one phrase can replace another in a given context. The corpus of cue phrases is given in Appendix A, and the taxonomy of cue phrases in Appendix B.

Chapter 5: Preliminaries for Defining a Set of Relations

This chapter outlines how the taxonomy of cue phrases can be used systematically to motivate a set of relation definitions. It is argued that the taxonomy lends itself very naturally to a conception of relations as **feature-based constructs**. Some general criteria for the individual features to be motivated are then discussed.

Chapter 6: Using the Taxonomy to Create Relation Definitions

In this chapter, a preliminary set of relation definitions is put forward. To begin with, the taxonomy of cue phrases is used systematically to motivate a number of orthogonal features: each feature is justified individually on the basis of appropriate extracts from the taxonomy. After this, the issue of how these features can be assembled to create complex relation definitions is discussed. The complete set of feature definitions is given in Appendix C, and a preliminary set of relation definitions is given in Appendix D.

The task of motivating features from the taxonomy is an arduous one, and there is plenty of room for improvement and addition to the sets of features and relation definitions reached in this chapter. However, a number of interesting new ideas do emerge as the theoretical interpretation of the taxonomy is developed. The fact that substitution methodology yields these fruitful results is an additional argument in its favour.

Chapter 7: An Evaluation of the Substitution Methodology

In this chapter, a number of problems with the substitution methodology are taken up. These have to do partly with the operation of the substitutability test itself, and partly with the set of relations which it eventually sanctions. None of these problems appears fatal to the proposed methodology, although they all point to further interesting avenues of research.

Chapter 8: Conclusions

The objectives of the thesis are quite straightforward. It attempts to remedy a current problem for theories of discourse, namely the confusing proliferation of coherence relations. It does so by giving firstly a concrete proposal about what relations should be thought of as modelling; secondly a clear methodology for investigating these entities; and thirdly, as a result of applying this methodology, an embryonic set of relation definitions.

The contribution of the thesis is twofold. The most tangible contribution is the set of relation definitions itself—these have several features not found in other sets of relations, and promise to be useful both in text analysis and in computational applications. However, it really requires more justification than this for bringing yet another set of relations into the world. The primary aim of the thesis is to establish and argue for a systematic methodology for *determining* a set of relations, based on a battery of fairly replicable linguistic tests. While it is unrealistic to suppose that every researcher who uses it will emerge with the same set of relations, it would at least be preferable if the differences between them could be traced to disagreements over the interpretation of concrete linguistic data, rather than being expressed in terms of intangible first principles.

1.5 The Scope of the Project, and Some Terminology

Before moving on, a word is in order about the type of texts which are within the scope of this study. I will only be looking at *monologue* in this thesis; and I shall concentrate (but not exclusively) on written monologue. Texts which incorporate devices from dialogue such as direct speech or question-and-answer patterns ('How do we know this? By looking at the evidence') will not be considered. I will also concentrate on *English* texts—although some cross-linguistic comparisons will be made in Chapters 3 and 7.

Finally, some terminology. I will use the words **text** and **discourse** interchangeably. Unlike Halliday and Hasan, for whom 'a text' is by definition coherent, I shall think of a text or discourse as any sequence of sentences produced by one writer in a particular context—texts can thus be ruled as coherent or incoherent.

I will refer throughout to the **readers** and **writers** of a text; these terms are intended cover hearers and speakers too, where applicable. To circumvent the biases of English, writers will be referred to generically as 'she', and readers as 'he'.

Chapter 2

Coherence Relations: A Survey of Research

2.1 Introduction

This chapter reviews the existing literature on coherence relations, placing it within the wider context of research into discourse coherence. A roughly historical perspective will be taken in presenting the different theories, beginning from the early work in text linguistics and finishing with the latest computational research.

Sections 2.2 and 2.3 provide some preliminaries to the review. In Section 2.2, a working definition of coherence relations is given, broad enough to encompass all the theories to be discussed here. Section 2.3 considers how (if at all) a purely relational theory must be supplemented in order to achieve a complete account of discourse coherence.

In Section 2.4, the many relational theories in the literature are presented. Recently, some of these theories have found useful application in computational domains, particularly in natural language generation: Section 2.5 describes how coherence relations have been adapted to this task, and surveys the different systems which have resulted.

Section 2.6 outlines the central problem to be addressed in this thesis: the proliferation of sets of relations, both in theories of coherence and in generation systems. The lack of a well-established, standard set of relations is already recognised in the literature: each theorist—and each generation system—uses a differently-defined set of relations. This makes it hard to draw comparisons between the various theories and systems; and worse, threatens to undermine the empirical content of the relation construct in general.

2.2 A Working Definition of Coherence Relations

As was noted at the very beginning, the point of departure for a discourse theorist is that a coherent text is more than just a concatenation of sentences. The properties of any given sentence in a text place certain constraints on those of its neighbouring sentences—in other words, there are constraints on the *relations* between the sentences in a coherent text.

However, to say at this point that 'a text is coherent because of the relations between its sentences' is not to make any particular theoretical claim; we are simply re-iterating the phenomenon to be explained. There are as many 'relations between sentences' as there are pairs of possible sentences—clearly what matters for a coherent text is that the relations be of the right kind.

A proper theory of coherence relations thus needs to propose a particular set of relations, and claim that texts are only coherent if they are composed of relations from within this set. At this point, the term 'relation' acquires a new, technical meaning, referring to a class of well-defined theoretical constructs.

Different theories have taken different approaches to the definition of relations. However, there are some concepts which we will take as central to all relational theories: these are outlined below.

2.2.1 A View to Explaining the Notion of Coherence

Relations have been used for a variety of purposes in the literature: for planning the structure of text, for working out the temporal relations between clauses or sentences, for identifying or generating the referents of anaphora, and so on (see Sections 2.4 and 2.5 for details of some of these). However, in this thesis it will be taken as central to a theory of relations that it is concerned at least in part with providing an explanatory account of text coherence. In practice, this conception of relations often underlies accounts developed for specific purposes. (For instance, if a theory of relations is useful in predicting the pattern of anaphora in a text, it is reasonable to assume that this is because it captures something about the ways a coherent text can be structured.) This concern will be a unifying thread behind all the accounts to be considered here.

2.2.2 Text Spans

A theory of relations must specify the units between which relations are required to hold in a text. In all the theories to be considered here, we can begin by thinking of relations as holding between **text spans**—units of text the size of a clause or bigger. In fact, theories often propose that relations hold between more abstract entities, such as propositions or intentions, which are *expressed* in text spans; however, the link between these entities and the spans which express them is sufficient at least to begin to make comparisons. In particular, a common graphical representation is possible for all coherence relations, as illustrated in Figure 2.1.[1] In this diagram, the horizontal lines represent text spans, and the curved line represents the relation between them. The line is labelled with the name of the relation (in this case, EVIDENCE,[2] a relation taken from Rhetorical Structure Theory—see Section 2.4.3 below).

---

[1] This representation is adapted from that used in Mann and Thompson (1988).

[2] The names of relations will appear in SMALL CAPITALS throughout the thesis.

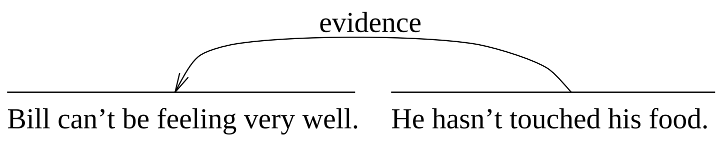

Figure 2.1: Graphical Representation of a Coherence Relation

Text spans can be of different sizes. In some theories, spans are always associated with individual clauses or sentences. In others, they can vary in size from single clauses to whole paragraphs and even larger sections of text. Still other theories allow relations to hold between elements within a single clause. In this thesis, we shall be mainly concerned with spans ranging in size from clauses to paragraphs. It is taken to be a defining feature of coherence relations that they can apply between text spans of any size within this range.

Insensitivity to span size is a useful feature of coherence relations. It permits a hierarchical, recursive account of text, in which small spans are joined together by relations to form larger spans, which are themselves be joined together by other relations. An example is given in Figure 2.2.

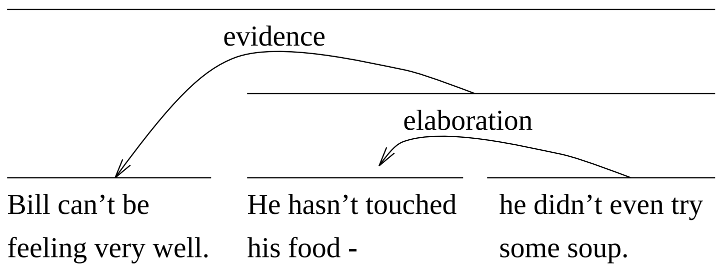

Figure 2.2: A Hierarchical Structure of Relations

2.2.3 Span Structure

As well as a providing a conception of text spans, a relational theory must specify the places where relations are needed in a coherent discourse. Whereabouts in a text must relations be present in order to ensure its coherence? Are they needed between every pair of clauses? Between every pair of paragraphs? How are the various spans in a text supposed to be connected up?

To answer these questions, we need what can be termed a theory of **span structure**, telling us, for any given text, where the coherence relations are expected to be found. Again, there is scope for considerable variation amongst such theories. In some, relations are not expected between every pair of clauses. In others, more than one relation can apply simultaneously between two spans; in others, several consecutive spans can be linked together by a chain of relations; other theories allow relations between discontinuous spans.

2.3 Relations and Focus: Alternative Metaphors for Discourse Coherence

Among other things, the theory of span structure determines how much of the coherence of a discourse is to be attributed to relations. Few, if any, researchers would claim that relations tell us everything we need to know about discourse coherence—they are typically held to account just for some aspects of it. To help delineate the territory to be explained by relations, it is useful to consider some of the other constructs used in theories of coherence; and in particular the concept of **focus**.

Many theories of discourse (eg Sidner (1983), Grosz, Joshi and Weinstein (1983), Brennan *et al* (1987), McCoy and Cheng (1991)) analyse a text in terms of the way its 'focus' moves from one entity to another. At any particular moment while a text is being read, certain objects in its domain of reference are 'in focus': the reader is attending to these elements in particular. The concept of focus has proved extremely useful in modelling anaphora resolution—focused entities are those which can be referred to anaphorically. But it is also used to talk about coherence: the idea is that a writer must obey certain constraints when *changing* the focus from one entity to another. Consider the following two texts:

(2.1) Dow Associates today launched a surprise bid for the beleagured Harris group. Harris reacted speedily, holding an upbeat press conference.

(2.2) *Dow Associates today launched a surprise bid for the beleagured Harris group. Beans, which contain lots of protein, are good for you.

The notion of focus can be used to explain why Text 2.1 is coherent and Text 2.2 is not. At any point in the text, the entities to which the focus can shift are grouped in a list, which Sidner calls the 'potential focus list' and Grosz *et al* call the 'forward center list'. The potential focus list associated with the first sentence in each case contains, among other things, *Harris*, but it does not contain *beans*; Text 2.1 shifts its focus to *Harris* and is thus coherent, but Text 2.2 makes an illegal shift to *beans*, and is thus incoherent.

Clearly, there is a significant possibility for overlap between the metaphors of 'relations' and 'focus'. The coherence of Text 2.1, for instance, might equally be attributed to the existence of a relation (perhaps called something like RESULT or REACTION) between its two sentences.

At the same time, the overlap between the two metaphors is not complete. For instance, the binary nature of relations makes them better suited to an analysis of local constraints on coherence, such as those which might apply between two neighbouring spans regardless of their context. Constraints operative over more than two spans (for example, prohibitions on 'straying from the original topic', or on 'returning to a topic previously closed') might be better dealt with using the focus metaphor. For another thing, relations (as we are thinking of them) do not hold between entities referred to in a text; rather, they link sections of the text itself (or at least the propositions or intentions which underlie it). 'Being in focus', on the other hand, is something more easily attributed to entities being discussed in a text than to portions of the text itself.

Of course, the aspects of coherence targetted by relations and by focus can vary considerably from theory to theory—there is a large degree of flexibility as to how the work should be divided up. The important thing is to make sure that between the two accounts, all aspects of text coherence are covered, and (equally importantly) that there is no redundancy between them. For this reason, when developing a theory of relations it is important to specify exactly which aspects of coherence it is intended to cover. Ideally, what is needed is a clear way of distinguishing the contributions to coherence made by focus and by relations, in such a way as to make best use of both metaphors.

This thesis is primarily concerned with developing a theory of coherence relations; the issue of focus will take a back seat in most of what follows. However, it will reappear with an important role in Chapter 6, when the new theory of relations is to be evaluated, and for this reason it is mentioned here.

2.4 Current Theories of Coherence Relations

To sum up so far: the requirements for a relational theory of coherence are (1) a set of relation definitions; (2) a conception of text spans; (3) a theory of span structure; and (4) some kind of delineation of the phenomena to be accounted for using the relational metaphor. The present section discusses the important relational theories in the literature in the light of these requirements, and examines the differences between them.

I will not go into the details of individual relation definitions in this review, although in many cases a rough idea of the nature of an individual relation is given by its name. My main purpose is to point out the diversity that exists between the many sets of relations that have been put forward.

2.4.1 Some Early Relational Accounts

Attempts to delineate the set of relations that can hold between sections of a discourse date back at least to the 1970s; perhaps the first of note is that of Ballard, Conrad and Longacre (1971). The aim in this study is to catalogue the 'deep' relations which underlie the 'surface' syntactic relations between clauses in complex sentences. Ballard *et al* note that there is no straightforward one-to-one mapping between surface conjunctions and the semantic relationships between the clauses they link—for instance, the conjunction *because* can be used to signal a relation of EFFICIENT CAUSE (*He did it because she wanted him to*) or of GENERIC-SPECIFIC (*They don't taste good because they are bitter*). Conversely, a single semantic relationship can be encoded in different ways: for instance, the sentences *I left before Mary came back* and *I left at about the same time as Mary came back* can be used to describe exactly the same situation, emphasising different aspects of it. The precise roles of deep and surface relations are not completely clear in this account; however, a distinction between the underlying relations in a text and the way they are signalled is retained in many subsequent theories.

The set of deep relations proposed by Ballard *et al* is reworked and refined in Longacre (1983). The new account of relations begins from first principles: a distinction is made between **predications** (typically expressed in clauses) and the **relations** between predications (typically expressed in more than one clause). While the predicate calculus is suitable for modelling clause-internal structure, relations between clauses are therefore described using a variant of propositional calculus. The operations in the propositional calculus—CONJOINING ($\wedge$), ALTERNATION ($\vee$) and IMPLICATION ($\rightarrow$)—are supplemented by a group of TEMPORAL relations to give what Longacre calls the set of BASIC relations; in addition to these he lists a set of ELABORATIVE relations, comprising PARAPHRASE, ILLUSTRATION, DEIXIS and ATTRIBUTION.[3] All of these relations are further subdivided; for instance TEMPORAL relations can be either of OVERLAP or of SUCCESSION. In addition, a parallel set of FRUSTRATED relations is given: these are found in texts where a relation which is expected does not occur. (For instance, the text *They set out for Paris, but didn't arrive* contains a relation of FRUSTRATED SUCCESSION, because there is an expectation that their setting out for Paris will be followed by their arrival.)

A diagram showing the main features of the taxonomy is given in Figure 2.3.

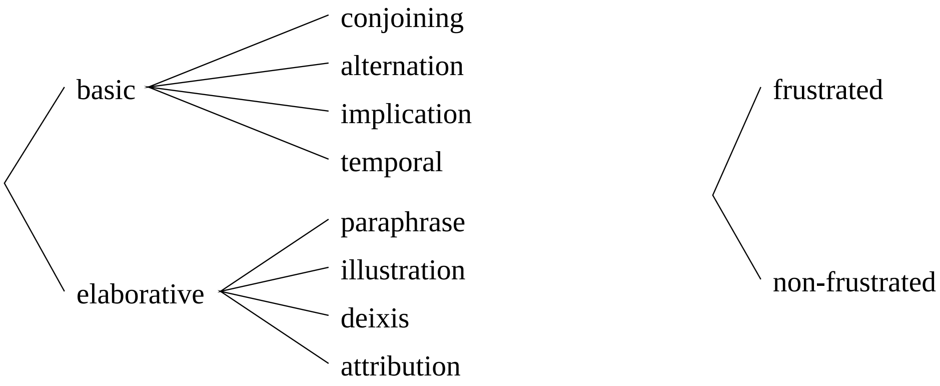

Figure 2.3: The Top Levels of Longacre's Taxonomy of Relations

Grimes' model of relations (Grimes (1975)) gives them a dual role: they provide information, just as clauses do, and they also organise groups of clauses into coherent discourses. Like Longacre, his basic units are clauses (embodying 'lexical predicates') and the relations between them (embodying 'rhetorical predicates'). However, this analysis admits of exceptions. Rhetorical relationships can be found within clauses; for instance, the relation of ALTERNATIVE, typically expressed between two or more clauses, is also found in the single clause *Let's have no more of your neither-here-nor-there observations*.[4] Moreover, the arguments of a lexical predicate can be spread out over several clauses; for instance, the arguments of the predicate 'drop' might be presented as follows: *The stone fell. It hit the ground. Zog made it happen.*).

[3] SMALL CAPITALS will also be used to refer to *groups* of relations identified by a particular theory. Hence, there will be no typographical way of distinguishing between an individual relation and a class of relations.

[4] Though in Grimes' example, it is not very clear what elements in the clause the relation applies between.

The main organising principle in Grimes' taxonomy of relations is a distinction between PARATACTIC and HYPOTACTIC relations. PARATACTIC relations are those whose predicates dominate their arguments in a coordinate fashion; so, for instance, in the ALTERNATIVE relation encoded in *He's in the study or he's in the living room*, the arguments are the two propositions expressed in the two clauses, and the predicate, dominating both propositions, states that exactly one of them is true. HYPOTACTIC predicates relate a central proposition to a subsidiary one; the central proposition then dominates both the subsidiary one and the hypotactic predicate itself. For instance, in the text *He saved the day; he made three touchdowns*, the predicate SPECIFICALLY specifies that the proposition in the second clause gives details about that in the first. This rhetorical proposition, and the specific proposition it refers to, are both subordinate to the proposition in the first clause. NEUTRAL relations are those which can be either PARATACTIC or HYPOTACTIC depending on the context: thus the relation COLLECTION can either group a set of entities together (*We went jogging*), or identify one entity in particular, and associate a secondary group of entities with it (*I went jogging with George and Henry*).

The distinction between PARATACTIC and HYPOTACTIC finds its way in various different guises into many subsequent theories. A second distinction between SYMMETRIC and ASYMMETRIC relations has been less widely adopted—briefly, a relation is SYMMETRIC if the order of its arguments is important and ASYMMETRIC otherwise. It is mainly of relevance to PARATACTIC relations; most HYPOTACTIC relations are ASYMMETRIC.

A representation of Grimes' Taxonomy of Relations is given in Figure 2.4.

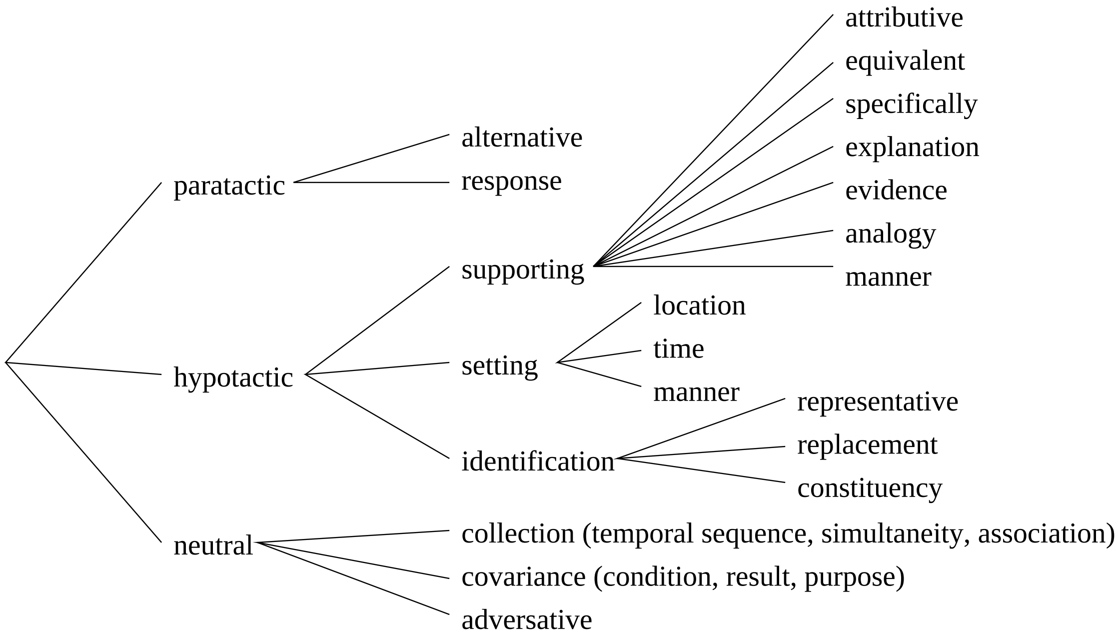

Figure 2.4: Grimes' Taxonomy of Relations

Although their taxonomies are quite different, the theories of Grimes and Longacre have much in common. For one thing, relations are defined independently of surface linguistic phenomena, in terms of the underlying semantics of the clauses which express them—they seem essentially to hold between *propositions*.

A second point of similarity between the two theories is the ability of relations to hold within clauses, as well as between them. Both theories, perhaps because they look beneath the surface structures in text, forfeit a solid conception of the 'text units' between which relations are construed to hold. For instance, Longacre sees an EFFICIENT CAUSE relation in the single clause *Stephen made Kathleen do it*; Grimes sees an ADVERSATIVE relation in the clause *I'd rather have coffee than tea*. It is not clear how, if at all, the semantic units being related in these examples map onto distinct, neighbouring sections of text. It is not even clear that the units being related are whole propositions. In such a situation, a theory of relations starts to encroach on a theory of the semantics of clauses; it becomes hard to see how the work is to be divided up betweeen the two types of theory.

Finally, while both accounts envisage relations applying between large sections of text such as groups of sentences, paragraphs and so on, neither of them has much to say about these higher levels—they both focus on the mechanisms of clause combination within complex sentences. For a fuller account of coherence, relations between sentence complexes and paragraphs must also be investigated.

2.4.2 Cohesive Relations

An alternative conception of relations downplays the importance of the 'deep structure' of text, concentrating on an analysis of the resources available in a language to signal relations explicitly in the surface structure. This approach is adopted by Halliday and Hasan (1976), and also (to some extent) by Martin (1992).

Halliday and Hasan's Conjunctive Relations

Halliday and Hasan (1976) set out to describe the **cohesive** resources in a language; that is, the resources available for linking sentences together in text. Cohesive relations are semantic in nature, and apply both within sentences and between them. However, when found inside sentences, they interact with a theory of sentence **structure** which is outside the scope of Halliday and Hasan's investigation; they are only studied in their 'pure' form, between whole sentences.

There are several types of cohesive relation: **reference**, **substitution**, **ellipsis**, **lexical cohesion**, and **conjunction**. The first four of these can be thought of as dependency structures, where one portion of a text can only be interpreted by reference to something else, either in the preceding or following portions of text or in the environment in which the text is found. Conjunctive relations, on the other hand, are not instructions about how to interpret a portion of text, but a specification of the way in which two already interpretable portions of text are to be linked together. These are the correlates of coherence relations.

Halliday and Hasan's study is primarily of **conjunctive elements** in a language; i.e. its resources for signalling conjunctive relations. Essentially, what is provided is a classification of sentence conjunctions. Several commentators have criticised their reliance on surface features in text for an account of coherence, arguing that texts can be coherent without any explicit signals. However, while a surface analysis certainly dominates in Halliday and Hasan, relations are not tied irrevocably to surface structure:

> it is the underlying semantic relation... that actually has the cohesive power. This explains how it is that we are often prepared to recognise the presence of a relation... even when it is not expressed overtly at all. We are prepared to supply it for ourselves, and thus to assume that there is cohesion even though it has not been explicitly demonstrated.
>
> *Halliday and Hasan (1976), p229*

Having said that, the idea of implicit relations is not discussed at length by Halliday and Hasan, and the mapping between conjunctive elements and 'underlying relations' remains unclear.

The actual classification of conjunctive phrases involves a fourfold distinction between ADDITIVE, ADVERSATIVE, CAUSAL and TEMPORAL relations. The ADVERSATIVE relations cover part of the ground covered by Longacre's FRUSTRATED relations; but here they are not factored out as an orthogonal dimension. An alternative orthogonal distinction is proposed instead, between INTERNAL and EXTERNAL relations. EXTERNAL relations hold between things referred to in a text, while INTERNAL relations hold between elements which are constitutive of the text itself—things such as speech acts. To give a canonical example, consider Texts 2.3 and 2.4:

(2.3) First he switched on the light. Next he inserted the key into the lock.

(2.4) First he was unable to stand upright. Next he was incapable of inserting the key into the lock.

Both texts contain TEMPORAL relations. But in 2.3 the relation is between two events in the world (and hence EXTERNAL); while in 2.4 it is between two *speech acts*, 'I assert that he was unable to stand upright' and 'I assert that he was incapable of inserting the key into the lock'[5] (and is hence INTERNAL). The distinction between INTERNAL and EXTERNAL relations is another one which is echoed in many subsequent relational theories.

The top levels of Halliday and Hasan's taxonomy are given (in slightly simplified form) in Figure 2.5.

By concentrating on an analysis of surface features in text, Halliday and Hasan's theory avoids some of the problems associated with Grimes' and Longacre's accounts. Conjunctive elements are concrete linguistic entities, and it is at least clear whereabouts in a text they are present (even if it is not so easy to decide exactly which units they are linking). But Halliday and Hasan admit that a complete account of text must make reference to unmarked relations. And it is not enough just to recognise the existence of such relations—they must be properly described and individuated.

---

[5] The interpretation in this case is presumably that the writer is enumerating two points in an argument.

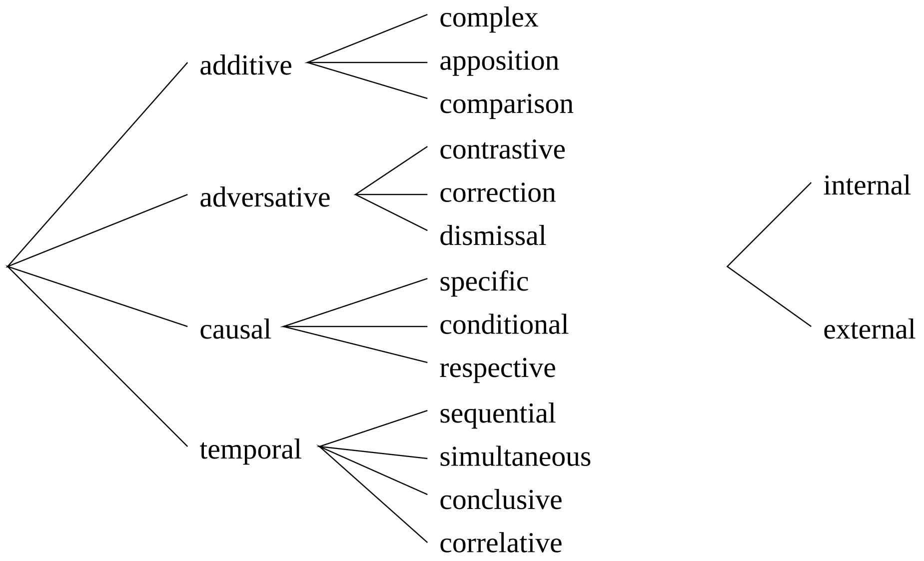

Figure 2.5: The Top Levels of Halliday and Hasan's Taxonomy of Relations

On a separate note, the decision not to consider the relations within complex sentences seems overly cautious. Admittedly, it avoids any possible confusion between a theory of clause semantics and a theory of discourse structure; however, there are so many similarities between clause-combining and sentence-combining phrases that generalisations will surely be lost if only the latter group is studied. Consider the phrases *although* and *however*—the former links clauses, the latter links sentences; but they have very similar functions. Why should they not be accounted for in the same theory?

It should be clear by now that the demands for structural simplicity in a theory and for a complete account of the underlying relations in a text pull in opposite directions. A compromise is needed between 'deep structure' accounts (which tend to fudge the dividing line between theories of clause semantics and discourse structure) and surface accounts (which often leave important features of a text unexplained). It remains to be seen whether a theory can be developed which avoids both of these problems.

Martin's Relations

Martin's (1992) theory of relations follows Halliday and Hasan in its emphasis on explicit linguistic conjunctions—in essence, it provides an alternative taxonomy of connectives. However, it expands on their theory in two respects. Firstly, it gives a better account of the relationship between 'deep' and 'surface' relations: basically, an **implicit relation** exists at a given point in a text if a conjunctive phrase can be inserted at that point.

> As a test for the presence of an implicit connection it can be required that the connection could have been explicit...
>
> *Martin (1992), p184*

This test gives a way not only for identifying the implicit relations in a text, but also for classifying them. The taxonomy of 'underlying relations' will basically mirror that of the devices for signalling them. The test runs into problems with some categories of relations (in particular with the INTERNAL and ADDITIVE relations), but it is nevertheless an improvement on Halliday and Hasan's account.

A second departure from Halliday and Hasan's theory is the decision to look within the sentence for methods of signalling relations. This expands the range of the analysis and enables more generalisations to be made. It also manages to avoid the danger of encroaching on an account of clause-internal semantics; although some clause-internal correlates of conjunctions are discussed (verbs such as *parallelled* and *precluded*, for instance), these are kept quite separate from the class of cohesive conjunctions.

The top levels of Martin's taxonomy of relations are given in Figure 2.6. The tax-

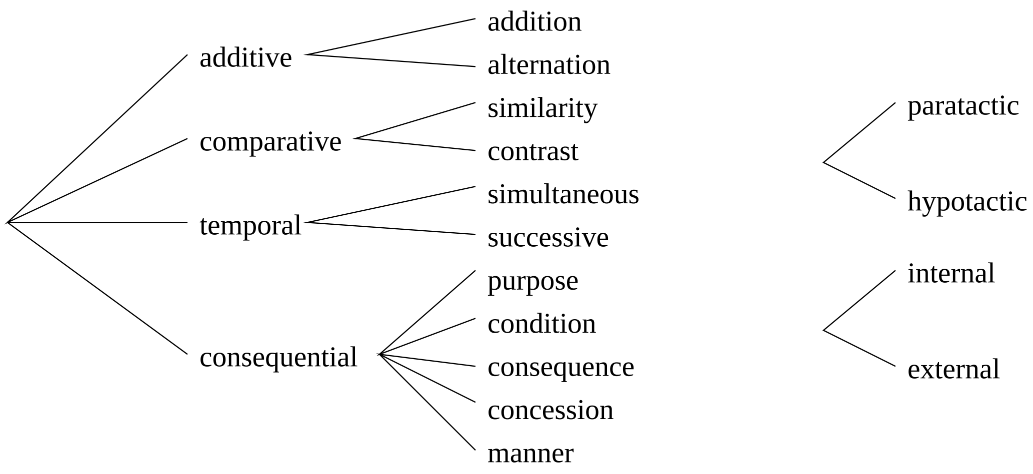

Figure 2.6: The Top Levels of Martin's Taxonomy of Relations

onomy is designed as a **systemic network**, framed in terms of the choices available to the writer. The writer begins by making high-level decisions about the nature of the relation, and these generate progressively lower-level choices until a decision about a particular conjunction can be made. The central division into ADDITIVE, COMPARATIVE, TEMPORAL and CONSEQUENTIAL relations is supplemented by orthogonal distinctions between INTERNAL and EXTERNAL (taken from Halliday and Hasan) and between PARATACTIC and HYPOTACTIC (reminiscent of Grimes's taxonomy). At lower levels, an extremely subtle classification is effected, which distinguishes around 100 different types of conjunction.

2.4.3 Computational Theories of Relations

In the 1980s, computational linguists became interested in coherence relations as potentially useful constructs in discourse processing applications. A number of distinctively 'computational' theories of relations appeared, characterised by the use of A.I. concepts such as knowledge representation, planning and recursion. Three of these theories will be outlined in this section.

Hobbs' Theory

Hobbs' theory (1985) emphasises the amount of world knowledge that is required to interpret discourse. He considers the following text by way of illustration:

(2.5) John took a book from the shelf. He turned to the index.

It is clear to the reader that *the index* referred to in the second clause is that of the book which John has just taken; but to make this inference automatically requires a great deal of knowledge about what people do with books, what indices are, and so on. The relations Hobbs proposes are defined in terms of the different kinds of inferences which the reader needs to draw in order to make sense of a text.

Hobbs identifies four types of inference, and accordingly, four categories of coherence relations. He notes firstly that a discourse can be coherent because it talks about coherent events in the world; events such that if one is known, the other one can be inferred given appropriate background knowledge. Two portions of text describing two such events are said to be linked by an OCCASION relation. This relation subdivides into relations like CAUSE and ENABLEMENT.

Secondly, discourse coherence can be due to the fact that the speaker has some rational structure of goals for producing a discourse. Sometimes these goals are referred to fairly explicitly in the text:

(2.6) Did you bring your car today? *Mine is at the garage.*

Hobbs accounts for such references with the class of EVALUATION relations, which associate some portion of discourse with a piece of 'metatalk' about why that portion is in the discourse.

Thirdly, a discourse will only be coherent if what the speaker says can be linked to what the hearer already knows. This idea provides the basis for the BACKGROUND and EXPLANATION relations, which perform this function.

Fourthly, the hearer of a discourse has only limited processing resources: therefore a coherent discourse makes explicit certain inferences which, given greater resources, he could be expected to make for himself. EXPANSION, the fourth class of coherence relation, links such inferences: two clear examples are PARALLEL and CONTRAST.

The set of relations in Hobbs' theory is given in Figure 2.7.

Hobbs' appreciation of the need for a knowledge-intensive approach to discourse interpretation is characteristic of a computational theory. Also characteristic is a well-articulated account of the compositional nature of relations:

> When two segments of discourse are discovered to be linked by some coherence relation, we can consider the two together as a single segment of

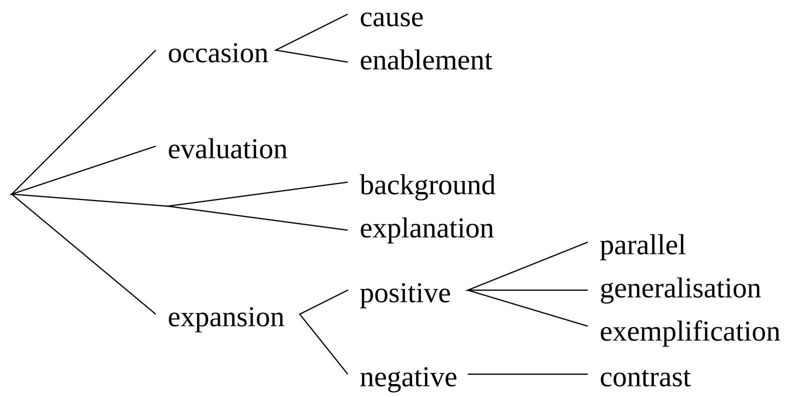

Figure 2.7: The Relations in Hobbs' Theory

> discourse. By recognising coherence relations between segments, we can thus build up recursively a structure for the discourse as a whole.
>
> *Hobbs (1985), p22*

The conception of relations as recursive mechanisms for building up large sections of text has two useful consequences. Firstly, it helps to give substance to the *structural theory* of text called for in Section 2.2.3. This theory, it will be recalled, is required to specify for any text whereabouts relations are expected to be found; Hobbs' suggestion is basically that a tree structure of relations must exist in a text for it to be coherent.

Secondly, the recursive conception of relations suggests for them a *procedural* role in constructing large sections of text. Recursively defined constructs are well suited, for instance, to hierarchical planning formalisms. And thinking about relations in the context of a planning paradigm also suggests how they could be grounded in a *psychological* theory of text processing—at one point Hobbs likens relations to 'text-building strategies', used by the writer to facilitate the task of the reader. This idea which will be taken up in much more detail in the next chapter.

Grosz and Sidner's Relations

Grosz and Sidner's (1986) theory also features recursively defined relations. In this account, **discourse segments** (DSs) are the principal units of structure, and relations hold between these to form larger DSs. However, the primitives used to define relations are different from those of Hobbs: they make reference solely to the **intentions** a writer has in creating a text. Relations actually apply between **discourse segment purposes** (DSPs); an assumption is made that a single overriding intention can be specified for each segment, and it is these intentions which are connected by relations. The fundamental metaphor is of a text embodying the execution of a **plan** pursued by the writer.[6]

[6] Note that although Grosz and Sidner frequently use examples from task-oriented dialogues, they take care in such cases to distinguish the plan required to carry out the task from the plan required to create the text. (See Litman and Allen (1990) for further discussion of discourse plans and domain-level plans.)

Using intentions in relation definitions follows quite naturally from thinking of them in the context of a recursive planning paradigm. Plans are produced to achieve user goals (or in the present case, writer goals); and do so by decomposing a principal goal into a hierarchy of subgoals. What is more, it is easy to see how the same intention can 'underlie' text spans of very different sizes. For instance, the same goal to convince the reader of some fact might require an extended argument for one reader, but a simple sentence for another reader who is already disposed to believe it. Note that it is much harder to claim that two differently-sized texts contain the same 'propositional content'. Intentions therefore seem to provide an ideal way of defining relations which must hold between spans of very different sizes.

There are only two relations in Grosz and Sidner's theory: DOMINANCE and SATISFACTION-PRECEDENCE (see Figure 2.8). Since these are the first intentionally defined relations

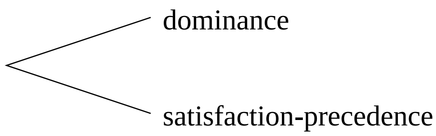

Figure 2.8: The Relations in Grosz and Sidner's Theory

we have come across, it is worth spelling them out in detail:

- DSP2 **dominates** DSP1 (and DSP1 **contributes to** DSP2) if the satisfaction of DSP1 is intended to provide part of the satisfaction of DSP2.
- DSP1 **satisfaction-precedes** DSP2 if both purposes are dominated by some other purpose DSP3, but in order to satisfy DSP3, DSP1 has to be satisfied *before* DSP2.

For example, Text 2.7 contains a DOMINANCE relation:

(2.7) Television is bad for children. They grow up on a steady diet of violence and advertising.

while Text 2.8 (second sentence) contains one of SATISFACTION-PRECEDENCE:

(2.8) Try out the gun by firing off a few rounds. First, release the safety catch; then squeeze the trigger gently.

As well as defining their relations differently, Grosz and Sidner also adopt a slightly different conception of compositionality to that proposed by Hobbs. While Hobbs sees a relation between two adjacent spans as forming a new composite span, Grosz and Sidner's composite discourse segments *include* the segments which they dominate. Thus in Text 2.7, the subordinate span is the second sentence, and the dominant span is the first and second sentences together.

A final attractive feature of Grosz and Sidner's theory is its account of the interaction between relations and focus. Associated with every discourse segment is a **focus space**, and at every point in a text a **focus stack** is given which models the reader's focus of attention as the discourse proceeds (elements at the top of the stack being 'more salient' to the reader than elements lower down). The metaphor of a stack is another import from computational theories. Its pushes and pops are determined by the DOMINANCE relations in the text: if the segment S_2 dominates a sub-segment S_1, then moving into S_1 causes the focus space associated with S_1 to be pushed onto the stack, and leaving S_1 causes it to be popped off the stack.

Grosz and Sidner's theory is the first to look in detail at relations between larger sized units of text; indeed most of their examples are of high-level relations. In fact, the theory does not provide a very complete account of lower level relations such as those between single clauses or sentences. Even 'atomic' discourse segments are often larger than a single sentence. The rationale for these units is based on studies of how 'naive' subjects segment discourses (1975) and of how speech rate correlates with segment boundaries (1975). However, whether or not these studies are convincing, there seems no reason in principle why the analysis of text units should not descend at least to the level of individual sentences.

Rhetorical Structure Theory

We turn now to the third computational theory to be discussed—rhetorical structure theory (RST). This theory, developed mainly by William Mann and Sandra Thompson, is presented in a number of papers; in this thesis I shall for the most part be drawing on the account in Mann and Thompson (1988), which is the most comprehensive.

The central constructs in RST are **rhetorical relations**. Text coherence is attributed principally to the presence of these relations; unlike Grosz and Sidner, Mann and Thompson do not envisage an important role for other constructs such as focus. The claim is that the relations in RST suffice to analyse 'the vast majority' of English texts; exceptions are only made for very unusual texts like poems and legal documents.

Rhetorical relations are defined functionally, in terms of the effect the writer intends to achieve by presenting two text spans side by side. In this respect, they resemble Grosz and Sidner's relations. However, there are also several differences between the two types of relation.

Firstly, RST relations do in fact make some reference to the propositional content of spans, as well as to the intentions of the writer in putting them forward. For instance, the MOTIVATION relation specifies that one of the spans 'presents' an action to be performed by the reader; the SEQUENCE relation specifies that a succession relationship must exist between the related spans. RST relations are in fact defined using five **fields**—only one of these explicitly represents the effect of the relation; the others represent the various different constraints that must be satisfied in order to achieve this effect, and these are specified using a mixture of propositional and intentional language.

Secondly, Mann and Thompson go out of their way to rule out a connection between the set of relations and the linguistic devices used to signal them. This goes beyond the claim that the relations in a text need not be signalled—they further suggest that

> some types of rhetorical relations have *no* corresponding conjunctive signals.
>
> *Mann and Thompson (1988), p45 (my italics)*

In this, their theory differs from Grosz and Sidner's (and many others besides), in which at least an informal link is made between underlying relations and the linguistic devices for marking them.

A third novel feature of RST is its concept of **nuclearity**. As well as representing the relationship between two text spans, rhetorical relations also convey information about which span is more central to the writer's purposes. The **nucleus** is the more central span, and the **satellite** is the less central one.[7] Mann and Thompson contend that the majority of text is structured using nucleus-satellite relations; although some relations—termed **multinuclear**—do not exhibit it. (There are two multinuclear relations: SEQUENCE and CONTRAST.)

The nucleus-satellite distinction is in some ways comparable to the PARATACTIC-HYPOTACTIC distinction of Grimes and others. But while these are expressed in semantic or even syntactic terms, RST's definition is functional, based on the idea that a writer has more important and less important goals when she sets out to create a text. Nucleus-satellite relations are in fact more reminiscent of Grosz and Sidner's class of DOMINANCE relations. But even here there is a difference: in Grosz and Sidner's model it is hard to talk about the purpose of the dominant span being 'more central' to the writer than that of the subordinate span, because the former purpose actually *includes* the latter.

RST provides a set of around 23 rhetorical relations. The numbers vary slightly from paper to paper, but the central core of relations as presented in Mann and Thompson (1988) are given in Figure 2.9. The top-level distinction in this taxonomy is between SUBJECT-MATTER and PRESENTATIONAL relations. SUBJECT-MATTER relations have as their effect that the reader *recognize* the relation in question; while PRESENTATIONAL relations have as their effect to *increase some inclination* in the reader. Thus SEQUENCE is a SUBJECT-MATTER relation (its effect is that the reader recognize that the two related spans present events occurring in sequence) and MOTIVATION is PRESENTATIONAL (its effect is to increase the reader's motivation to perform the action presented in the nucleus span). To some extent, this distinction mirrors Halliday and Hasan's distinction between INTERNAL and EXTERNAL relations. But again, the similarity is far from complete.

Like the other computational theories of relations, RST has a strong structural account of text. It begins with an independent definition of 'text span'—for Mann and Thompson, the size of the atomic units of text analysis is arbitrary, but they should have independent functional integrity. The clause is selected as the minimal unit of organisation; thus text spans are clauses, or larger units composed of clauses. Unlike Grosz and Sidner, relations must hold between non-overlapping text spans. (An ex-

[7] A test for nuclearity is to delete the satellite span of a given relation and see if the resulting text still makes sense. The prediction is that without the nucleus, the significance of the material in the satellite will not be apparent, while the nucleus should be able to stand by itself. This test is rather blunt, and not completely reliable, but it is useful at least in giving some substance to Mann and Thompson's notion of nuclearity.

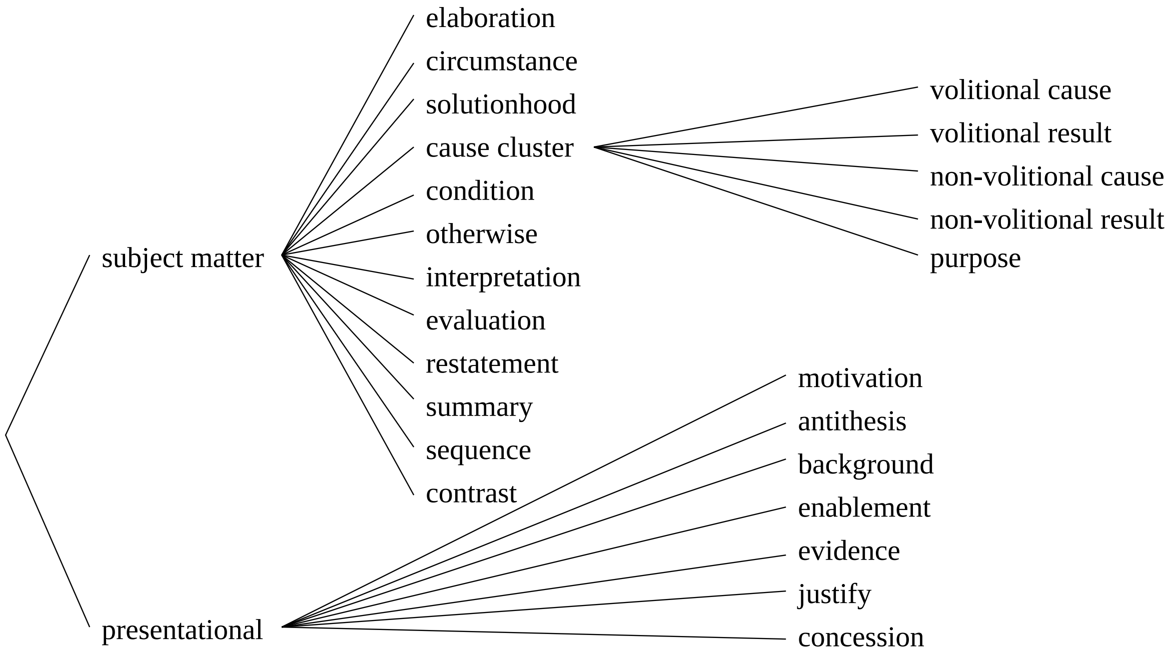

Figure 2.9: Mann and Thompson's Relations

ception to this rule is made for non-restrictive relative clauses: relations are permitted to hold between a matrix clause and a subordinate clause.)

In RST, relations are not mapped directly onto texts; they are fitted onto structures called **schema applications**, and these in turn are fitted to text. Schema applications are derived from simpler structures called **schemas** (see Figure 2.10).

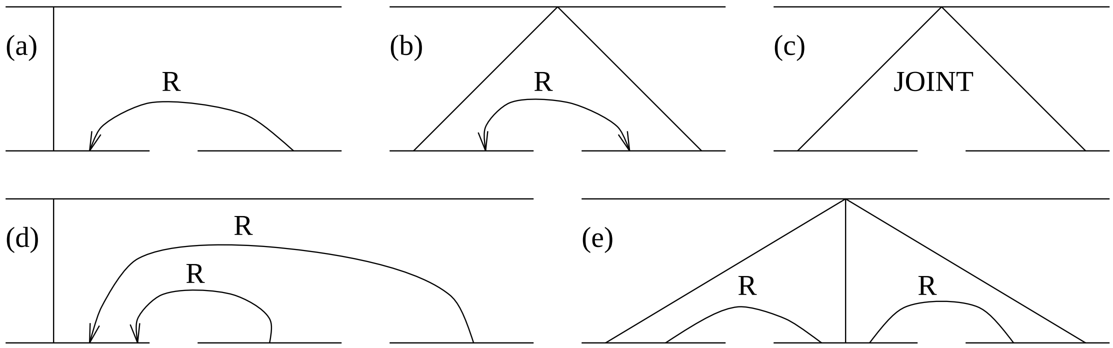

Figure 2.10: The Types of Schema in RST

In this diagram, horizontal lines depict text spans, the labelled lines depict relations between spans, nuclei are picked out by the vertical lines (they are diagonal for multinuclear relations), and all other spans are satellites. From these structures, schema applications are formed, by rearranging the spans in any order and by duplicating spans any number of times. (For the schemas with satellites, only the satellite spans can be duplicated.) Relations are then fitted to the schema applications thus formed—relations which take a nucleus and a satellite are fitted to the single or dual relation schema applications, and the specialised CONTRAST and SEQUENCE relations are fitted to the individual schemas (b) and (e) respectively. The 'joint' schema is for linking pieces of text which are not linked by any RST relations, and is essentially used for representing lists.

A **rhetorical structure tree** is a hierarchical system of schema applications. A schema application links a number of consecutive spans, and creates a complex span which can in turn be linked by a higher level schema application. This enables tree structures to be built—it is a central claim of RST that the structure of every coherent discourse can be described by a single rhetorical structure tree, whose top schema application creates a span encompassing the whole discourse.

An interesting aspect of RST's structural account is its use of the multisatellite schema (illustrated in Figure 2.10 (d)). Note that in this case, there is no relation linking the second and third spans; although they are adjacent, they are only linked indirectly, by virtue of both being related to the first span. (They might, for instance, be two separate JUSTIFICATIONs for a claim expressed in the first span.) Other theorists have chosen to model such structures with a new type of relation; for instance, Halliday and Hasan's INTERNAL SEQUENCE accomplishes exactly this function. The respective merits of the two approaches will be discussed in more detail later in the thesis, in Section 6.2.4.

RST has proved a very influential theory amongst computational linguists, as the next section will attest. Its popularity is perhaps best attributed to a combination of features: the emphasis on a functional conception of relations; the carefully presented set of relation definitions; the simply stated structural theory. It is doubtful whether anyone believes the claims made in RST—but at least it is clearly enough expressed for people to be able to frame their objections to it.

2.5 The Uses of Relations in Automatic Text Generation

In the last few years, coherence relations have begun to feature prominently in studies of natural language generation (NLG). Successive generation workshops (Dale *et al* (1990), McKeown *et al* (1990), Dale *et al* (1992), Zock *et al* (1994), McDonald *et al* (1994)) contain accounts of relations being implemented in generation systems. And two recent workshops, in Maratea (Scott and Hovy (1993)) and Columbus (Rambow (1993)) have been largely given over to a discussion of relations.

The two most popular theories for implementation have been Mann and Thompson's RST and Martin's systemic model. However, as in the theoretical field, no standard set of relations has emerged—the relations in both of these theories have been considerably and variously adapted for their new procedural role. In the following review, I will again be emphasising the differences between the various conceptions of relations which have emerged in the literature.

2.5.1 Hovy's Implementation: Relations as Planning Operators

The original relation-based text structurer was Hovy's (1988) implementation of RST. This system develops the idea of relations as planning operators, defining them in terms of **preconditions** and **postconditions**. Both preconditions and postconditions are expressed in terms of communicative goals—the idea is that a communicative goal can be satisfied in simple cases by producing a single clause, and in other cases by 'applying' an RST relation, whose nucleus and satellite are each characterised in terms of simpler communicative goals. For example, the definition of the operator for the SEQUENCE relation specifies a complex goal as its 'result' (or postcondition), and simpler goals as its 'requirements' (or preconditions):

```
            ___SEQUENCE___
Results:
        ((BMB SPEAKER HEARER (SEQUENCE-OF ?PART ?NEXT)))
Nucleus requirements/subgoals:
        ((BMB SPEAKER HEARER (TOPIC ?PART)))
Satellite requirements/subgoals:
        ((BMB SPEAKER HEARER (TOPIC ?NEXT)))
```

('`BMB SPEAKER HEARER X`' can be read 'achieve the state in which the speaker and the hearer mutually believe that X'. '`TOPIC X`' is satisfied by associating the clause which translates X with the nucleus or satellite of the relation.)

Of course, an operator can only be applied if the appropriate semantic relation holds between its nucleus and satellite. So another part of the precondition for SEQUENCE is as follows:

```
Nucleus+satellite requirements/subgoals:
        ((NEXT-ACTION ?PART ?NEXT))
```

('`NEXT-ACTION`' is a semantic predicate which can be checked by inspection of a database.)

Hovy's structurer takes as input one or more communicative goals, and a set of clause-sized 'input entities' which contain the material to be generated. On the basis of these, a simple rhetorical structure tree is formed, consisting of one relation, one nucleus and one satellite. The tree is then progressively expanded by a process of adjunction: one of its leaf nodes is selected, and replaced by a new relation whose nucleus is the original leaf node and whose satellite is one of the other input entities to be incorporated. The process continues until all the input entities have been used up.

To produce a piece of text, the finished tree is traversed left-to-right, and the leaves are fed to the sentence generator PENMAN (Mann and Matthiessen (1983)). The sentences thus produced are linked using prototypical keywords; for instance, keywords for the SEQUENCE relation include *then* and *next*. The resulting text should in theory satisfy the original communicative goal, include all the material in the input entities, and in addition, be coherent.

Hovy's system is the first to exploit a powerful new technique. Although it is rudimentary, it demonstrates how the power of a hierarchical planning system can be of real use in text structuring problems—the ability of a planner to deal with complex constraints and to consider all possible solutions to a given problem make it well suited to such tasks. And using relations as operators seems to provide just the right amount of flexibility for the planner: if relations do indeed underlie the phenomenon of coherence, then the space of texts to be searched will include all and only those texts which are coherent. However, Hovy's system is only a start; a number of problems can be identified in its operation.

For one thing, in constraining the set of texts produced from a single input, the planner makes use of the notion of **possible growth points**—a notion not apparent in any of the RST literature. The idea is that the nucleus and satellite of a given relation can only be expanded in certain ways; for instance, the nucleus of SEQUENCE can expand into CIRCUMSTANCE, ELABORATION-ATTRIBUTE or PURPOSE and its 'satellite' can expand into ELABORATION-ATTRIBUTE, ELABORATION-DETAILS or another SEQUENCE. There seems no justification for these constraints—texts which violate them can easily be thought up.

In other respects, the generation process appears underconstrained. Constraints to do with the **focus** of the text being built are not addressed, and the texts produced often seem to violate these, switching in an unprincipled way from subject to subject. A system implementing both relational and focus-based constraints is presented by Hovy and McCoy (1989), which resolves some of these problems.

Perhaps the most serious problem with Hovy's system is the way it characterises the text structuring problem in the first place, in terms of a communicative goal and a set of clause-sized input entities. Beginning from these elements begs the question—how are they themselves determined? It is not realistic to think of text structuring happening after the elements of content to be expressed are decided; for instance, we might first decide we need to justify a claim, and only then search for the material to use in the justification.

2.5.2 Moore and Paris: Relations for Guiding Content Selection in Dialogue

The planner developed by Moore and Paris (1989) (see also Moore (1989), Moore and Paris (1993)) addresses the problem of choosing the material to be generated as well as that of deciding how the material should be structured.

> We believe that the tasks of choosing what to say and selecting a strategy for saying it cannot be divided.
>
> *Moore (1989), p67*

Their planner forms part of the interface for an expert system which gives advice in various different domains, and reacts to users' followup questions. It is thus able to draw on the domain-specific information needed to decide on what advice to give.

The technique for interleaving content selection and text structuring tasks centres around the representation of the **intentions** of the text to be produced. For Moore and Paris, it is important that the intention behind every part of the text be represented in the text plan, so that if the plan fails in some respect and the reader has to ask for clarification, it can be determined what went wrong and how to put it right. In Hovy's structurer, the intention underlying the whole text is represented—it is given in the communicative goal specified at input—but the motivation for including additional input entities is not. If some part of the text is unsuccessful and the reader queries it (for instance, by saying 'I don't understand that last point'), what is required is a new attempt to achieve the intended effect of that portion of text. However, if its intended effect is not explicitly represented, this will not be possible: knowing the relation which links the text segment to the rest of the discourse is not sufficient, because frequently what is called for is a new explanatory strategy involving other relations. For instance, if the reader does not understand a concept by use of an ANALOGY, a CONTRAST might be given, although the intention (to explain the concept) is the same in both cases.

Thus it is claimed that there is no simple one-to-one mapping between the relations in a text and the intentions that underlie them. In one sense this is confusing: both Hovy and Moore and Paris work with RST relations, which are defined precisely in terms of the effects they are intended to achieve. The point becomes clearer when it is realised that the intentions behind a span of text can be described at many different levels. Consider Text 2.9:

(2.9) The window's open.

Underlying this text are any (or all) of the following:

- The intention to say "The window is open".
- The intention to convey to the hearer that the window is open.
- The intention to convey to the hearer that the room is cold.
- The intention to make the hearer close the window.
- The intention that the room warm up.

Moore and Paris' point is that the intentions used to define RST relations are often not abstract enough to allow effective backtracking in the case of a local failure.

In a later paper (Moore and Paris (1993)) this point is refined. They claim that for the PRESENTATIONAL relations in RST, the effects specified in the definitions are deemed adequate for generating alternative strategies. (The effects of all these relations, it will be recalled, are to 'increase some inclination in the reader'.) For the SUBJECT-MATTER relations, whose effects are simply that the reader recognise the relation in question, a specification of deeper level intentions is demanded to avoid the backtracking problems outlined above.

Clearly, the intentions used in defining some RST relations are insufficient for some purposes. However, the level of abstraction at which intentions must be specified might be to some extent genre-dependent—in an explanatory dialogue, we can expect PRESENTATIONAL relations to predominate. But in other genres, such as narratives, it is difficult to specify a more abstract goal than the simple representation of content. To give an example: the CAUSE relation can be used as part of an argument, as in Text 2.10, or in a story, as in Text 2.11:

(2.10) Bill must have been absent, because he was sick.

(2.11) Bill was absent from school that day, because he was sick.

The argument in 2.10 hinges on the fact that being sick typically causes absence. But expressing the causal relation is not the writer's primary objective; if the reader is unconvinced, an alternative strategy (perhaps to report that no-one saw him at school) could be attempted. However, if the causal relation is presented as part of a story being told to the reader, as in 2.11, it is hard to identify a goal beyond that of simply relating the cause of Bill's absence. To take another example—the goal behind a SEQUENCE relation in a story is likely to be just to tell the reader what happened next.

2.5.3 Systems using Multiple Levels of Analysis for Relations

A further extension of the notion of intentions is suggested in Moore and Pollack (1992). Here, it is proposed that every discourse should be analysed on two levels, firstly in terms of its intentional (presentational) relations and secondly in terms of its informational (subject-matter) relations. The arguments given are similar to those in the previous section. While the case of purely narrative texts seems to argue against identifying an intentional and an informational relation in every case, there certainly seems some merit in factoring out the two components.

A similar decomposition of relations is proposed by Maier and Hovy (1991) and Maier (1993), and implemented in Hovy *et al* (1992). In these studies, three types of relation are identified, mirroring Halliday and Hasan's (1976) classification of the three functions of language. IDEATIONAL and INTERPERSONAL relations correspond roughly to RST's SUBJECT-MATTER and PRESENTATIONAL relations respectively. TEXTUAL relations serve to 'structure text'—the archetypal examples seem to be the signallers of Halliday and Hasan's INTERNAL SEQUENCE (*firstly*, *secondly*, *moreover* and so on). Maier and Hovy allow that more than one type of relation can hold between two spans in a text; however, there is no need for every type of relation to be present at every point. Thus they allow that descriptive texts are relatively poor in INTERPERSONAL relations.

However, some aspects of this classification make for confusion. For one thing, the class of TEXTUAL relations seems rather different from the other two. It contains not only 'linearising' relations such as those marked by *firstly* etc, but also relations such as 'relnamecomparative', 'relnameconcessive' and 'relnametemporal', which seem to perform much more than a purely ordering function. It is plausible to claim that spans linked by *moreover* are not related by any IDEATIONAL or INTERPERSONAL relations, but merely by the fact that they appear as consecutive items in a text; but these other relations seem quite clearly to convey either propositional or intentional content. There thus seems to be considerable redundancy within the system of relations as a whole.

A second objection to the classification comes from Moore and Pollack (1992): they object to the large number of possible combinations of relations that results from the assumption of three independent relation types. However, this objection seems less telling; Moore and Pollack's use of two independent levels of analysis gives rise to a similar combinatorial problem. Clearly what both theories need to provide is a specification of any contingent *constraints* between the independent components of the analysis. In fact, both theories do provide some initial specifications of such constraints (see Moore and Paris (1993) p670, Maier (1991) pp42–43); but neither of them is yet complete.

2.5.4 Relations in Text Realisation

Attention has also been focused on RST relations as guides to the linguistic marking of text structure.

Although Mann and Thompson strenuously avoid any reference to surface syntactic phenomena in their relation definitions, many correspondences can in fact be found. Several techniques for signalling RST relations are presented by Scott and de Souza (1990), who outline a number of heuristics for guiding the textual realisation process, informed by a combination of stylistics and psycholinguistics. The heuristics, motivated individually, include using embedded relative clauses to signal ELABORATION, and using paratactic and hypotactic coordinators to signal MULTINUCLEAR and NUCLEUS-SATELLITE relations respectively. The utility of such heuristics has given many researchers cause to question Mann and Thompson's policy of downplaying the link between relations and surface linguistic structures. It should certainly not be assumed that any one-to-one mapping between relations and linguistic forms will be found; as Scott and Paris (1995) point out, it is often impossible to predict the structure of a text on the basis of its surface characteristics alone. However, the development and refinement of sophisticated heuristics for marking relations continues to prove useful—see in particular the system proposed by Moser and Moore (1995).

On another front, Rösner and Stede (1992, 1992) and Knott (1991) have suggested ways of making the realisation of a relation sensitive to its hierarchical position in a rhetorical structure tree. Different strategies are appropriate for different levels; of particular interest are high-level marking techniques which involve whole clauses, such as *there are two reasons for this* or *this happened as follows*. At the other end of the spectrum, researchers (notably Vander Linden *et al* (1992)) have investigated ways of signalling RST relations within single clauses. Available techniques include nominalisation and the use of adverbial phrases; for instance the PURPOSE relation can be marked as follows:

(2.12) Follow the steps in the illustration below *for desk installation.*

A final issue of importance is the decision about when to mark a relation and when not to. If relations are always signalled using the most specific available conjunction, the texts which result are often stylistically awkward, making explicit information which is easily inferrable by the reader. Oberlander and Lascarides (1991; see also Lascarides and Oberlander, 1992) introduce the notion of **laconic text** to represent text in which the only relations to be marked are those not inferrable from context and world knowledge. They propose an algorithm for generating such texts using a system of defeasible rules to represent the different types of knowledge that are needed.

2.5.5 Relations in Multilingual Generation Systems

A final use of coherence relations in generation systems is as a language-independent intermediate representation of text structure. Such a representation is particularly useful in multilingual generation systems—relations model the intentions underlying a text, and it seems plausible that these are invariant during translation. A modular approach to multilingual generation is then afforded, whereby it is only in the textual realisation of rhetorical structure trees that different strategies are required for different languages.

Several multilingual systems currently make use of relations in this way; among them Rösner and Stede (1992), Bateman et al (1993). However, there are also studies which call into question the invariance of relations across translations—see for example Delin *et al* (1994). The debate here seems to hinge on the level of abstraction at which relations are intended to represent the information conveyed by a text.

2.6 The Proliferation of Relations, and its Problems

We turn now to the central point to be made in this chapter—that the diversity amongst the many alternative sets of relations, as well as being confusing in its own right, is symptomatic of a deeper confusion about what it is about a text that relations are actually modelling.

I will begin in Section 2.6.1 by looking at the diversity of relations in computational applications. In Section 2.6.2 I will turn to the more serious differences that exist between relational theories, and make some suggestions about what might be responsible for them.

2.6.1 Differences between Generation Systems

All the systems reviewed in Section 2.5 draw principally on RST's set of relations. But despite this, there is a surprising amount of variability between researchers in the sets of relations they use. No two systems use exactly the same set of relations; and no system uses exactly the set of relations proposed in RST.

Departures from RST's original set are of many different types. RSTrelations have been subdivided—for instance, Rösner and Stede's STEP-SEQUENCE is a specialisation of the RST relation SEQUENCE. They have also been amalgamated to form new relations—for instance, Scott and de Souza combine the relations VOLITIONAL-CAUSE, NON-VOLITIONAL-CAUSE, VOLITIONAL-RESULT, NON-VOLITIONAL-RESULT and EVIDENCE, for the purposes of textual realisation. In other cases, relations seem to be defined orthogonally to those in RST: for instance, Hovy et al's (1992) ANALOGY covers some of the ground covered by the RST relations RESTATEMENT, EVALUATION and ELABORATION; but needless to say, these latter relations are not *always* analysable using ANALOGY. Finally, some new relations have appeared which have no obvious connection with RST relations at all—for example, Rösner and Stede define a relation called UNTIL, which is customised for instructional texts where a given action must be performed until a certain condition is met. In short, there are currently hundreds of relations in use—clearly, a significant departure from RST's original set of 23.

This proliferation is partly to be expected: its origin can be traced to a number of features of RST. For one thing, Mann and Thompson are themselves quite flexible about the set of relations in their theory; they are more concerned with establishing the rhetorical relation in general as a useful tool for text analysts.

> Relation definitions have the status of applications of the theory rather than elements of the theory. One might want to change or replace the definitions...such changes are to be expected and do not cross the definitional boundaries of RST.
>
> *Mann, Matthiessen and Thompson (1989), p 48*

The flexibility of RST is in fact held by some to be an advantage of the theory:

> Generally a new domain has dictated modifications to the inventory of relations, but this very adaptability is one of its most useful features.
>
> *Vander Linden et al (1992), p 184*

At the same time, giving theorists such a free hand to choose relations creates some serious problems. For one thing, it makes it hard to compare generation systems amongst each other. But more importantly, it tends to undermine the empirical content of the relation construct in general. The hypothesis that 'virtually any text can be analysed by representing its coherence relations' becomes much less strong if relations can be created whenever they are needed: it is hard to think what evidence could be found which could disprove it. Coherence relations at this point hardly seem to be saying any more than speech act theory; that we must take intentions into account when representing text. The extra claim in RST—that text is coherent by virtue of the relations between its intentions—is virtually unfalsifiable without a method for specifying what is to count as a relation in the first place.

Even incoherent texts can be analysed according to the relations between the intentions in their spans. For instance, the text in (2.13) seems incoherent at first sight:

(2.13) John broke his leg. I like plums.

Yet we could still define a relation which holds between the intentions underlying the spans in this text: perhaps we could call the relation INFORM-ACCIDENT-AND-MENTION-FRUIT. The relation might be multinuclear, with one nucleus intended to convey information about an accident, and the other intended to convey information about fruit. Clearly, we do not want to include these sorts of relations in any principled set of coherence relations.

To place the relational claim on a sounder empirical footing, we need to tighten the constraints on relationhood—for instance, by giving a fixed set of relations, or by giving rules for picking out a set of relations from some larger set of 'potential relations'.

2.6.2 Differences between Relational Theories

A diversity of relations exists not only amongst the various implementations of RST but between RST and the other relational theories. This should be clear from Section 2.4. The number of relations posited in a theory ranges from two to over a hundred; the primitives used for their definition are themselves very diverse; so too are the principles by which relations are organised into taxonomies. Naturally, some of the differences between theories are due to the fact that different theories have different aims—for instance, Grosz and Sidner's theory is geared principally towards modelling the pattern of anaphora in a text, rather than to providing a full set of resources for a text planner. And yet all these theories have, broadly speaking, a common objective—to model the coherence and structure of discourse. They are all apparently trying to explain the same phenomenon; thus it is odd that no consensus about a set of relations is emerging.

One reason for the many different sets of relations is perhaps that none of them seems quite right for expressing all the data. Many theorists acknowledge this themselves: for instance, Grimes (1975) admits that he is 'not completely satisfied with the basis of classification' for his HYPOTACTIC predicates; Martin (1992), in comparing his relations with those of Halliday and Hasan, mentions several areas where 'neither categorisation is completely satisfactory'. The difficulties in finding a suitable classification scheme suggest perhaps that the data to be accounted for is not well captured by a hierarchical taxonomy at all, and that some different formalism might be more appropriate. This idea will be developed in Chapter 5 and beyond.

However, the proliferation of relations also highlights a more fundamental problem, not just to do with the accuracy of relational theories, but again to do with their empirical status. It seems strange, from a theoretical point of view, that we should have a *choice* about which set of relations to use. Theoretical constructs are typically treated as corresponding to *real phenomena* underlying the data they describe, rather than as being purely synthetic: otherwise there seems little point in using the constructs at all. But relational theories tend to downplay this 'realist' conception of relations; consequently, little attention is paid to the question of what it is underlying a text that relations actually model. And yet it is a serious question. 'Coherence' is not just a label applied to an arbitrary group of texts by text analysts—coherent texts are actually produced, and appear in books, newspapers and so on, while incoherent texts (by and large) do not. There must be some reason for this, and if we are looking to relations for an account of coherence, we should expect them to make some reference to it.

In fact, the real situation is not quite as arbitrary as the preceding discussion would imply. There is still some degree of consensus between researchers about which relations to use. (For instance, nearly all of them identify concepts like causality and sequence as important.) But this very fact suggests that intuitions are at work which are not being acknowledged: while researchers rarely attempt justification for their choices beyond an adherence to the needs of descriptive adequacy, their choices are not as diverse as they might be given *only* this criterion.

2.7 Summary

This chapter has reviewed a large number of alternative theories of coherence relations, highlighting the diversity of the different sets of relations that have been proposed. The diversity of relations has of course been noted before—for instance, Hovy (1990) discusses it at length—and it is clear that agreement on a standard set of relations would be a considerable step forward for discourse theorists. Apart from anything else, there is a danger that unless a standard set begins to emerge, the research programme based on coherence relations might grind to a halt. If what one researcher calls a RESULT the next calls a REACTION, and the next an ELABORATION, we will be bound to ask whether there is any point in using such constructs at all.

Clearly what is needed is a standard set of relations. But perhaps more importantly, we need a way of *justifying* a standard set of relations, so that we can begin to argue that one set is more appropriate than another. It is to this latter question that we turn in the next chapter.

Chapter 3

Strategies for Motivating a Set of Relations

> Lacking a clear way to continue our description of the underlying conceptual system, we turned to analyses of the parts of it frozen by social convention into the English lexicon.
>
> *Miller & Johnson-Laird,* Language and Perception*, p697*

3.1 Introduction

In the previous chapter, a great many sets of coherence relations of different kinds were described, and the need was emphasised for a means of justifying one such set over the others. The justification should ideally perform several functions. It should provide an independent definition of relations, which makes clear *why it is* that relations underlie discourse coherence. It should also impose some kind of limit on the *size* of the set of relations, so that a theory explaining coherence in terms of relations is not unfalsifiable. It should also make clear how a given set of relations is *more appropriate than any other set* for the theoretical task it is to perform.

This chapter is given over to examining various strategies that have been proposed in the literature for justifying a set of relations. The first of these is to pick a set of relations that allows an adequate description of all the discourses that the theory purports to explain. The second strategy suggests that connective **cue phrases** such as *because* and *however* can be used to motivate a set of relations. A final strategy is based on the contention that relations model psychological constructs used in human text processing. Each of these strategies has advantages and shortcomings, which will be discussed. In the second half of the chapter, a new strategy for justifying a set of relations is proposed, which (it will be argued) combines the advantages of the above methods, and minimises their shortcomings. According to this strategy, cue phrases can be used as evidence for relations *precisely if* relations are considered as psychologically real entities. This is the central claim of the thesis: the rest of the chapter will be devoted to clarifying it, arguing for it, and defending it against various objections.

3.2 Choosing a 'Descriptively Adequate' Set of Relations

The most fundamental rationale for a set of coherence relations is that they enable an analysis of all the texts which the theory using them purports to account for. Some sets of relations are tailored specifically to particular registers of discourse (for instance, those in Rösner and Stede's (1992) planner, and in that of Vander Linden *et al* (1992) are just for processing instructional texts) while other sets (such as that of RST[1]) are claimed to be almost register-independent. In all cases, a claim is made that a given set of relations is sufficient to enable an analysis of a previously-specified set of texts. Clearly, any theory invoking relations must address the question of how well they can be made to fit texts, and so must make this claim in some way.

However, it remains to be determined what it means to say that a set of relations is 'sufficient' to analyse a text. As noted in Section 2.2.3, such a claim must be interpreted in the light of a theory of span structure, which specifies the places in a coherent text where relations are expected to be found. RST makes a rigorous prediction, holding that *each clause* in a text is linked to some other portion of it by a rhetorical relation; the same goes for each 'composite' span created by a schema application (except the top span, of course). Other theories impose lesser constraints: for instance, Grosz and Sidner's (1986) relations link 'discourse segments', typically composed of several clauses—which means that there are many clauses between which relations are not construed to apply. At the other end of the scale, accounts like Vander Linden *et al*'s (1992) suggest that relations can even exist *within single clauses*—although they do not yet specify exactly when such relations will be found.

How should we decide how densely a text must be filled with relations in order for it to be 'adequately described'? It seems that this question can only be addressed when we know more about what it signifies for two text segments to be connected by a relation. It makes no sense to set a criterion for 'descriptive adequacy' which cannot *itself* be further justified; except perhaps as a working hypothesis.[2] On the other hand, if we had a theory about text processing in which relations played a role, this might well be able to tell us where to *expect* to find relations in a text. Say the theory states that a reader links each paragraph to the previous one by means of relations. The notion of descriptive adequacy would be different for this theory than for one which claimed that the processing of individual clauses is mediated by relations.

There is a second problem with descriptive adequacy as the sole criterion for judging a set of relations. Many different sets of relations can be used to describe any given text: the subtlety of the distinctions between relations is not constrained. For instance, while Mann and Thompson choose to split 'causal' relations into five sepa-

---

[1] RST analyses are claimed for 'virtually every text' (Mann and Thompson (1988) p20), though exceptions are mentioned for some registers, such as legal documents and some kinds of poetry.

[2] This suggestion is at odds with Mann and Thompson's (1988) position. They claim that the decision about the size of text units to be related is 'arbitrary'—their only caveat being that the units should be defined in some theory-neutral way (p6).

rate groups (VOLITIONAL-CAUSE, NON-VOLITIONAL-CAUSE, VOLITIONAL-RESULT, NON-VOLITIONAL-RESULT, and PURPOSE), others have chosen a less fine-grained distinction: Scott and de Souza (1990) group all these relations together for the purposes of textual realisation. Vander Linden *et al* combine the VOLITIONAL-RESULT and NON-VOLITIONAL-RESULT categories to create a simple RESULT relation, but keep PURPOSE separate. Each of these models seems feasible as a 'descriptive' framework for text: it is equally feasible to split up Mann and Thompson's relations still further, creating relations like IMMEDIATE-VOLITIONAL-CAUSE, DISTANT-VOLITIONAL-CAUSE and so on. If we are looking to justify a set of relations, we need to have a way of deciding on an appropriate level of detail.

It could be argued that although a *tight* specification of the requisite level of detail is not feasible, some sort of approximation can still be arrived at. For example, it is possible to describe any text just using the relations 'CAUSAL' and 'NON-CAUSAL'; we might argue that such a description is a clear case of 'descriptive inadequacy'. But this is to miss the point: Grosz and Sidner use just two relations, but their relations are adequate (it is claimed) to achieve the task for which they were designed, namely accounting for the pattern of pronominalisation in texts. The essential point is, again, that it is only when relations are given some sort of *theoretical role* that we can even begin to talk about descriptive adequacy. The standards of adequacy are set by the demands of the theory in which the relations figure. The theory will determine what information about a text relations are supposed to capture; we can *then* ask whether the description they provide is in fact sufficient to capture that information.

In short, the criterion of descriptive adequacy, while in some sense essential for any set of relations, is far from being able to stand on its own as a method of justification.

3.3 Associating Relations with Cue Phrases

Cue phrases[3]—clausal/sentence connectives such as *but* and *because*—have provided another source of evidence for justifying a set of relations. Such phrases are often conceived as *signalling* relations in a text: thus, for instance, *because* can be used to signal the presence of a CAUSE relation:

(3.1) Jane fed Lars *because* he was getting so hungry.

It is important to recall that we are interested in looking at the *set* of cue phrases, in order to motivate a *set* of relations. As emphasised in Section 1.2, this task is distinct from the 'first-order' task of identifying relations in actual texts. A study of cue phrases will certainly not suffice for this latter task; it is widely accepted that relations can be

[3] What I am calling 'cue phrases' have been given many different names in the past: 'conjunctive elements' (Halliday and Hasan (1976)), 'clue words' (Cohen (1984)), 'cue phrases' ((1986)), 'discourse markers' (Schiffrin (1987)), 'meta-technical utterances' (Zuckerman and Pearl (1986)). Various different ways of defining them have been suggested, and the different definitions pick out slightly different sets of phrases. For the moment, the class of cue phrases can be identified by typical examples, such as *however*, *then*, *previously*, *or*, *next*, *while*. I will propose a more rigorous definition of my own in the next chapter.

unmarked in text, so cue phrases will only be a partial source of information in this regard.

3.3.1 An Atttractive Source of Evidence for Relations

Cue phrases would certainly be convenient as a source of evidence for a set of relations. There are a large number of them, and the differences between them can be quite subtle; associating cue phrases with relations would yield a sophisticated classification of relations. For instance, Elhadad and McKeown (1990) have noted some subtle differences between *but* and *although*, phrases which are often inter-substitutable:

(3.2) a. He failed the exam *but* he's smart. Let's hire him.
b. *He failed the exam *although* he's smart. Let's hire him.

We would expect these differences to be reflected in the set of relations if we constructed it to mirror the set of cue phrases.

At the same time, using cue phrases as a source of evidence would give us a way to decide at what level of detail to *stop* making distinctions. If alternative cue phrases exist to pick out two similar relations (as in the above cases), then they can be distinguished. But there would be no need to create two separate relations if no cue phrases exist for distinguishing them in text. Thus, for example, there would be no need to distinguish between 'FEMALE-VOLITIONAL-CAUSE' and 'MALE-VOLITIONAL-CAUSE'.

Finally, cue phrases simply provide an *extra source of information* when it comes to working out relation definitions. Many systems for justifying a set of relations start 'from first principles', without any preconceptions about which relations are eventually going to be decided on. Thus Maier and Hovy's (1991) taxonomy of relations is based on Halliday and Hasan's (1976) analysis of the functions of language as 'ideational, interpersonal and textual'; Hobbs' (1985) classification is based on similar abstract considerations about 'the situation in which discourse between a speaker and a listener takes place'. It is certainly essential to have such high-level concerns in mind when working out relation definitions—but the task would be considerably eased if we could also make use of information about the applicability of cue phrases. This is because we can find out about all the situations in which a relation can be used, *in advance of working out its definition*—we just need to examine the range of ways that the appropriate cue phrase can be used in discourse. For instance, having posited that a relation exists which is signalled by the cue phrase *although*, we can consider the different contexts in which *although* can be used, and try to abstract from these to create a definition of the relation itself. The ability to draw on concrete linguistic examples is likely to be of considerable help.

3.3.2 Previous Work with Cue Phrases

As was seen in Chapter 2, a number of researchers have made use of cue phrases in determining a set of relations. While Halliday and Hasan (1976) are only interested in classifying the linguistic resources available for signalling relations, other theorists have used the cohesive resources in a language as 'evidence' for a set of underlying coherence relations, which can themselves be marked or unmarked. The work of Ballard *et al* (1971) and Longacre (1983) follows this strategy, appealing to surface syntactic phenomena in the motivation of 'deep' interclausal relations:

> It is our contention... that a surface taxonomy of form within a language determines a similar taxonomy of deep relations, and that the two taxonomies stand and fall together.
>
> *Ballard* et al *(1971), p75*

According to this approach, the existence of a cue phrase in a language is testimony to the existence of a particular type of interclausal relation, and this relation can apply between clauses even in the absence of the cue phrase. Among the attractions of this method, Ballard *et al* note one I already mentioned—that it enables us to fix the level of detail of the analysis:

> The deep grammar... stops short of dissolution into general semantic or logical categories. It stops in fact where the structure of a given language indicates a cut-off point in that it sets up no more deep structure categories than are required to account for surface encodings.
>
> *Ibid.*

Martin (1992) follows a similar strategy, suggesting that the relations in a text are in principle markable by surface conjunctions. His classification of cue phrases is extremely detailed, and is a clear testimony to the benefits of cue phrases for creating a subtle taxonomy.

Several other researchers make reference to cue phrases when putting forward a set of relations. For instance, Hobbs (1985) uses cue phrases as an informal method of deciding on which of his relations applies in a text—if you can insert *then* between two segments, then the OCCASION relation 'is an excellent candidate'. But he is emphatic that such tests do not figure in the *definitions* of the relations he presents. Scott and de Souza (1990), in a study of how RST relations can be textually marked, found that a large number of them map closely onto cue phrases. Sanders *et al* (1992, 1993) also identify 'prototypical markers' for each relation in their taxonomy, although the main justification for the taxonomy (to be examined in Section 3.4) is not concerned with linguistic issues. Even Mann and Thompson's (1988) relations, which are expressly defined without reference to surface linguistic phenomena, can often be associated with classes of cue phrase.

3.3.3 Problems with Reliance on Cue Phrases

To sum up: there are many advantages to be gained in using cue phrases to decide on a set of coherence relations; and cue phrases have been quite widely used for this purpose. However, an important problem remains for all of these attempts to date: the decision to link relations to cue phrases *itself* needs to be justified. Without such a justification, we still lack a reason to prefer the chosen set of relations over other possible sets.

To illustrate the problem, consider the text in 3.3.3:

(3.3) Bob cooked supper that night. His wife had been working hard at the office all day.

The current suggestion is to account for the coherence of this text by saying that the two clauses are linked by 'some relation which can be signalled by a cue phrase'. (In this case, *because* seems the most likely cue phrase for the job.) But why should this *because*-based relation be any more appropriate in explaining coherence than relations like VOLITIONAL-CAUSE or CAUSE-OF-COOKING-EVENT, which don't happen to correspond to cue phrases? How can we use the association with a cue phrase to *argue* against someone who proposes an alternative relation?

Somehow an argument must be given that relations linked to cue phrases tell us something about the text that relations otherwise classified do not. But neither Longacre nor Martin (who are most explicit in their appeal to cue phrases) provide such an argument: in fact, justification seems to remain principally in terms of descriptive adequacy; and it has already been argued (in Section 3.2) that such an appeal is of little use by itself. To reiterate the point made in Section 3.2: the notion of 'descriptive adequacy' cannot really be used on its own to evaluate a descriptive formalism. It needs to go hand in hand with a *theory* about the thing described: then we can ask whether the formalism is adequate to provide a description in the appropriate theoretical terms. So for a sounder justification of the reliance on cue phrases, what is needed is some sort of theory about text, which would show why linking relations to cue phrases makes them *particularly revealing* as descriptive constructs. At that point, it would be reasonable to prefer them over any set of relations which could not be given such a theoretical role. But such a theory still needs to be provided.

3.4 Looking for 'Psychologically Real' Relations

A third method for justifying a set of relations begins immediately by giving them a theoretical role. The central idea is that relations model *psychological constructs*—that is, they tell us something about the psychological processes which occur in people when they create and interpret text. A given relation can then be justified by producing evidence that it is one that people actually use when processing text. This is why it makes sense to use the relation to link two text spans: because it models part of the process which *actually led* to these spans being juxtaposed as they are. By using a relation conceived of as psychologically real, we are not just describing the text in an arbitrary manner, but contributing to an *explanation* of why the text is the way it is. Clearly, this conception of relations gives us exactly the sort of 'theory about text' that we need in order to justify the use of relations as an appropriate descriptive device.

Many theories of relations make an appeal to psychological notions. As mentioned in the previous chapter, Hobbs (1985) thinks of his relations as 'text building strategies', used by the speaker to facilitate the job of the listener. And he makes some suggestions (p22) about the nature of these strategies. To re-iterate the quote given at the outset:

> It is tempting to speculate that... 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.

Mann and Thompson's (1988) relations also embody psychological insights. Relations are functional constructs, associated with the particular effects a writer intends to achieve; relation definitions make extensive reference to the psychological states of the reader and writer. Yet no *evidence* is given that the RST relations are the ones people use: justification of the relations is again purely in terms of their 'descriptive adequacy' in the hands of discourse analysts.

By far the most thorough investigation of the idea that relations are psychologically real comes from Sanders, Spooren and Noordman (1992, 1993). Any theory proposing that relations 'model psychological constructs' must address two key issues. Firstly, some kind of account of human text processing is needed, which makes it clear what role is played in the mechanism by 'relation-like' constructs. Secondly, there has to be some way of *investigating* these psychological constructs, and working out what they are. Sanders *et al* address both of these requirements.

3.4.1 An Overview of Sanders *et al*'s Work

Parameterising the Space of Relations using 'Cognitively Basic' Primitives

Sanders *et al* approach the issue of psychological reality from the perspective of text understanding. Understanding a discourse involves constructing a coherent mental representation of it, and this in turn involves setting up appropriate *links* between the representations of its various segments.

These coherence links, it is argued, are likely to be established using *general cognitive resources*. During comprehension, the segments of the discourse are integrated into a language-independent representation; part of the reader's general framework for making sense of the world. Coherent texts are likely to be structured in such a way as to facilitate this integration—so Sanders *et al* propose that coherence relations should be investigated by looking for the 'cognitively basic' features which must underlie them.

Four 'cognitively basic' primitives are identified, according to which relations can be classified: these are described below.

- Basic operation. Every relation is deemed to have either a causal or an additive component. causal relations are those where a 'relevant' causal connection exists between the spans; all other relations are additive. a is an example of a causal relation; b is an example of an additive one.

    a. The drive to the arrivals was closed so that nobody could leave the terminal.

b. Centraal Beheer's turnover is about 2.4 billion guilders. In 1988 the profits increased from 75 million to 103 million guilders.

- Source of coherence. Every relation is coherent on semantic or pragmatic grounds. It is semantic if the spans are related in terms of their propositional content and pragmatic if they are related because of their illocutionary force. a is an example of a semantic relation; b is an example of a pragmatic one.

  a. Theo was exhausted because he had to run to the university.

  b. Theo was exhausted because he was gasping for breath.

- Order of segments. This distinction only applies to causal relations; they are deemed to have basic order if the antecedent is on the left, and non-basic order if it is on the right. a is an example of a **basic** relation; b is an example of a non-basic one.

  a. The drive to the arrivals was closed, so nobody could leave the terminal.

  b. Nobody could leave the terminal, because the drive to the arrivals was closed.

- Polarity. A relation is positive if its basic operation links the content of the two spans as they stand, and negative if it links the content of one of the spans to the negation of the content of the other span. Negative polarity relations typically involve either a violation of expectation, where the expectation derives from a causal basic relation; or a contrast, where the basic relation is additive. a is an example of a positive relation; b is an example of a negative one.

  a. Because he had political experience, he was elected president.

  b. Although he had no political experience, he was elected president.

These four parameters can combine to form twelve 'complex' relation types. For each type, Sanders *et al* provide one or more sample RST-like relations: for instance, CAUSE-CONSEQUENCE (basic operation = causal; source of coherence = semantic; order = basic; polarity = positive); or CLAIM-ARGUMENT (basic operation = causal; source of coherence = pragmatic; order = non-basic; polarity = positive). Each of the relations is associated with a 'typical' connective word used for marking it.

Evidence for the Parameterisation

Support for the four parameters is provided by a number of empirical experiments. The first of these used discourse analysts as subjects: they were given definitions of all the relations, and asked to decide which relations were appropriate for a number of sample texts. The second experiment used 'naive' subjects who did not know about the relation definitions: they were shown sample texts without explicit connectives, and had to decide which connective word was most suitable. Both experiments were designed to test how much agreement there is on how to use the relations. In both cases it was found that there was a fair amount of agreement between subjects. Equally importantly, where there was disagreement over which relation to use, it tended to be over the value of a single parameter only—this provides support for the independence of the decisions about the different parameters.

However, in both experiments, there were differences in the *strength* of evidence for different parameters. In the first experiment, there was hardly any confusion over the value of the polarity parameter. But for negative polarity relations, there was a great deal of confusion about the source of coherence parameter: in fact, this parameter is the least agreed upon for all classes of relations. This finding is replicated in second experiment: agreement is lowest for source of coherence.

A third experiment also used discourse analysts as subjects: here the task was to *compare* the coherence relations in a number of different sample texts, and to group those texts which used the same relation. The results showed four distinct clusters of relations: positive causal relations, positive additive, negative relations, and 'conditional' relations (a subtype of causal relation). Again, source of coherence is not well distinguished—there is no evidence for this parameter amongst negative relations, and between positive relations there is not much. Finally, in this experiment, no evidence at all is found for the order of spans parameter.

The last experiment was targeted specifically at the source of coherence parameter. It was hypothesised that confusions regarding this parameter would be lessened if relations were presented in contexts rich enough to disambiguate them. The subjects were again discourse analysts; the task was similar to that in the first experiment. Only those relations with positive polarity were examined. For these relations, under these conditions, it was found that a distinction can indeed be made between semantic and pragmatic relations.

3.4.2 Some Problems with Sanders *et al*'s Parameterisation

While the initial idea that 'relations are psychologically real' provides a very promising method for justifying a set of relations, deciding on exactly *which* set of relations is psychologically real presents problems of its own. The experiments reported above can be criticised on a number of grounds: they do not provide conclusive evidence for Sanders *et al*'s four-way parameterisation.

Problems with Specific Parameters

Firstly, two of the parameters are supported much less strongly than the others. The first three experiments give only weak support for source of coherence. The final experiment only gives support for this parameter for a subset of the relations in the set; and then only under ideal conditions.

Questions can also be raised about order of spans: there was no evidence at all to support this parameter in the third experiment. The first experiment showed that analysts *could* distinguish between basic and non-basic order; but this just shows that they could use the relation definitions they were given, it does not legitimise these definitions.

How do you Choose a Set of Parameters to Test?

The problems with source of coherence and order of spans throw up a more general problem with the experimental methodology—that of *deciding* on a set of parameters to test. Sanders *et al* (1993) take a lot of trouble giving *a priori* justifications for each parameter; but there is still no *systematic* approach towards deciding on the set. For instance, we have no way of knowing that *all* the relevant parameters have been identified. Maybe there are more than four: 'hypotheticality' could be another candidate, in view of the results of the third experiment.

One argument given for the four chosen parameters is that they result in a *productive* taxonomy—in other words, that every combination of parameter values seems to yield a plausible relation. This point is cited as one reason why temporal relations are not included as a parameter. But for a start, Sanders *et al*'s four chosen parameters do not currently result in a *completely* productive taxonomy: the fact that order of spans is only defined for causal relations means that it has four 'empty' slots. Moreover, why should we expect that the cognitive factors which underlie coherence relations will result in a neat parameterisation? It seems perfectly plausible that some factors are only relevant for a *particular* type of coherence relation.

Some Questions about the Experimental Paradigms

Two kinds of experiments are carried out by Sanders *et al.* In one kind, the subjects are discourse analysts and the task makes explicit reference to coherence relations: texts have to be analysed using relations, or sorted into groups on the basis of the relations they use. In the other kind of experiment, the subjects are 'naive' about theories of discourse, and the cue phrases they use are taken as evidence of the relations that they perceive in a text. Both types of experiment are open to question. It is not certain that results obtained from discourse analysts, thinking explicitly about relations, can be taken as evidence for the kind of relations that people normally use when they process text. Neither can the evidence from cue phrases be taken as conclusive: why should we suppose such a tight association between cue phrases and the constructs we make use of when processing text? At the very least, an argument must be given for this policy. Otherwise, we are making unfounded assumptions about the very constructs we are investigating.

The Grain-Size of Relations in the Taxonomy

For Sanders *et al*, the twelve combinations of parameter values do not pick out individual relations; rather *classes* of relations. Thus, for example, two 'prototypical relations' are identified which are causal, semantic, basic and positive—CAUSE-CONSEQUENCE and CONDITION-CONSEQUENCE. How are these individual relations identified? How many of them are required? Sanders *et al* talk about these extra relations as needed to achieve 'descriptive adequacy'; but once again, it is unclear how this criterion is to be interpreted. *What* needs to be described? At the outset, relations were conceived of as modelling the cognitive constructs involved in human text processing. Does this mean that such constructs exist at a finer level of detail than can be specified by the parameters? If so, additional principles are surely needed for picking them out. And if not, then what is the point in refining the taxonomy beyond the level of detail provided by the parameters?

To sum up: a number of problems remain with Sanders *et al*'s justification of a set of relations—although their enterprise of looking for evidence for cognitive text-processing constructs certainly seems the most attractive way of effecting a justification. In the next section, an alternative source of evidence for these cognitive constructs is proposed, which (hopefully) overcomes many of the problems that have so far been raised.

3.5 A New Motivation for Relations: Linguistic Evidence for Psychological Constructs

So far, we have looked at two methods for justifying a set of relations. One suggests using the space of cue phrases in a language to work out a taxonomy. This enables an extensive and detailed taxonomy to be worked out; but it is unclear what explanatory role relations thus justified are to play in a theory of discourse coherence. The other method is based on the idea that relations model the psychological constructs which mediate the production and interpretation of discourse. Here, the theoretical role of relations is clear; but difficulties arise in the attempt to *discover* what these constructs might actually be.

It will be noticed that the advantages and the problems for these two approaches are complementary. The central idea in what follows is that the approaches can be *combined*, so as to capitalise on their advantages and minimise their drawbacks, by taking cue phrases as evidence for cognitive text-processing constructs.

3.5.1 The Central Argument

The claim to be established is the following: that the existence of a cue phrase in a language is good grounds for inferring the existence of a corresponding 'relational' construct in the cognitive apparatus of those who use the language. It will be argued that a language is likely to contain resources for making explicit all the relations which play an important part in human discourse processing. An additional argument will be given as to why cue phrases are a particularly appropriate kind of 'linguistic resource' to study.

The implications of this argument can be illustrated by giving an example. It would mean, for instance, that the existence in English of the word *however* points towards the existence in speakers of English of a *text structuring strategy* which can be signalled in text by using that particular phrase. Again, it is important to note that I am *not* suggesting that cue phrases can be relied upon to identify the relations in *actual* texts—as has already been pointed out, relations will not always be explicitly signalled. My suggestion is rather that the *set* of cue phrases in a given language can be used to determine a *set* of relations: how these relations are identified in particular instances is a separate question.

The claim being made is far from self-evident. Two quite different sets of things are being associated; a set of clause/sentence connectives and a set of cognitive strategies. Why should we assume that there is a connection? Or at least, why should we assume that the connection is strong enough to warrant cue phrases being used as a central source of evidence? Some researchers do assume such a connection without discussion—for instance in Sanders *et al*'s experiments, subjects' use of cue phrases is taken to reflect the relation they are thinking of. Yet it could be that the mapping between cognitive representations and cue phrases is not just a simple one-to-one: the relations people use might be more subtly delineated than is suggested by the range of cue phrases in a language; alternatively, they might be more broadly classified, so that the distinctions between cue phrases give a false impression of the accuracy with which relations are specified. These are real possibilities: if we want to make a direct link between cue phrases and cognitive constructs, we must provide an argument for so doing.

The claim that cue phrases mirror people's text-structuring mechanisms will be supported in three stages. Firstly, a clearer idea will be sought about what such mechanisms might be expected to be like, and what their role could be in the tasks of text creation and text comprehension. Next, a model of the process of 'communication via a text' will be advanced, in which the *communication of relations* between a writer and a reader plays an important role. Lastly, it will be argued that since the communication of relations is an important feature of communication via a text, it is to be expected that language contains ways of making relations explicit.

3.5.2 What Are 'Psychologically Real Relations'?

In this section, the nature of the psychological constructs we are looking for is examined in more detail. We are interested in how people represent the relations between segments of text, for the purposes of text comprehension and text construction, and how the representations are used in these tasks.

In one sense, the idea that 'people use relations to structure text' is almost trivially true: clearly, people form *some* representation of the relations between text segments, because texts are more than just collections of clauses, and people can recognise the difference between a coherent text and a collection of clauses. Somehow, the way clauses are combined in coherent text is being modelled: the real issue to address is what these models are like, and how the models influence text processing. And here there are many open questions. Do people use the same structuring strategies for reading as for writing? Do we really use a 'fixed set' of strategies, or are more general mechanisms in operation? Do we represent inter-segment relations using the same resources we use to represent the content of sentences, or are they treated in a different way? These issues and others will be addressed below.

It is important to note that we are not yet asking about *how to define* the relations which people use—this question is dealt with in the two following chapters. The present concern is rather to ask what sort of psychological constructs relations might be, and what processing purposes they might serve.

Relations In Discourse Interpretation

A basic preliminary observation is that people find it *easier* to process texts if they are more than just a collection of clauses. It is not just that we can recognise the difference between a coherent text and a collection of clauses; we actually *use* some representation of the relations in a text to help us process it. This can be seen, for instance, in a study by Meyer and Freedle (1984). Pairs of texts were prepared which differed only in the relations between groups of sentences: for example, in one text, the elements of content were linked by CAUSAL relations, and in the other, they were just presented as a collection. In this case, subjects' recall was much better for the former texts; this suggests that the presence of CAUSAL relations somehow facilitates text processing or storage.

Further studies show that the signalling of relations in a text facilitates its interpretation. For instance, Haberlandt (1982) shows that reading time is improved by the addition of linguistic markers; Segal *et al* (1991) show that the presence of interclausal connectives in a text helps subjects decide how to classify the connections between its clauses. We should not assume in interpreting such studies that *particular* surface cues mark *particular* coherence relations—this is the very claim that we are trying to justify. But the *general* finding that connectives facilitate discourse processing can be taken as a sign that relations between spans of text are somehow involved in the process: these experiments again show that it is important for readers to work out how text segments are linked together.

Alternative Conceptions of Psychological Reality

Sanders *et al* base their taxonomy of relations on a 'psychologically plausible' account of how the relations in a text are interpreted. They begin by claiming that relations of the kind proposed in RST are implausible as psychological constructs, because they are treated as unanalysed units:

> from a psychological point of view, Mann and Thompson's ideas are not very convincing, because they assume that all relational propositions are cognitively basic. If, for example, a relation like EVIDENCE occurs in a discourse, people interpret the discourse by referring to the cognitively basic notion of the EVIDENCE relation... Such an assumption is rather implausible.
>
> *Sanders* et al *(1992) p4*

The idea that people should have 'in the head' a complex construct such as the EVIDENCE relation is seen as unlikely. For Sanders *et al*, it is more plausible to *decompose* relations according to more 'general' principles of cognition such as causality and polarity: it is then claimed that such principles are used *jointly* to *infer* relations such as EVIDENCE. This is thought to be more plausible than the idea that each relation is a completely separate purpose-built construct.

However, it is not clear that general principles of cognitive organisation are the *only* ones which can be considered 'psychologically plausible'. A much broader conception of psychological reality seems possible, especially when we consider that reading and writing are highly skilled, practised activities.

Psychological studies of skill acquisition frequently point towards a model in which practice at a task leads to the development of *specialised mechanisms*, specifically tailored for performing the task in question. Such mechanisms have been posited in many different domains. For instance, Reason (1979) suggested that everyday tasks like driving and cooking are carried out by a system of **motor programs**, operating with a certain degree of autonomy. Models of linguistic processing also commonly involve sets of specialised constructs evolving during the course of practice: PDP models are a case in point. Consider, for instance, Rumelhart and Norman's (1982) model of the performance of skilled typists: here, every word in the lexicon is associated with an **action schema**, which when activated, sets off a chain of events leading to a sequence of keypresses. Such schemata develop as a typist's skill increases; they are clearly not 'cognitive primitives' in Sanders *et al*'s sense. And yet, they (or something like them) definitely seem to be 'used by people' in performing tasks.

The idea of a set of specialised constructs capturing regularities at the lexical level is extensible upwards to larger levels of structure. It is possible to imagine constructs corresponding to coherence relations evolving in the same way: producing and understanding large pieces of text are highly practised tasks, and it does seem plausible that mechanisms are developed specifically for them. When thinking of psychologically real relations, therefore, we do not have to limit ourselves to thinking about general 'cognitively basic' principles. However, this is not to say that general cognitive principles should not feature at all in a psychological theory of relations. A complete theory might well envisage relations *partly* in terms of general cognitive resources, and *partly* as learned strategies for structuring discourse.

Relations in Discourse Production

Sanders *et al*'s account of psychological reality is mainly based on the task of discourse interpretation: how coherence relations are inferred from passages of text. It is suggested that relations have a role in discourse production as well, but this role is not examined in any detail. This may be because less research has been done on how extended passages of discourse are produced—nevertheless, there are some studies, particularly about written discourse production, and it is instructive to consider how coherence relations can be fitted into existing theories.

Psychological theories of writing are still at a relatively early stage compared to theories of discourse interpretation, the difficulty being how to get an experimental handle on the process. Some studies (eg de Beaugrande (1984)) have used films of writers at work, and analysed pauses and crossings-out; but this source of evidence is unlikely to be rich enough to provide a full account of the processes involved. Most current theories are based on another technique—**protocol analysis**. In this paradigm, subjects are asked to 'verbalise' while they write, about what they are thinking about and how they are performing the task that they have been set.

Again, this source of evidence is questionable: it is unlikely that all the mechanisms involved in writing are amenable to verbalisation. Despite this, some initial theories of written composition have been formulated, which look plausible as far as they go. Perhaps the best known theory is that of Flower and Hayes (eg Flower and Hayes (1980), Hayes and Flower (1980)). This theory draws on spoken protocols, and also on the the notes taken by writers. It is suggested that the task of writing involves three separate sub-processes:

Planning: this process itself has three components.

- In the **generating** component, information relevant to the writing task is retrieved from long-term memory. The topic of the text is the initial search key; thereafter, the search key is the last item to be retrieved. The process stops when irrelevant items begin to be produced, and loops back to previous items to look for other relevant information in the same way.
- In the **organising** component, the most useful generated items are selected and organised into a writing plan. A number of **operators** are used to effect the organisation—for instance, 'identify as a possible first or last topic', 'search for a previously noted topic subordinate to present topic', 'order with respect to a previously noted topic'.
- In the **goal-setting** component, criteria are identified by which to judge the text, and they are stored for later use in **editing**.

Translating: this involves creating sentences out of the organised material.

Reviewing: again, there are two components to this process, **reading** and **editing**. The material so far produced is read, segment by segment, and each segment is edited in turn. The editing process detects and corrects inaccuracies in meaning, violations in writing conventions, and mismatches between the writer's intentions and those apparently expressed in the text.

The order of the processes is roughly as outlined above—although the editing process can cause the system to be re-entered at various different stages. In fact, editing and generating can interrupt any of the other processes.

Much of this model seems little more than a 'common sense' view of writing. But while it is clearly not very elaborate, it seems likely that the writing process is broken down in something like the way proposed. A number of hypotheses correlating stages in the process with the form of subjects' notes and protocols seem to be borne out. In which case it must be asked: how (if at all) do coherence relations fit into the model?

There are several possible answers. For one thing, relations could be involved in the **generation** component, as aids to the retrieval of relevant material from memory. If the content of one portion of text has been decided, then a coherence relation could give us useful information about other items of content which could feature in adjoining portions of text. A set of relations could provide a standard repertoire of methods for accessing relevant content. Evidence for this idea comes from a study of young writers by Bereiter and Scardamalia (1987). Children experience particular problems in finding content in their compositions: this study examines the effect of cueing children with suggestive prompts when they 'dry up'. The prompts given are predominantly cue phrases—for instance *even though*, *also*, *for example*. Given such cues, children are frequently able to produce more material for their compositions. Bereiter and Scardamalia suggest that the cues 'appear to stimulate the children to search for new nodes in memory that meet the logical requirements of the sentence openers' (p62). More mature writers, it might be supposed, will have internalised such cues, so that they can initiate the search for content themselves instead of waiting for an external prompt. We would not want to suggest that these internalised cues can be mapped directly onto surface cues. But this experiment does show that whatever form they take, methods for accessing new and relevant content are of great use in creating text.

Another possible role for relations is in Flower and Hayes' **organisation** component. The operators used in this component are used to structure elements of content already accessed. Many of them work with pairs of topics, for instance by deciding which of two topics to mention first, or whether one topic is subordinate to another. Coherence relations (which might well feature ordering constraints or incorporate hierarchical concepts like subordination) could be involved in making such decisions.[4]

It is interesting to note that the above uses for relations correspond quite closely to the uses found for relations in current text generation programs. Hovy (1988) uses relations as planning operators for working out text structure; Moore and Paris (1989) incorporate relations into planning operators, and in addition use these operators to access new elements of content to be generated. Many other subsequent systems have a similar design. It would be disingenuous to draw conclusions about human text processing mechanisms by looking at the way current text generation programs operate—but when constructs used in generation systems are found to resemble constructs hypothesised in psychological models, it seems worth mentioning the fact.

To sum up: it is plausible to think of coherence relations as modelling a set of strategies used by people to access and organise elements of content when planning text. The conception of relations as planning operators seems useful for psychological modelling as well as in computational systems.

A 'Basic Level' of Relations

One problem for the above idea is that it is hard to find any single relation that is not *sometimes* going to be useful for 'accessing and organising material'. For instance,

[4] A third possible role for relations is in the **translation** component, where the plans thus far built are converted into sentences. If relations were involved in the construction of plans, and at the same time associated with surface linguistic expressions, then the task of translation would be considerably facilitated: the relevant linguistic expressions would be *predetermined* by the plans. However, this idea cannot be used in the present argument, because—again—it assumes the conclusion we are trying to reach; that cue phrases can be used as evidence for relations.

consider a 'possible' relation like 'BROTHER-EVENT', where one span introduces a person, and the other span presents something happening to that person's brother. Such a relation would be helpful, almost by definition, in generating Example 3.4:

(3.4) Lars fixed the boat. His brother had holed it by bashing into the jetty.

Having decided on producing the first clause, BROTHER-EVENT could be used as a search cue to retrieve the material in the second clause. If coherence relations are to be thought of as strategies for accessing and organising material, why shouldn't BROTHER-EVENT count as a coherence relation?

One answer is that this relation is *seldom* a useful one when it comes to producing a coherent text. In most cases, if this relation is used, an incoherent text results:

(3.5) *Lars fixed the boat. His brother was a sergeant in the Danish police.

Furthermore, in the coherent cases, it is likely that an alternative relation can always be found. In 3.4 above, other relations such as BACKGROUND or CAUSE could have been used; and these are more often useful when it comes to structuring text.

But this explanation is not yet completely convincing. It could still be claimed that BROTHER-EVENT is useful *in particular cases*, such as Example 3.4. The relation 'BROTHER-EVENT' must have *some* kind of mental representation: if the reader or writer did not appreciate that it was Lars' *brother* who had holed the boat, then something would be missing from their representation of the text.

Of course, we can appreciate that in this example, Lars' action is not caused by just any event, but by an event involving his brother. Likewise, *each* CAUSE relation will be unique in certain ways: the point is that CAUSE is an *abstraction* from particular instances of relations. When I suggest that CAUSE has 'psychological reality', I am not claiming that we *only* represent relations at this level of abstraction, but that it is used *for some purposes*.

It is useful to take an analogy from another field of psychology at this point: Rosch's theory of categorisation (eg Rosch *et al* (1976), Rosch (1978)). Rosch claims that in order to form strategies for dealing with the infinite variety of stimuli we are faced with in the world, we have to work with abstractions, since we are finite processing devices, and we cannot have a particular strategy for each stimulus. Thus we have to treat some stimuli as equal: Rosch calls the level of abstraction at which we operate **the basic level**. The basic level is thought of as optimising the trade-off between useful categories and general categories. For instance, 'chair' is a basic level category: if we have to write down all the things you can do with a chair, the list will be much longer than the list of all the things you can do with a piece of furniture (a superordinate category), and not much shorter than the list of things you can do with an armchair (a subordinate category). Rosch claims that for some purposes, we work with the concept 'chair' rather than with more specific concepts. Yet at the same time, of course, we can recognise individual chairs and tell them apart. Different tasks call for reasoning at different levels of abstraction.

A similar point can be made for relations. It's just for some components of the text generation process that we abstract to the level of CAUSE or BACKGROUND. This does not mean that we can't recognise and differentiate individual instances of these relations for other purposes—of course we can tell the difference between individual instances of a particular relation. But text generation is a hard task, with lots of simultaneous constraints, demanding lots of processing: the ability to work with abstractions would be a useful one. In order to decide which abstractions to use, it would make sense to look again at the trade-off between utility and generality. Thus it is likely that BROTHER-EVENT is not so good at retrieving relevant information as CAUSE: so it would be less useful to work with this concept. In the same vein, a more specific relation like CAUSE-THROUGH-BROTHER'S-ACTION is not likely to be much more useful than CAUSE as a search cue for relevant material. Thus we would want to include CAUSE in our set of coherence relations, but not BROTHER-EVENT.

The above ideas should help to give substance to the idea that a set of coherence relations can be taken to model mechanisms 'used by people' when they produce text.

3.5.3 The Communication of Relations

Thus far, the production and the interpretation of discourse have been considered separately. In this section, we consider how the two processes come together, in what we might call 'communication via a text'.

For the purposes of the argument, it is not only important that writers use relations when creating a text, and that readers use them when interpreting it, but that they use the same relations. Otherwise it is impossible to argue that it would be helpful for writers to signal the relations they use in surface text. In order to make this argument, the model needed is something like the one in Figure 3.1: here, the *communication of relations* is seen as an intermediate step in the communication of a writer's goals and ideas to the reader. In such a model, the identification of relations is something which

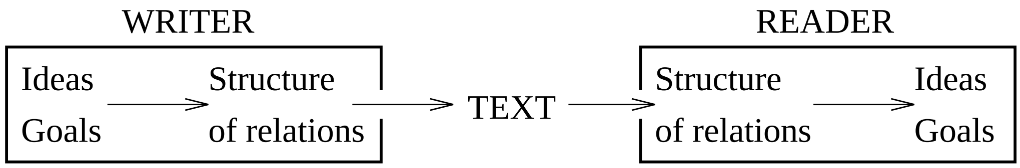

Figure 3.1: A Model of Communication Via a Text

really matters: readers need to be able to do it. It is this which makes it likely that ways exist for identifying relations explicitly.

What arguments can be given for the idea that writers and readers use the same relations? A number of points can be made. Firstly, if we assume that writers learn how to structure their texts by reading other peoples' texts, it seems likely that there will be an overlap in the methods used in reading and writing. More importantly, arguments can be given from the perspective of computational efficiency. The knowledge that a fixed set of relations is used to structure text permits a big reduction in the search space for both reading and writing. When a reader is working through a text, (s)he will know that the *next* segment of text will be related to the current segment in one of a fixed number of ways. If the set of relations includes a CAUSAL relation, then this is one of the alternatives the reader must always consider. Relations which are *not* in the set do not even need to be considered. Imagine a 'possible' relation which is unlikely to feature in the actual set; for instance the 'BROTHER-EVENT' relation mentioned above. The argument is, that at some stage of processing, the question of whether BROTHER-EVENT holds is not one which needs to be asked; whereas a relation like CAUSE is always going to be a possibility. Take a text like 3.6:

(3.6) Lars woke up early one morning. His brother was being noisy in the kitchen.

The obvious interpretation of the relation in this text is CAUSE, with the first span as the nucleus. But if BROTHER-EVENT were a relation in the set, then this would be an alternative: in this case, we wouldn't know which kitchen was involved, or when Lars' brother was being noisy in it. The text would be much more ambiguous, and it would be likely to be harder to process. It is a sign of how heavily we can rely on the conventional set of relations that we find it hard to even imagine a relation like 'BROTHER-EVENT'.

Writers will also profit from the conventional use of one particular set of relations. In example 3.6, the writer knows that enough has been done to disambiguate the relation in question, because the reader will not be expecting a relation like BROTHER-EVENT: so there is no need to make the causal relation explicit. In order to achieve coherence, texts only need to be specified with sufficient detail to allow the reader to work out which of the limited number of relations is being used.

From the point of view of computational efficiency, then, a strong case can be made for the use of a standard, smallish set of relations, by both readers and writers.

3.5.4 The Need to Signal Relations in Text

The final stage in the argument builds on the idea worked out in Section 3.5.3, that the communication of relations 'really matters'. It is shown how this idea can be taken to support the strategy of using cue phrases as evidence for the relations that readers and writers use.

The argument is as follows: if people actually *use* a certain set of relations when constructing and interpreting text, it is likely that the language they speak contains the resources to signal those particular relations explicitly. If people plan texts by building a structure of relations, and understand texts by working out this structure, then being able to mark relations explicitly in text will facilitate the communication process, by making it easier for a writer to indicate to a reader which relation is intended. As a consequence, ways of signalling relations in text would be extremely useful.

Of course, relations do not *always* need to be signalled in text. Often, they will be inferrable without explicit signals, as in example 3.6 above. The inference can be due to contextual information, or to the reader's general knowledge, or to the conventional use of a subset of possible relations. However, it is unlikely that any relation exists that is *always* inferrable in these ways. For one thing, the amount of relevant knowledge that the reader has is not always under the writer's control. Consider a text like 3.7:

(3.7) Bill was laughing. Frank was angry with him...

Here, if we don't know more about the situation, we don't know whether Bill is laughing because Frank is angry, or whether Frank is angry because Bill is laughing. It is useful to have linguistic devices, like the cue phrases *because* or *as a result*, to distinguish these possibilities.

It is not strictly true to say that the reader's knowledge is beyond the control of the writer. In the above example, the writer could have given enough prior context to disambiguate one or the other reading. But using cue phrases is a much simpler way of providing the necessary information; from the point of view of efficiency, it is an attractive option.

So—again from the point of view of efficiency—we can argue that if the communication of relations is important, then simple linguistic means (such as cue phrases) will exist for identifying them in text. If this is indeed the case, then we can look at the range of cue phrases in a language to give us an indication of the relations that people use in constructing and interpreting texts in that language.

3.5.5 Summary

To sum up the argument that has been presented:

- In Section 3.5.2, the idea that 'people use a set of relations when they process text' was fleshed out. A conception of relations as constructs developed during the course of practice was proposed. Rosch's notion of the basic level was invoked, to illustrate how a particular level of abstraction is suitable for particular tasks. It was also used to emphasise that just because we work at this level of abstraction for some purposes, it does not mean that for other purposes we cannot represent the relations in discourse with more finer detail.

- In Section 3.5.3 it was suggested that if readers and writers use relations, then it is likely they use the same set. The argument was on the grounds of efficiency: the conventional use of one particular set would reduce the search space for both reading and writing tasks.

- Finally, in Section 3.5.4, it was argued that if the communication of relations is of real importance in the communication of a writer's ideas via a text, then it is likely that simple linguistic means exist for identifying relations explicitly. Again, the argument was on the grounds of efficiency: language, if it is an efficient communicative tool, should contain such devices.

- If the above arguments go through, then it is permissible to take connective devices like cue phrases as *evidence* for the set of relations that people use.

3.6 Some Objections to the Argument

The argument might be considered quite speculative in some places; particularly those which make reference to considerations of processing efficiency, or which propose the investigation of psychological constructs through an analysis of language. Some of these issues are addressed in this section.

3.6.1 Can you really investigate psychological constructs without doing any psychological experiments?

As a method of studying psychological constructs, the technique being proposed is very unusual. How can we expect to find out about human processing mechanisms just by studying linguistic phenomena? How could we possibly confirm or refute any conclusions we came to? Almost without exception, psychological theories are based on data about how humans perform experimental tasks. In this technique, 'a language' is the source of all relevant data: is this really permissible?

Admittedly, the technique is unconventional as a psychological methodology. However, many objections have already been mentioned concerning the experimental paradigms used to study human text structuring mechanisms. To sum up some of the points in Section 3.4.2: when subjects are asked to talk explicitly about the relations they use while they are processing text, it is not certain that the tasks they perform are exactly those they normally perform. If subjects are 'naive', it is not certain that their understanding of what coherence relations are is sufficiently clear; if they are discourse analysts, it is possible that their intuitions are tainted by biases towards one theory or another. There are methods of studying relations without making subjects think about them explicitly—in particular, recall and reading time can be examined. But these methods are indirect: they can only be used to choose between various hypotheses, not to form hypotheses in the first place. And finding an experimental indicator of the relations used in generation promises to be an even harder task. Cue phrases cannot themselves be used, until an argument is given to suggest why cue phrases and cognitive constructs should stand in a one-to-one relation with each other. But if we have such an argument, then there is little point in conducting an *experiment* using cue phrases as evidence for relations: the argument allows us to look *directly* at the cue phases in the language to find out about the constructs people employ.

To put the positive case for studying language to find out about psychologically real relations: it just seems that language provides an enormously rich source of evidence for studying the ways in which people structure text, and it would be a pity to ignore it. As the next two chapters show, the set of cue phrases in English is very large and diverse, and it is structured in all sorts of interesting ways. Some very subtle distinctions between relations are captured by the different phrases—in comparison, the experimental methodologies outlined above seem relatively impoverished sources of data. Of course I am not saying that psychological experiments cannot *also* be used to investigate the relations people use. I am just arguing that the method of analysing cue phrases is a legitimate (and attractive) alternative.

3.6.2 It might be *useful* if there were a cue phrase for every relation people use: this doesn't mean there *will* be one.

This objection goes to the heart of the argument being advanced. It suggests that we are not permitted to rely on language being perfectly efficient as an instrument for the communication of relations. What grounds are there to suppose that cue phrases *will* exist in a language just because it would be *useful* if they did? For one thing, we need to be certain that language is flexible enough to adapt to the requirements of its users. On top of that, it may not be just a matter of flexibility: maybe efficiency isn't the only constraint on the way languages change. Other conflicting factors may also be involved—for instance, a stylistic trend in favour of terse, simple sentences—such that the end result only *partially* satisfies some constraints. The argument could, in short, be considered Panglossian, affirming that 'language has evolved to be the way it is, and therefore the way it is must be the most efficient'. If this were the case, it might well put the proposed methodology in jeopardy.

We would not want to have to rely on 'the efficiency of language' as an article of faith. Fortunately, there do seem to be at least some indications that cue phrases have evolved through considerations of efficiency. For one thing, many cue phrases are single words, or **idiom chunks**,[5] which are conventionally treated as single words. Moreover, the etymology of many cue phrases suggests that they have evolved from longer, less formulaic phrases. For instance, Halliday and Hasan (1976, p230) list several connective words which were originally more complex phrases, containing anaphora, and making use of the 'compositional' resources of the language. Such words include *therefore*, *thereupon*, *whereupon*, and so on. In addition to this, there are cases of new cue phrases being invented where there is a need for them. For instance, the phrases *iff* and *just in case* have been coined in logical and philosophical genres of text respectively, to replace the longer and more unwieldy phrase *if and only if*. There are many examples of cue phrases evolving in this way: they provide some support for the idea that the set of cue phrases will reflect the set of relational constructs we use.

3.6.3 Cue phrases aren't the only way of signalling relations.

As far as it goes, the argument gives no reason why cue phrases should be studied to the exclusion of other linguistic devices for signalling text structure. In fact, the principal criticism made about the present methodology (Bateman and Rondhuis (1994), Seligman (1994)) has been that it concentrates exclusively on cue phrases.

There are certainly other means of signalling discourse structure. A great many researchers (see e.g. Moens and Steedman (1988), Lascarides and Asher (1993)) have shown that tense and aspect are cues to the temporal structure in a text. Scott and de Souza (1990) have explored a large number of syntactic devices as signallers of RST relations. Several researchers (e.g. Sidner (1983), Grosz and Sidner (1986)) have suggested that the pattern of pronominalisation in a text provides information about its thematic structure. Delin and Oberlander (1992) investigate the discourse structures

[5] Phrases whose meaning is not a function of the semantics of the individual words of which they are composed.

signalled by *it*-clefts. In short, it is uncontroversial that discourse structure can be expressed through a wide range of surface linguistic devices.

In the light of these examples, the present emphasis on cue phrases should really be seen as just a starting point in working out a theory of discourse coherence. Ideally, other cohesive strategies should also be considered. However, the general idea that the investigation should be driven by the range of available resources remains the same. In addition, there are some good reasons for beginning by looking at cue phrases. For one thing, cue phrases are a relatively homogenous set of linguistic devices, whose different effects can be fairly easily compared amongst one another. Furthermore, many alternative cohesive strategies also have cue phrase counterparts. For instance, temporal relations can be signalled by phrases like *previously*, *afterwards* or *while*. Finally, it can be argued that structures which cannot be signalled by cue phrases should not be modelled using coherence relations, but with some other theoretical construct. This argument will be taken up in Section 7.2.

3.6.4 Different languages have different cue phrases.

In cross-linguistic studies, it has been found (eg Ballard *et al* (1971), Longacre (1983)) that different languages often have different sets of cue phrases. Longacre, in justifying his set of relations, uses cue phrases from a number of different languages as evidence. This goes against the argument being suggested here: if a cue phrase doesn't exist in a certain language, we would have to say there was no need for it. This in turn would mean advocating different structuring mechanisms in people speaking different languages.

However, this objection is not insurmountable. For one thing, the differences between the sets of cue phrases in different languages are not *that* great. It is quite surprising how much similarity there is between English cue phrases and those in the Philippine languages which Ballard and Longacre study. And between European languages, which are much more closely related, it is likely that the differences are even smaller.

Furthermore, differences between the sets of relations in two languages might be attributable simply to the different registers of discourse used in these languages. In spoken language, for instance, phrases like *lastly* or *to summarise* are very rare; maybe we would not expect to find them at all in a language with a largely oral tradition, or in one without any writing at all. All this means is that speakers of this language have not needed to internalise the constructs developed by English speakers in the course of learning how to read and write.

Finally—to pre-empt some of the discussion in later chapters—while cue phrases from one language may not always translate directly into cue phrases in another language, it may still be that the dimensions along which cue phrases *vary* are the same in the two languages. Thus, at a more abstract level, interesting similarities may still be found. To take a simple example, there is no single cue phrase in English to translate the German cue phrase *wenn*; two phrases (*if* and *when*) are needed to do the job. However, there are *other* phrases in English—for instance *then*—which seem to manifest just the kind of ambiguity that *wenn* does. See Chapter 6 for further discussion of this idea.

In any case, it is not inconceivable that different language communities use different structuring mechanisms even for producing similar registers of discourse. As mentioned in Section 3.5.3, one psychological role for relations could be as *conventions*, such that the reader knows what set of relations the writer is working with, and the writer knows which relations the reader will be expecting: different languages could make use of different conventions. It is important to remember that 'psychological constructs', in the sense that we are talking about them here, do not *have* to be 'cognitive primitives' posessed by all humans. An equally plausible idea is that relations model a mixed bag of constructs, some of which are cognitive primitives and some of which are acquired through exposure to a particular language community.

3.6.5 What about relations between large segments of text?

One of the attractive features of coherence relations is their insensitivity to span size. Indeed, this feature was taken in Chapter 2 to be central to the notion of coherence relations. However, cue phrases seem primarily designed for linking clauses or sentences together. In linking relations to cue phrases, do we not risk ignoring the issue of higher level relations?

This objection is telling if we adopt a straightforward conception of cue phrases as clause or sentence conjunctions; however, we will actually be working with a more general and informative definition of cue phrases (to be outlined in the next chapter, in Section 4.2). This definition moves away from a syntactic conception of cue phrases, allowing more complex phrases such as *this is because* and *following this*. Stock phrases such as these, which make use of propositional anaphora, can signal relations between large spans of text, and hence (according to the argument in this chapter) can be used as evidence for identifying those relations.

3.7 Summary: A New Proposal for Motivating Relations

In this chapter, a methodology has been suggested for working out a 'standard' set of coherence relations. Coherence relations are thought of as psychological constructs used in planning and interpreting discourse, and it is suggested that the cue phrases in a language can be used as evidence for these constructs.

This argument leaves us with a fairly strong general prediction; namely, that the set of relations which corresponds to the set of cue phrases will suffice to describe the relations in all coherent texts. In the remaining chapters, this prediction will be refined and tested.

Chapter 4

A Data-Driven Methodology for Motivating a Set of Relations

4.1 Introduction

In the previous chapter, it was argued that linguistic devices (in particular, cue phrases) can be taken as evidence for relations, provided these are thought of as constructs which people actually use when creating and interpreting text. This chapter describes how a set of relations can be determined and justified in the light of this argument. The methodology is incremental—it consists of a series of relatively simple linguistic tests, which can be performed quite systematically, with a minimum of inter-analyst disagreement.

To begin with, in Section 4.2, a suitable method is sought for defining 'cue phrases', without relying on terminology from existing theories of discourse. Cue phrases must be characterised independently in order to avoid circularity: since they are to be used to motivate relation definitions, no reference can be made to their role in signalling relations. Instead, a linguistic **test for cue phrases** is proposed, which makes use of readers' intuitions about the coherence of certain constructed mini-discourses.

Section 4.3 describes how a corpus of cue phrases is gathered using this test, and provides some preliminary discussion of its size, and the variation within it. In Section 4.4, a second linguistic test is presented, for classifying cue phrases into groups of synonyms and hyponyms: the test basically determines whether one cue phrase is **substitutable** for another. The results of the test are presented in the form of **substitutability diagrams**, which are explained and illustrated in Section 4.5. Using the test, a **taxonomy** of cue phrases is constructed: this taxonomy is described in Section 4.6.

4.2 Firming Up the Notion of 'Cue Phrase': A Test for Relational Phrases

The first task is to formulate a precise definition for the class of phrases which is under investigation. Some definitions of 'cue phrases' exist already, but these are often internal to the theory of discourse being proposed. For instance, Cohen (1984) defines 'clue words' as 'special words or phrases directly indicating the structure of the argument to the hearer'; Hirschberg and Litman (1993) define cue phrases as 'words and phrases that directly signal the structure of a discourse'. With such definitions, in order to decide what counts as a cue phrase, we already need to know what 'the structure of a discourse' is. In order to avoid circularity, the constructs used in the discourse theory must be justified using some other criterion.

As an alternative to this approach, cue phrases, if given an independent definition at the outset, can be used in motivating the constructs used in the discourse theory. This is the approach adopted here, and sanctioned by the arguments in the previous chapter.

In an attempt to come up with a precise yet theory-neutral definition of cue phrases, a linguistic test is proposed which picks out a certain set of phrases as they occur in natural discourse. The test is given in Figure 4.1 below. It is designed to pick out all sentence and clause connectives, but to stay away from methods of realising relations within a single clause.

In order to avoid any terminological confusion, we can refer to the class of phrases which pass this test as the class of **relational phrases**—although since this is quite a mouthful, the term 'cue phrase' will continue to be used, with this new technical meaning.

The central idea behind the test is that cue phrases have a function which extends beyond a single clause. They link clauses and sentences together to create larger units of text; therefore they cannot be made sense of when associated with one clause in isolation. Thus the clause

(4.4) *Because* Bill owed John money

is impossible to understand without prior linguistic context, but can be understood when the cue phrase is removed:

(4.5) Bill owed John money.

In order to make the test work, any anaphoric or cataphoric expressions in the clause to be isolated must be replaced by their referents; otherwise it would be impossible to interpret out of context regardless of whether or not it contained a cue phrase. However, propositional anaphora *within* the candidate phrase should not be substituted: thus complex constructions like *because of this* or *for this reason* will also be identified as cue phrases. There are good grounds for opting to allow anaphoric expressions within cue phrases: as Halliday and Hasan (1976) point out, many *bona fide* cue phrases

1. Isolate the phrase and its **host clause**. The host clause is the clause with which the phrase is immediately associated syntactically; for instance, if the passage of text to be examined is

   (4.1) ...John and Bill were squabbling: John was angry *because* Bill owed him money. That was how it all started ...

   then the isolated phrase and clause would be

   (4.2) *because* Bill owed him money.

2. Substitute any anaphoric or cataphoric terms in the resulting text with their antecedents, and include any elided items. For the above clause, this would result in

   (4.3) *because* Bill owed John money.

   Propositional anaphora *within the candidate phrase itself* should not be subsituted, however. Thus if the candidate phrase is *because of this*, the propositional anaphor *this* should remain.

3. If the candidate phrase is indeed a relational phrase, the resulting text should appear **incomplete**. An incomplete text is one where one or more extra clauses are needed in order for a coherent message to be framed. The phrase *because Bill owed John money* is incomplete in this sense: it requires at least one other clause in order to make a self-contained discourse. Even the fact that it could appear by itself on a scrap of paper (say as an answer to a question) does not make it complete; the question is essential context if it is to be understood.

   Note that it is only additional clausal material which is to be removed in the test. Any additional contextual information necessary for the comprehension of the clause (for instance, knowledge of the referents of definite referring expressions like *John* and *Bill*) can be assumed to be present.

4. Any phrases which refer directly to the text in which they are situated (such as *in the next section*, *as already mentioned*) are to be excluded from the class of relational phrases. Such phrases pass the test—but only because their referents have been expressly removed through the operation of the test itself.

5. Phrases which pass the test only because they include comparatives (for instance *more worryingly*, *most surprisingly*) are also to be excluded from the class of relational phrases. Stripped of the comparatives, such phrases do not pass the test. Comparatives like *more* and *most* introduce a very wide range of adverbials, bringing the compositional resources of the language quite strongly into play. Since we are more interested in stock words and phrases that have evolved to meet specific needs, phrases involving comparatives will not be considered as relational phrases.

6. Sometimes, more than one cue phrase can be found in the isolated clause (eg *and so*, *yet because*). In such cases, both phrases should pass the test when considered individually in the same context. In other words, the host clause should appear incomplete with either phrase.

Figure 4.1: Test for Relational Phrases

derive etymologically from phrases involving anaphora (consider words like *therefore* and *thereby*); so it seems reasonable also to allow as cue phrases expressions with an explicit anaphoric component.

The test is designed to give a reasonably objective way to pick out a set of phrases to act as the object for further study. It calls on analysts' intuitions, but there is no need for 'coherence relations' to be explicit in their minds: judgements are purely about whether or not given clauses make sense when isolated from their context. There are of course many cases where the test is hard to apply. One problem in the application of the test is to decide whether the required context for a mini-discourse is linguistic or non-linguistic. For instance, consider this discourse:

(4.6) *But* you can't just leave us here!

It is possible to imagine this discourse with no previous utterances at all. All the same, it needs to be interpreted as a reaction to a previously existing propositional attitude (in this case, perhaps an intention to leave), and so can arguably be interpreted as part of a relation between propositions.

It might be considered that the test is overly restrictive in some cases. For instance, as it stands, several phrases used to signal 'purpose'-type relations are excluded. Consider the following two cases:

(4.7) Bill escaped from prison *by* bribing a guard.

(4.8) Bob used the crowbar *to* lever open the window.

Bribing a guard is not a full sentence, and cannot stand alone; *Lever open the window* can be interpreted as a sentence, but only as an imperative, quite different from its original meaning. Other phrases used by Vander Linden (1992) to signal purposive relations are even more clearly ruled out, such as the preposition *for*:

(4.9) Bob used the crowbar *for* extra leverage.

Since the relation here is realised within a single clause, the candidate phrase's host clause contains both parts of the relation, and can *only* be interpreted when the phrase is present!

It would be useful to have a way of expanding the test to allow for additional phrases such as these. But the decision here has been to keep the test reasonably simple, rather than to extend it until it covers exactly the range of phrases we think should be included. Hopefully, when the present corpus of phrases has been analysed, a more principled method for identifying cue phrases can be found.[1]

[1] Note that as it stands, the test works less well in other languages, where connectives often exert a grammatical influence on their host clauses. (For instance, German subordinators can alter the position of the verb; conditional phrases in many languages require a clause in the subjunctive.) It may be that the standardisations of clauses required to overcome these problems are also sufficient to expand the test's English coverage in some of the ways required.

4.3 Gathering a Corpus of Cue Phrases

Using the test, a corpus of cue phrases has been gathered. In order to limit the scope of the investigation, the source texts were all from the same genre of discourse: 'academic' writing, such as can be found in journals or academic books. It is likely that different registers of text require slightly different sets of cue phrases: for instance, phrases like *just then*, *whereupon* and *sure enough* occur in narrative discourse but are unlikely to appear in academic articles. At the same time, the texts analysed in this study occasionally switch into different genres; for instance, a narrative genre is often adopted to describe a case study. (For this reason, the corpus contains some phrases which might not seem typical of 'academic' writing.)

All corpus analysis was carried out by the author. 226 pages of text were analysed altogether, from twelve different authors. This yielded a corpus of around 200 phrases. There was found to be extensive use of a core of phrases across all the authors: for instance, *and*, *since*, *if*, and *but* were used by all twelve; *on the other hand*, *however*, and *also* were used by eleven; and *then*, *for example*, *because*, *when*, and *although* were used by ten.[2]

Following the study, the corpus was somewhat enlarged, again by the author, as new phrases not encountered in the original analysis were discovered. Some of these phrases might well have been found if a larger amount of text had been searched. Others are not typically found in 'academic' discourse, but have been included because they provide interesting contrasts for subsequent discussion. For each new addition, contexts have been found in which new candidate phrases pass the cue phrase test. The enlarged corpus, containing some 350 phrases, is given in Appendix A. While there are doubtless other phrases still to be included, it is the largest corpus of connective phrases that I am aware of in the literature.

Sections 4.3.1 and 4.3.2 report the results of two preliminary analyses carried out on the corpus.

4.3.1 The Syntactic Diversity of cue Phrases

Cue phrases fall into five syntactic classes (as defined by Quirk *et al* (1972)):

- **Coordinators:** these always appear in between the clauses they link; the clauses can be in separate sentences or in the same sentence. If in the same sentence, no punctuation is required in addition to the coordinator; and if combined in a sequence with other cue phrases, coordinators always appear leftmost in the sequence. For example:

  (4.10) a. An object may move *but* it remains the same object...

  b. A general rule is needed to prevent comparative constructions. *Or* some rule is needed that will say: 'if a word cannot ...

[2] Of course, it might be objected that *if ... [then]* is really the cue phrase, rather than simply *if*; the test for cue phrases does not capture this as it stands. The issue of the inter-dependence of cue phrases is raised in Section 4.4.

- **Subordinators:** these introduce subordinate clauses in complex sentences. The subordinate clause can be on the left or the right of the main clause, but the subordinator is always on the left of the subordinate clause. For example:

  (4.11) a. *Although* it is common sense that labels are related, this is a difficult idea to explicate.
  b. One further illocution should be considered *before* we discuss some variants.

- **Conjunct adverbs:** these modify whole clauses, and can appear at different points within them, although there is often a default position for particular phrases. There are also syntactic constraints on exactly which positions conjunct adverbs can occupy: at the beginning of a clause, between subject and verb, between any auxiliary verbs, between auxiliary verb and main verb, after a copula if there is one, before a sentential complement if there is one. For example:

  (4.12) a. The parallel between permissibility and possibility has been exploited by many linguists. There are, *however*, two important distinctions between them ...
  b. We will select only those hypotheses we deem relevant. *As a consequence,* our discussion differs from the usual views ...

- **Prepositional phrases:** these often contain propositional anaphora referring back to the previous clause. For example:

  (4.13) a. It has a high degree of opacity. *In that respect* it resembles glass.
  b. The plate extends as far as the Pacific coast. *At this point* it slopes down.

  The distinction between prepositional phrases and conjunct adverbials is often hard to make. I have tended to include phrases in the latter category if they are best analysed as idiom chunks, and in the former category if they retain a fair degree of compositionality—see Section 4.3.2 for further details.

- **Phrases which take sentential complements:** these often introduce a particular intentional stance with respect to the content of the clause they introduce. For example:

  (4.14) a. An act that is physically impossible cannot occur. (...) *It follows that* the language used ... is often straightforward.
  b. *It may seem that* we are making too much of orientation; but characteristic orientation is not an idiosyncrasy.

4.3.2 The Space of Cue Phrases

One finding that emerged from a study of the corpus was that the cue phrases fell into two groups. It was possible to envisage a compositional semantics for some phrases: for instance, the semantics of the phrases *many years later*, *a few years later*, and *twenty-five years later* can be seen to depend on the semantics of the constituent phrases *many years*, *a few years*, and *twenty-five years*; and these in turn depend on the semantics of the determiner phrases *many*, *a few*, and *twenty-five*.

Other phrases in the corpus, which we might refer to as simple phrases, are impossible to break down in this way. This might be because they are single words, or alternatively because they are idiom chunks, which are defined precisely as multi-word phrases whose semantics is not compositional. Examples of idiom chunks in the corpus include *on the other hand* (in contrast with the ungrammatical *off the other hand*), *after all* (in contrast with *before all*), and *given that* (in contrast with *taken that*).

A great many phrases seem to be *partly* compositional—for instance, the meanings of *on the one hand* and *on the other hand* can be thought to hinge on the meanings of *one* and *other*, but not on the meaning of *hand*: the phrase *on the other foot* is nonsense as a marker of a relation. There are many other phrases of a similar 'semi-compositional' status; but there seem to be no hard-and-fast rules for working out how such phrases are formed, and it is easiest at the outset simply to treat them as unanalysed atomic elements.

The existence of compositional cue phrases has an important consequence: it makes the class of cue phrases infinite in size. Phrases like *very very ... very many years later* are technically members of the class, even though in practice they will never occur. This means that in order to describe the class, it is necessary to lay down rules for how compositional cue phrases can be constructed. These rules will be syntactic in nature. For instance, the following two rules are helpful in expanding the charted space of cue phrases:

- There is a class of words which modify all subordinators and only subordinators; these words are *even*, *just*, *except*, *only* and *especially*. Thus we can construct cue phrases like *only where*, *except before*, and *just on the grounds that*. There are exceptions to this rule (for instance, **except in case*), but it still provides a useful generalisation.

- Temporal phrases can also be modified in a systematic way. The conjunct adverbs *earlier*, *afterwards* and *later*, as well as the phrases *before* and *after* (which can be conjunct adverbs or subordinators), can all be modified by any expression denoting a length of time; for instance *three days after*, *a minute earlier*, and *some time before*. The modifiers always precede the head phrases.

The general syntactic concepts of **head** and **modifier** can be used to analyse any phrase, regardless of its syntactic category. Compositional cue phrases can typically stand alone without modifiers—for instance, *later* and *after* by themselves are still cue phrases. In what follows, modifiers have been stripped wherever possible. To reiterate the point made in Section 4.2: we are not interested in phrases in which the full compositional power of the language is brought to bear; we are interested in the 'stock' words and phrases, which have evolved to meet specific communicative needs.

4.4 Organising the Corpus: A Test for Substitutability

Thus far, we have used a simple test for detecting cue phrases in text, and on the basis of this we have gathered a corpus of cue phrases. The phrases have been classified according to their syntactic properties, so that an idea can be obtained of the complete space of phrases. But since we are principally interested in cue phrases as signallers of discourse structuring strategies, a classification of phrases according to their function in discourse is our central objective.

In keeping with the data-driven methodology adopted thus far, the classification will be made by means of a simple linguistic test, rather than by making theoretical claims about the semantics or pragmatics of the phrases in the corpus. The test is to do with **substitutability**. Very broadly, if two phrases are inter-substitutable in a passage of discourse then they should be classified in the same category. If one phrase can always be substituted for another, but not *vice versa*, then the latter phrase should be classified in a category subordinate to that of the former phrase. In this way a taxonomy of synonyms and hyponyms can be constructed. It will also be interesting to represent those groups of phrases which can *never* be substituted for each other, and those which can *sometimes* be substituted for each other, in certain contexts.

The approach here is similar in many ways to that taken in the WordNet project (Beckwith *et al* (1990), Miller *et al* (1990)). WordNet is a lexical database organised on psycholinguistic principles: it comprises taxonomies of nouns, verbs and adjectives, which represent various different relationships between words, such as synonymy and antonymy. The classification of cue phrases makes use of different relationships, but the idea of a hierarchical taxonomy of words and phrases is the same.

The test for substitutability is given in detail in Figure 4.2. The main idea is that the tester considers a cue phrase in a context where it naturally occurs, and then considers which other phrases (s)he, *as a writer*, would be prepared to use in its place. This is a task which occurs quite regularly during the course of normal writing. The tester might imagine that the original phrase has been used recently in the preceding discourse, and needs to be changed for reasons of 'elegant variation'.

As it will be seen, the conditions for substitutability are slightly less constrained than those under which one phrase can simply replace another. To begin with, we are not interested in whether two phrases can take the same *grammatical* position in a clause; rather, we are interested in whether they have the same function in signalling discourse relations between the clause and other units. For instance, a conjunctive adverb like *nevertheless* might have the same function as a coordinator like *but*, but the latter can only appear at the beginning of a clause, so simple replacement will not always be possible. In view of this, candidate phrases can be substituted in the clause in a different position, from the original phrase, if necessary.

For another thing, when it comes to comparing the original text with its counterpart containing the candidate phrase, there are some factors which are not taken into account. Stylistic mismatches are ignored; *moreover* is thus a legitimate substitute for *and* in some contexts, even though the latter may be less formal. The following examples of substitutability are therefore legitimate:

$$\text{Grease is the time, is the place, is the motion;} \left\{ \begin{array}{l} \textit{and} \\ \checkmark\ \textit{moreover,} \\ \checkmark\ \textit{furthermore,} \end{array} \right\} \text{Grease is the way you are feeling.}$$

The size of the units of text being linked is also a factor to be disregarded. *Because* tends to connect smaller spans than *this is because*; but other than this, there is little difference between the two phrases. Finally, the amount of background knowledge possessed by the reader is treated as a variable in the test. The phrase *and* can be substituted for the phrase *because*, but only if we assume that the reader can *infer* that a causal relation is being expressed. See Figure 4.2 for further explanation of all these points.

Some Terms Based on the Notion of Substitutability

The test in Figure 4.2 identifies when one candidate cue phrase is **substitutable** for another phrase in a given context. If we generalise over all possible contexts, three different relationships between two cue phrases can be distinguished:

$always(x, y)$ $\Leftrightarrow$ in every context where y appears, x is substitutable for y

$sometimes(x, y)$ $\Leftrightarrow$ x is substitutable for y in some contexts where y appears, but not in all of them

$never(x, y)$ $\Leftrightarrow$ in every context where y appears, x is not substitutable for y.

These three relationships exhaust the possible substitution relationships between x and y: for any pair of cue phrases, exactly one of them holds.

The definitions of *always*, *never* and *sometimes* form the basis of four composite relationships between cue phrases:

x and y are **synonymous** $\Leftrightarrow$ $always(x, y) \wedge always(y, x)$

x and y are **exclusive** $\Leftrightarrow$ $never(x, y)\ (\wedge never(y, x))$

x is a **hyponym** of y (y is a **hypernym** of x) $\Leftrightarrow$ $always(y, x) \wedge sometimes(x, y)$

x and y are **contingently substitutable** $\Leftrightarrow$ $sometimes(x, y) \wedge sometimes(y, x)$

Again, for any pair of cue phrases, exactly one of these relationships holds. The concepts of synonymity, hyponymity/hypernymity, exclusivity and contingent substitutability will be used in most of the discussion which follows.

Examples of Substitutability Relationships

In order to present examples of substitutability relationships, diagrams such as that in 4.20 will be used:

$$\text{First span of text.} \left\{ \begin{array}{l} \textit{C1} \\ \checkmark\ \textit{C2} \\ \checkmark\ \textit{C3} \rightarrow \\ \checkmark\ \textit{(C4)} \\ \#\ \textit{C5} \end{array} \right\} \text{second span of text.} \tag{4.20}$$

1. Consider any cue phrase from the corpus in a text where it naturally occurs. Imagine you are a writer who has just produced this text, but needs to choose an alternative phrase (perhaps because you have just used the original phrase, and do not want to repeat it).
2. Remove the cue phrase from its host clause, and insert any other phrase from the corpus (the **candidate phrase**) into the same clause, at any appropriate position.
3. If need be, the punctuation of the new discourse can be altered to make it more suitable for the candidate phrase. For example, if the phrase *so* is being replaced by the phrase *this implies that*, it may be necessary to replace a comma with a full stop and create a new sentence.
4. If need be, the new discourse can be supplemented with additional or alternative cue phrases in other clauses. There are sometimes dependencies between the cue phrases in a text (for instance between *if* and *then*, or between *either* and *or*), so changing one phrase might require changes to others.
5. If it is possible to use the resulting discourse in place of the original discourse, then the candidate phrase is said to be **substitutable** to the original phrase in that context.

   The notion of 'being able to use one discourse in place of another' is expanded below.

   - It is not sufficient that the new discourse can be used to describe the same set of events in the world as those which the old one describes. For instance, the adverb *afterwards* and the subordinator *before* are truth-functionally equivalent in that they are both suitable for describing two events in temporal succession. But they are not always equally appropriate:

     (4.15) Bill was always interested in books. He could read *before* he could walk.

     (4.16) Bill was always interested in books. He could read; *afterwards* he could walk.

     In addition to describing the same eventualities, it must be ensured that the new discourse achieves the same goals as the old discourse achieved.
   - Some differences between the two discourses can nevertheless be overlooked—for one thing, *stylistic* discrepancies can be disregarded. The cue phrase *hence* can often be substituted for the cue phrase *so*, the only difference being in the 'formality' of the resulting discourse:

     (4.17) I'm just back from a holiday in France so there'll be no need to bring wine.

     (4.18) I have just returned from France; hence there will be no need to bring wine.

     Stylistic changes to the new text may thus be needed in order to accommodate the candidate phrase.
   - Different cue phrases are appropriate for linking portions of text of different sizes. For instance, *because* typically links clauses within a compound sentence; *this is because* typically links whole sentences. Such differences are to be overlooked in the test. In some cases, changing the punctuation is sufficient to accommodate the candidate phrase; but in others, it might be necessary to alter the length of the spans of text being linked, by substituting a précis or by adding additional relevant material.
   - A final factor to be disregarded is the amount of *background knowledge* the reader is assumed to possess. For instance, *and* can often be substituted for *yet*, but only if the reader will be able to infer the appropriate contrastive relation. Consider Example 4.19:

     (4.19) Mike was ravenous, yet he ordered watercress salad.

     Here, a substitution by *and* is only permissible if we can assume that the reader knows that petit fours are snacks, and hence that Mike's order comes as a surprise.

Figure 4.2: The Test for Substitutability

The items between the braces are all cue phrases. $C1$ is the original cue phrase; $C2$ is a phrase which in this context is substitutable for $C1$. $C3$ is substitutable for $C1$ in the given context, but it must first be moved to a different position in $C1$'s clause (in this case to the right). $C4$ is also substitutable for $C1$ in the given context, but it requires either a change to one of the other cue phrases in the text (due to a dependency between cue phrases), or a change to the size of the spans involved (due to the suitability of different cue phrases to spans of different sizes). All of these changes are permitted by the test for substitutability. Finally, $C5$ is not substitutable for $C1$ in the given context, even allowing for the changes the test allows.[3]

It should be noted that the text on either side of the braces can in principle be as long as is needed to make the original context clear. In practice, one or two clauses' worth of context will normally be given on each side. As was noted in Section 1.1.2, the idea of presenting the context of a text 'in its entirety' is problematic; however, it is hoped that the contexts provided in the examples which follow will be sufficient to give the reader a good idea of the texts.

A few examples of the test for substitutability can now be given. For instance, the phrases *to start with* and *to begin with* are intersubstitutable in all contexts, and hence termed synonymous: two examples of their intersubstitutability are given in texts 4.21 and 4.22.

$$\text{Cyril set about preparing breakfast.} \left\{ \begin{array}{l} \textit{To start with,} \\ \checkmark\ \textit{To begin with,} \end{array} \right\} \text{he put some porridge on to boil. Next, he set out four bowls...} \qquad (4.21)$$

$$\text{Sid's got no hope of winning the race.} \left\{ \begin{array}{l} \textit{To begin with,} \\ \checkmark\ \textit{To start with,} \end{array} \right\} \text{he's out of training. For another thing he'll be running against Otto Schultz, who hasn't lost all season...} \qquad (4.22)$$

In the case of *to start with* and *for a start*, the relationship is not bidirectional: the former phrase is more general than the latter.

$$\text{Sid's got no hope of winning the race.} \left\{ \begin{array}{l} \textit{To start with,} \\ \checkmark\ \textit{For a start,} \end{array} \right\} \text{he's out of training. For another thing he'll be running against Otto Schultz, who hasn't lost all season...} \qquad (4.23)$$

$$\text{Cyril set about preparing breakfast.} \left\{ \begin{array}{l} \textit{To start with,} \\ \#\ \textit{For a start,} \end{array} \right\} \text{he put some porridge on to boil. Next, he set out four bowls...} \qquad (4.24)$$

In texts such as 4.24, *for a start* does not seem an appropriate substitution—it gives the text an argumentative tone which is lacking in the original, which is purely narrative. From examples such as these, we can conclude that *to start with* is a hypernym of *for a start*.

[3] It should be borne in mind that the text may still be *grammatical* with $C5$; it may even still *make sense* with $C5$. The point is that just $C5$ cannot be used as a *replacement* for $C1$ in the context. The hash sign used to indicate non-substitutability must therefore be interpreted somewhat differently from hash signs as conventionally used in linguistic examples, which often denote 'ill-formed discourse'.

The phrases *lastly* and *moreover* are contingently substitutable. In some contexts they are both appropriate:

$$\text{Sid's got no hope of winning the race. For one thing, he's out of training. For another thing, he's best at altitude, and he'll be running at sea level.} \left\{\begin{array}{l}\textit{Moreover,}\\ \checkmark\,\textit{Lastly,}\end{array}\right\} \text{he's pitted against Otto Schultz, who hasn't lost all season.} \tag{4.25}$$

But in some contexts, *lastly* cannot be replaced by *moreover*:

$$\text{Cyril set about preparing breakfast. To start with, he put some porridge on to boil. Next, he set out four bowls.} \left\{\begin{array}{l}\textit{Lastly}\\ \#\ \textit{Moreover}\end{array}\right\} \text{he sliced some bread ready for toasting.} \tag{4.26}$$

And in other contexts, *moreover* cannot be replaced by *lastly*:

$$\text{Sid's got no hope of winning the race. For one thing, he's out of training.} \left\{\begin{array}{l}\textit{Moreover}\\ \#\ \textit{Lastly}\end{array}\right\} \text{he's best at altitude, and he'll be running at sea level. In addition, he's pitted against Otto Schultz, who hasn't lost all season.} \tag{4.27}$$

When applying the test for substitutability, a question arises as to how subtle we should be in distinguishing between cue phrases. It is often noted that 'true synonyms' are extremely rare: indeed, in some of the above examples where substitutability is claimed, one phrase might appear slightly more appropriate to some readers even though no particular reason suggests itself. Typically, a *rule* can be envisaged which relates various features of a text to the cue phrases which are most appropriate. But in a context where both phrases are acceptable, one being just marginally better than the other, generalisations are often hard to make: In such cases, we will err on the side of generality, and allow that substitutability is possible. If, subsequently, we are able to find a reliable rule, of course this decision can be reversed, and subtler distinctions made. But it should be borne in mind that we are principally concerned with making broad classifications within the set of cue phrases, rather than descending into the minutiae of 'descriptive linguistics'—and the test for substitutability is perfectly adequate for this task.

4.5 Substitutability Diagrams

In this thesis, a diagrammatic representation of substitutability relationships is used. The diagrammatic notation allows information about many pairs of cue phrases to be presented simultaneously, in a form which is relatively easy to understand.

The diagrams consist of **nodes** containing (possibly empty) sets of cue phrases, connected by a structure of directed **arcs**. Figure 4.3 shows the simplest structural relationships that can exist between two nodes A and B: $hypo(A, B)$, $excl(A, B)$ and $cs(A, B)$. Informally speaking:

- If $hypo(A, B)$, then phrases which are in A are hyponyms of phrases in B.

- If $excl(A, B)$, then phrases which are in A are exclusive with phrases in B.
- If $cs(A, B)$, then phrases which are in A are contingently substitutable with phrases in B (provided no other relationship between A and B is documented—see Section 4.5.1).
- Phrases which appear at the same node are synonymous.

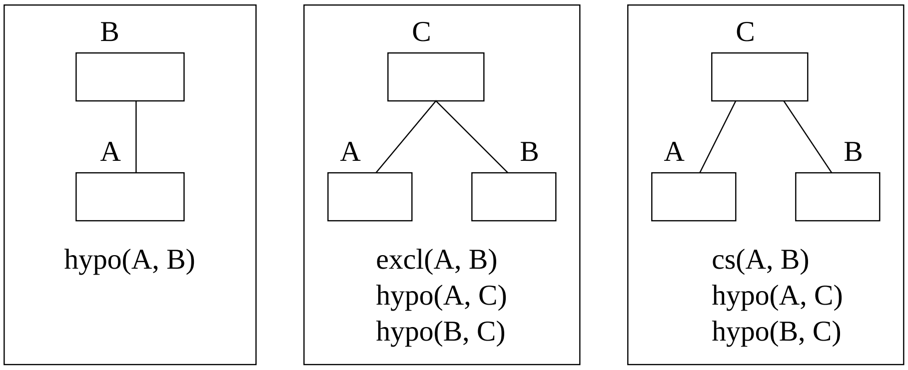

Figure 4.3: Three Possible Structural Relationships Between Nodes

Note that the *excl* and *cs* relationships apply between 'sister' nodes; i.e. nodes which have a common mother node. The difference between $excl(A, B)$ and $cs(A, B)$ relates to whether or not their arcs meet on the mother node. This notation is chosen to allow the representation of hypernyms shared between exclusive or contingently substitutable phrases.

4.5.1 Contingent Substitutability Relationships

The relationship of contingent substitutability is overridden by other relationships in diagrams where a conflict is present. For instance, in Figure 4.4, x and y are represented as exclusive (through the arcs that touch on the mother node), but also as contingently substitutable (through the arcs which do not touch). In cases where a conflict such as

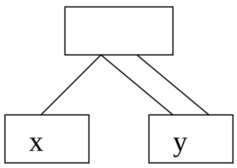

Figure 4.4: An Overridden Contingent Substitutability Relationship

this one is present, the contingent substitutability relationship is overridden. (In this example, of course, the additional arc representing contingent substitutability does no useful work, and it would be much clearer to leave it out; the diagram is just used as a simple illustration of how a contingent substitutability relationship can be overridden. But see the following section for cases where overridden contingent substitutability relationships do have a useful role.)

4.5.2 Complex Substitutability Diagrams

Complex substitutability diagrams involving many cue phrases can be created by combining the structures presented in Figure 4.3. These diagrams make use of inheritance: the phrases in a daughter node inherit the exclusivity and hyponymity relationships of the phrases in their mother node. Thus in Figure 4.5 (i), z is a hyponym of x, so by inheritance, z is a hyponym of w and exclusive with y.

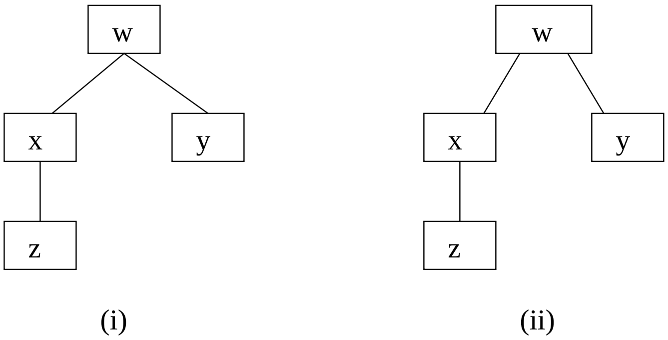

Figure 4.5: Two Examples of Inheritance

Contingent substitutability relationships are also inherited from phrases in a daughter node to phrases in its mother node. Thus, in 4.5 (ii), z and y are contingently substitutable. Note again, however, that these relationships can be overridden if they conflict with other relationships. Thus, in Figure 4.6, while x is contingently substitutable with y, z does not inherit this property because it is explicitly shown to be exclusive to y. The inherited contingent substitutability relationships between x and z, and between z and itself are also overridden.

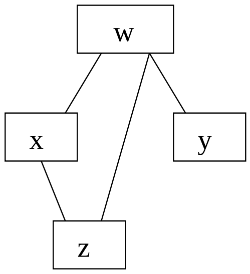

Figure 4.6: Overridden Inherited Contingent Substitutability Relationships

There are, however, cases where inheritance in a substitutability diagram causes genuine contradictions. For instance, Figure 4.7 is an illegal diagram: z is represented by inheritance as exclusive with itself; and exclusivity cannot be overridden.

Substitutability diagrams are intended to represent the relationship between *each pair* of phrases which appear in it—in other words, to provide all the substitutability information that it is possible to provide about the phrases involved. A final requirement to this end is to specify that diagrams must have a single top node. The diagram in

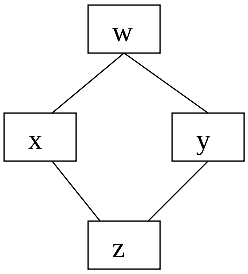

Figure 4.7: An Illegal Substitutability Diagram

Figure 4.8 is not permitted, because it does not document the relationship between x and y. Requiring that a single node dominates both x and y ensures that their

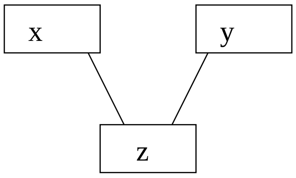

Figure 4.8: Another Illegal Substitutability Diagram

relationship will be represented.

4.5.3 Formalising the Semantics of Substitutability Diagrams

A set of rules for determining the relationships between phrases in a (legal) substitutability diagram is given in this section. The rules draw on the definitions of *always*, *never* and *sometimes*, which are re-iterated below:

$always(x, y)$	$\Leftrightarrow$	in every context where y appears, x is substitutable for y
$sometimes(x, y)$	$\Leftrightarrow$	x is substitutable for y in some contexts where y appears, but not in all of them
$never(x, y)$	$\Leftrightarrow$	in every context where y appears, x is not substitutable for y.

The following rules are for deriving new substitutability relationships from existing ones:

$$\begin{aligned}
always(x, y) \wedge never(x, z) &\Rightarrow never(y, z) \\
always(x, y) \wedge always(y, z) &\Rightarrow always(x, z) \\
never(x, y) &\Rightarrow never(y, x) \\
sometimes(x, y) &\Rightarrow sometimes(y, x) \\
always(x, y) \wedge sometimes(x, z) \wedge \neg always(z, y) \wedge \neg never(y, z) &\Rightarrow sometimes(y, z)
\end{aligned}$$

The following rules are for deriving substitutability relationships from structures in a substitutability diagram. (They should be seen as replacements for the informal definitions given in Figure 4.3.)

$$\begin{aligned} x \in A \wedge y \in A &\Rightarrow always(x, y) \\ x \in A \wedge y \in B \wedge hypo(A, B) &\Rightarrow always(y, x) \\ x \in A \wedge y \in B \wedge excl(A, B) &\Rightarrow never(x, y) \\ x \in A \wedge y \in B \wedge cs(A, B) \wedge \neg always(x, y) \wedge \neg never(x, y) &\Rightarrow sometimes(x, y) \\ x \in A \wedge \neg(y \in A) \wedge always(x, y) &\Rightarrow sometimes(y, x) \end{aligned}$$

The intended definition of $sometimes(x, y)$ relies on a closed-world assumption about *always* and *never* relationships. In order to compute the complete set of relationships in a diagram, all the *always* relationships should first be computed, then all the *never* relationships, and finally the *sometimes* relationships.

4.5.4 Empty Nodes

Some nodes in a diagram do not contain any cue phrases at all. At the very top of the hierarchy, an empty category is necessitated by the formalism chosen for depicting substitutability relationships: if two phrases are exclusive or contingently substitutable, a common superordinate category must be shown whether or not they have a common hypernym. We can use the graph-theoretical category **top** (or $\top$) to fulfill this purpose (see Figure 4.9).

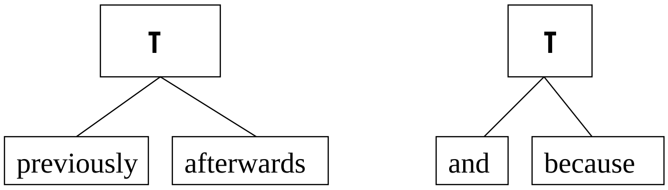

Figure 4.9: Two Uses of the Empty 'Top' Category

Empty nodes can also appear lower down in the taxonomy. These are not essential—a diagram can always be redrawn without them—but they often make diagrams easier to read. Imagine we have three phrases, X, Y and Z. X and Y are exclusive, so are X and Z; but Y and Z are contingently substitutable. Figure 4.10 shows two alternative ways of representing all these relationships: in many cases, the method involving the empty category is neatest.

4.6 The Taxonomy of Cue Phrases

The central task is now to incorporate as many of the cue phrases as possible into a single substitutability diagram. Ideally, the aim is to document the substitutability relationship between each pair of phrases in the corpus. Of course, this is a huge number; assuming there are N phrases in the corpus, the total number of relationships

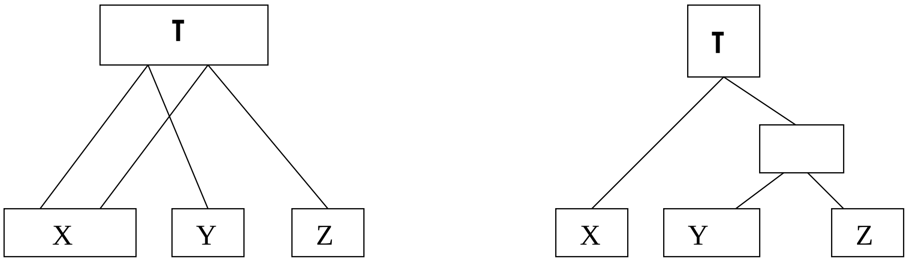

Figure 4.10: Empty Categories Lower Down in the Taxonomy

documented will be $(N * (N - 1))/2$. For $N \approx 330$, as in the present case, the number of relationships will be over 50,000. Clearly it makes sense to begin by looking for all the relationships within a *subset* of the complete corpus.

Currently, around 150 phrases have been incorporated into a single substitutability diagram, documenting some 11,000 relationships, which will be referred to as the **taxonomy** of cue phrases. The complete diagram is given in Appendix B, along with details of its organisation, and copious examples to motivate it. The following sections present an extract from the taxonomy, and summarise some of its most important characteristics.

4.6.1 Construction of the Taxonomy

As with the corpus of cue phrases, the taxonomy of cue phrases was constructed entirely by the author. It would have been preferable to construct the taxonomy on the basis of the judgements of a sizeable group of people (ideally, people without any theoretical experience of discourse analysis, given that the substitutability test is designed to recreate a task that forms a part of ordinary writing). However, the amount of data needed in order to build a taxonomy of any reasonable size from scratch makes such an experiment quite infeasible, bearing in mind the huge number of relationships that must be documented. Instead, the decision was taken to build a taxonomy reflecting the author's own intuitions, which could then be used and tested more systematically in subsequent experiments on groups of naive readers and writers. Such experiments have yet to be carried out; however, they would be very valuable as a follow-up to the present study.

The amount of data required to build the taxonomy also dictated that most of the examples used to motivate substitutability relationships were hand-crafted. It would have been preferable to search for appropriate examples in a corpus, but again, this would have been prohibitively time-consuming. A corpus-based study would certainly shed useful light on the taxonomy as currently constructed; but again remains to be pursued in follow-up work.

4.6.2 An Extract from the Taxonomy

A small portion of the taxonomy, dealing with some of the phrases which signal position in a sequence, is given in Figure 4.11. Pre-theoretical titles have been assigned to some

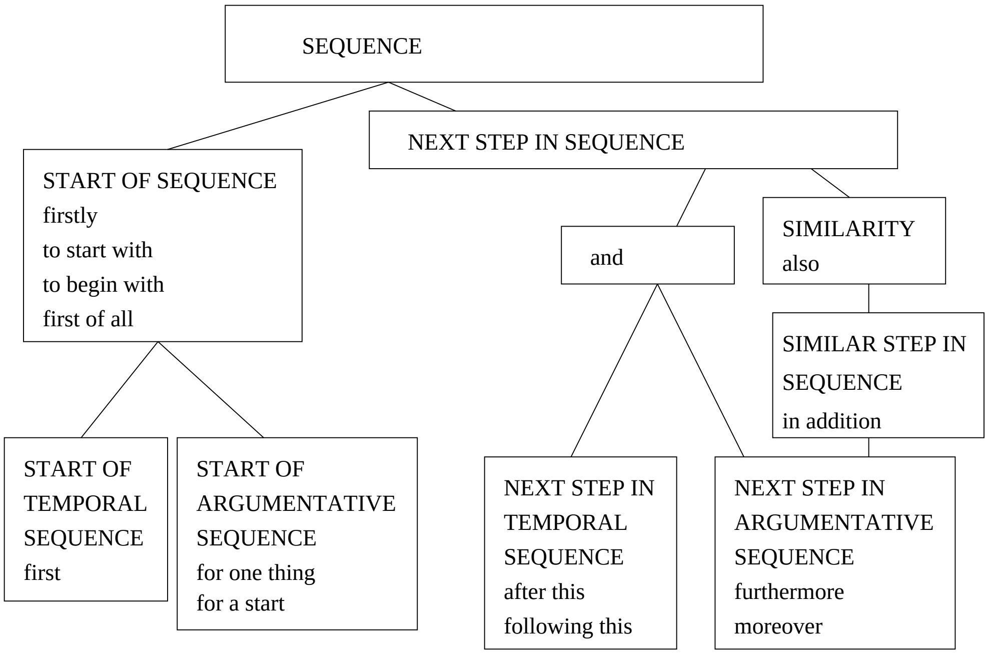

Figure 4.11: A Portion of the Taxonomy of Cue Phrases

of the categories; but this is just to give some idea of what functions the phrases might have: the taxonomy still just represents substitutability information.

Some examples of the substitutability relationships in Figure 4.11 are given below.

$$\text{Bob set about cleaning the house.} \left\{ \begin{array}{l} \textit{To start with,} \\ \checkmark\ \textit{To begin with,} \\ \checkmark\ \textit{First,} \\ \#\ \textit{For one thing,} \\ \#\ \textit{Furthermore,} \\ \#\ \textit{And,} \end{array} \right\} \text{he swept the floors; next he washed them; and lastly, he tidied the cupboards.} \tag{4.28}$$

$$\text{Bob set about cleaning the house. To start with, he swept the floors and washed them;} \left\{ \begin{array}{l} \textit{in addition,} \\ \checkmark\ \textit{after this,} \\ \checkmark\ \textit{following this,} \\ \checkmark\ \textit{(and)} \\ \checkmark\ \textit{also} \rightarrow \\ \#\ \textit{For one thing,} \\ \#\ \textit{Furthermore,} \end{array} \right\} \text{he tidied the cupboards.} \tag{4.29}$$

$$\text{Television is bad for us. It kills creativity;} \left\{ \begin{array}{l} \textit{and} \\ \checkmark\ \textit{furthermore,} \\ \checkmark\ \textit{also,} \\ \checkmark\ \textit{moreover} \\ \#\ \textit{after this} \\ \#\ \textit{for one thing} \end{array} \right\} \text{it promotes an unhealthy kind of `crowd mentality'.} \tag{4.30}$$

$$\text{Television is bad for us.} \left\{ \begin{array}{l} \textit{For one thing,} \\ \checkmark\ \textit{To begin with,} \\ \checkmark\ \textit{Firstly,} \\ \#\ \textit{Moreover} \\ \#\ \textit{And} \\ \#\ \textit{First} \end{array} \right\} \begin{array}{l} \text{it kills creativity; in} \\ \text{addition it promotes an} \\ \text{unhealthy kind of `crowd} \\ \text{mentality'.} \end{array} \tag{4.31}$$

$$\begin{array}{r} \text{Jim jumped off the cliff, so} \\ \text{Bill} \end{array} \left\{ \begin{array}{l} \textit{also} \\ \#\ \textit{and} \\ \#\ \textit{in addition} \\ \#\ \textit{furthermore} \\ \#\ \textit{for one thing} \end{array} \right\} \text{jumped off.} \tag{4.32}$$

4.6.3 Some General Remarks about the Taxonomy

Two of the taxonomy's most significant characteristics should be mentioned straight away, as they have an important bearing on its organisation, and on the theoretical interpretation it will subsequently be given.

For one thing, a degree of hierarchy is found throughout the taxonomy. Chains of 2 or 3 hyponymic nodes are fairly common. The 'most general' cue phrase is *and*, which has over 30 separate hyponyms. In other words, the degree of generality of cue phrases in the taxonomy is an interesting variable to study.

Another important finding is that the taxonomy does not divide neatly into large exclusive subgroups of phrases. For any candidate grouping, many phrases can be found which fit into more than one group. (*And* and *then*, for instance, have many other uses aside from signalling position in a sequence.) In fact, most of the variation between cue phrases is represented at a relatively low level, in the microstructure of the taxonomy. This is interesting, because it already suggests that a classification scheme based on the taxonomy is unlikely to identify any one dimension of variation amongst relations as 'dominant'—an assumption which is characteristic of many existing classifications of relations.

4.6.4 The Global Organisation of the Taxonomy

The task of representing *all* the substitutability relationships between *all* the phrases in the corpus is an extremely complex one. One of the main difficulties is the fact just alluded to, that the phrases in the corpus do not separate neatly into exclusive categories. This lack of modularity makes it difficult to work with a subset of the phrases in isolation.

To solve this problem, phrases are organised at a high level into a number of *non-exclusive* categories: SEQUENCE PHRASES, CAUSE PHRASES, RESULT PHRASES, RESTATEMENT PHRASES, TEMPORAL PHRASES, NEGATIVE POLARITY PHRASES, ADDITIONAL INFORMATION PHRASES, HYPOTHETICAL PHRASES, SIMILARITY PHRASES, and DIGRESSION PHRASES. These categories have no theoretical significance at all, and should just be thought of as providing an expedient way for spreading the taxonomy over several pages. Two types of cue phrase are then identified: **exclusive phrases**, which belong just to one category; and **multicategory phrases**, which belong to two or more categories. This distinction introduces a certain amount of modularity into the taxonomy, and greatly reduces its complexity. To begin with, there is a diagram showing the exclusive relationship between the 'exclusive phrases' in every category. Then a substitutability diagram is given for each separate category, showing the relationships between its exclusive phrases, and between its exclusive phrases and all the multicategory phrases. Finally, there is a substitutability diagram for all the multicategory phrases. In this way it is ensured that the relationship between each phrase and each other phrase is represented.

The reader is referred to Appendix B for a closer look at the taxonomy. It is not yet perfect, of course: there are still cue phrases in the corpus which have not been incorporated; and it is still not very hard to find counterexamples to some of the relationships it documents. But at least a reasonably clear method exists for querying and improving it: questions about the placing of a given phrase will be decided on the evidence of concrete linguistic data.

4.7 Summary

This chapter has described the incremental construction of a taxonomy of cue phrases. Initially, a **test for cue phrases** was employed to gather a corpus of cue phrases from naturally occurring texts. Then this corpus was organised into a taxonomy, using a **test for substitutability**. Using these two tests, it should be possible for several people to arrive at very similar taxonomies of cue phrases. And where there are discrepancies, the reliance in both tests on concrete linguistic examples should provide a convenient way for alternative analyses to be discussed.

The next chapter shows how the taxonomy of cue phrases can be used to motivate an isomorphic taxonomy of coherence relations.

Chapter 5

Preliminaries for Defining a Set of Relations

In this chapter, a framework for using the taxonomy to create relation definitions is proposed. A fundamental theoretical assumption about the taxonomy is made—that it lends itself to a conception of relations as composite constructs, made up of a number of independent **features**. This conception is argued for in Section 5.1: the argument turns on the existence of hypernymic cue phrases in the taxonomy.

In Section 5.2 some general principles governing the motivation of features are proposed. For one thing, the features chosen must be sufficient not only to distinguish between the different cue phrases in the taxonomy, but also to determine all and only those contexts where a given cue phrase can be used. For another thing, the set of features eventually chosen must be **productive**; that is, the alternative values of each feature must make sense in combination with all the other possible values of all the other features.

Section 5.3 presents the bare bones of the knowledge representation system to be used for expressing feature definitions. This system is able to represent the text itself, eventualities in the world, and the goals and beliefs of the reader and writer.

5.1 Relations as Feature-Based Constructs

As already noted in Section 4.6.3, the taxonomy is very much a hierarchical structure: hypernymity and hyponymity relationships amongst cue phrases are common. Some cue phrases are much more general in their application than others: a phrase like *and* is substitutable for a wide range of more specific phrases, such as *next*, *whereas* and *thereby*. A question which immediately raises itself is, why do such 'hypernymic' cue phrases exist? Our initial rationale for looking at cue phrases was to explore the linguistic resources for making relations explicit. But if *and* is appropriate in such a wide range of cases, it cannot serve to make any one relation explicit. Why, then, do we find hypernymic phrases like *and* in the taxonomy?

5.1.1 Cue Phrases for Signalling Components of Relations

We can begin by considering the case of the hypernymic cue phrase *and*. It should be noted that even this most widely applicable of cue phrases is by no means an appropriate substitute for *all* other cue phrases. There are many phrases for which *and* cannot be substituted, such as *for example*, *before* or *because*. Indeed, *and* is only an appropriate substitute for a fairly small subset of the overall corpus of cue phrases. It thus goes some of the way to making a relation explicit. Nevertheless, we are still faced with the question of why writers do not always choose to be as explicit as possible when they select a cue phrase.

One answer to this question can be given by appealing to Grice's 'maxim of quantity' (Grice (1975)), which states that contributions to discourse should be as informative as required, but no more informative than required. If we think of relations as *composite* constructs, then it is possible to imagine a situation where some **features** of a relation are easily inferrable by a reader from context or background knowledge, while others are not thus inferrable. Using the most specific cue phrase in such a situation would violate the maxim of quantity, by providing some information twice. But if there were a cue phrase which specified *just those features of a relation which were not inferrable*, the maxim of quantity would not be violated.

To give a concrete example, consider the text in 5.1:

(5.1) It was time for punishment to be meted out. Bob decided on the cat o' nine tails, *and* Frank took his place at the gizzern.

Here, because the reader does not know what a gizzern is, or who Bob and Frank are, the hypernymic cue phrase *and* is insufficient to specify one particular relation, and the text is ambiguous. For instance, the reader does not know whether Bob and Frank are both to be punished and have opted for different punishments, or whether Bob is charge of deciding on a punishment for Frank. The two readings are much more clearly distinguished if more specific cue phrases are used:

(5.2) Bob decided on the cat o' nine tails, *so* Frank took his place at the gizzern.

(5.3) Bob decided on the cat o' nine tails, *whereas* Frank took his place at the gizzern.

On the other hand, if the relation is clear from context, a general cue phrase is acceptable:

(5.4) The captain decided on the cat o' nine tails, *and* Frank wrapped his arms round the mast in readiness...

If a much more specific cue phrase is used, then the same information is effectively provided twice, violating the maxim of quantity, and resulting in a text which is (at least) stylistically awkward:

(5.5) The captain decided on the cat o' nine tails. *As a result*, Frank wrapped his arms round the mast in readiness...

The situation can be compared to that of referring expressions. Why do writers ever use pronouns? Why do pronouns exist at all? They are often ambiguous, and it is surely asking for trouble to introduce them. The answer, again, seems to be that writers try to avoid repeating information that can be easily inferred by the reader. Language contains the resources to create referring expressions with subtle variations in explicitness. (Consider the following expressions: *a big pot*, *a big one*, *the pot*, *such a pot*, *it*.) In a similar way, the hierarchical structures in the taxonomy permit relations to be signalled more or less explicitly.[1]

5.1.2 Using Features to Explain Patterns in the Taxonomy

The taxonomy lends itself well to a conception of relations as feature-based constructs. We can think of cue phrases as signalling particular **values** for some features of a relation, and as being **undefined** for other features (that is, it does not matter which value they take). On this basis, the different substitutability relationships between phrases can each be given a plausible feature-theoretic interpretation:

- If two phrases are **synonymous**, then they are defined for exactly the same set of features, and take the same values for all of these features. This explains why they can always be used in the same contexts.
- If two phrases are **exclusive**, then they are defined for at least one feature in common, and signal different values of this feature. This inconsistency explains why they can never be used in the same contexts.[2]
- If phrase X is a **hypernym** of phrase Y (and Y is a **hyponym** of X), then Y is defined for all of the features for which X is defined, and takes the same values for all of these; in addition, Y is defined for at least one other feature, for which X is not defined. Y, being more tightly defined, can only be used in a subset of the contexts in which X is appropriate.
- If X and Y are **contingently substitutable**, then they are both defined for some set of features, and signal the same values for these features; in addition, Y is defined for some feature f_1 for which X is not defined, and X is defined for some feature f_2 for which Y is not defined. Thus X can be used in some contexts where Y cannot (those which take the wrong value of f_1), Y can be used in some contexts where X cannot (those which take the wrong value of f_2), and in other contexts (taking suitable values for both f_1 and f_2), X and Y can be substituted.

[1] Dale and Reiter (1992) have pointed out that when people generate referring expressions, they do not conform completely to Gricean maxims: they are often more explicit than they have to be. This may well be the case in the way people signal relations too. However, the maxim of quantity does at least offer an explanation of why people are not *always* maximally explicit, and this is the important point for present purposes.

[2] Note that feature values are not typed in the system I am proposing, and so different values do not unify. See Section 5.2.4 for a more detailed account of how this system differs from typed feature hierarchies in the tradition of ALE (Carpenter and Penn (1994)).

For a summary of these interpretations, see Figure 5.1.

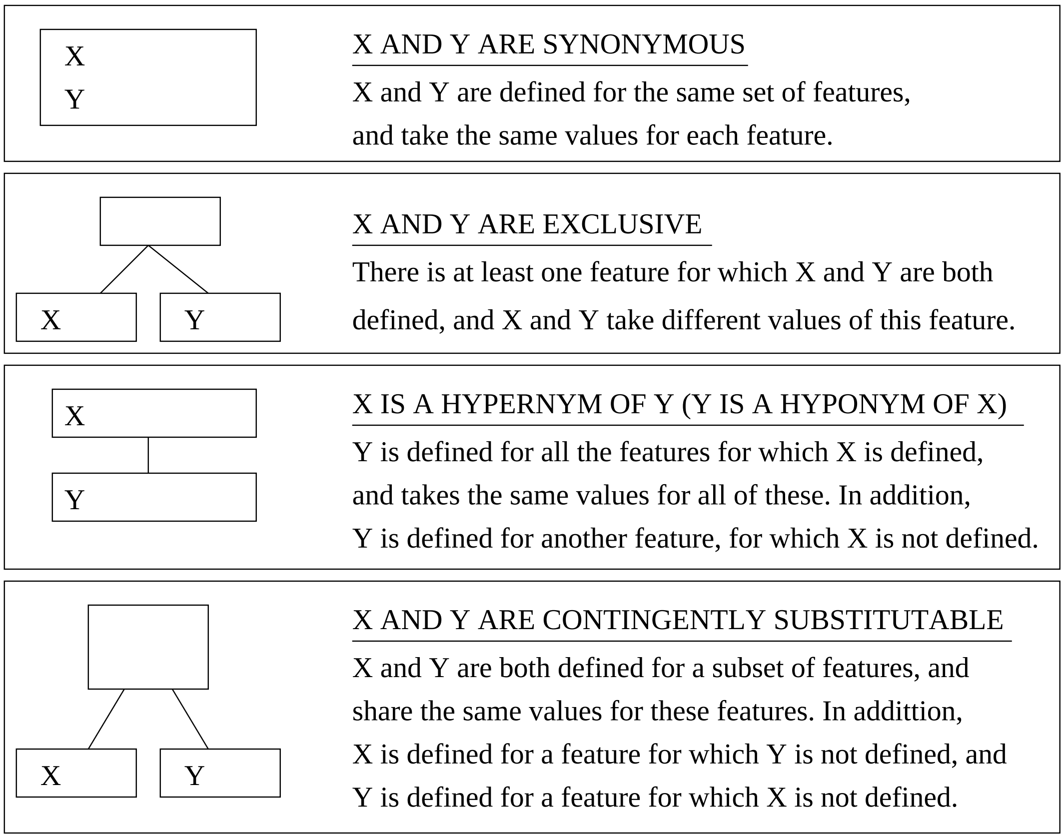

Figure 5.1: Feature-Theoretic Interpretations of Substitutability Relationships

The concept of inheritance can likewise be given a feature-theoretic interpretation. The taxonomy can be seen as an **inheritance hierarchy for feature values**:[3] phrases in a daughter category inherit all of the feature values associated with its mother phrases, and in addition are defined for new features. We can also give an account of why it is that inherited contingent substitutability relationships can be overridden. Consider Figure 5.2. Phrase D inherits all the feature values of B (that is to say, f_1+ and f_2+); in addition, it is defined for a new feature, f_3, for which it signals the value $-$. Phrase C also inherits f_1+ from A, and is also defined in addition for f_3, but it signals the alternative value $+$. B and C are contingently substitutable, since they are defined for different features, but D does not inherit contingent substitutability with C because they signal alternative values of the same feature.

[3] The notion of 'feature inheritance' is also quite different from that used in typed feature hierarchies. gain, see Section 5.2.4 for an account of how the formalisms differ.

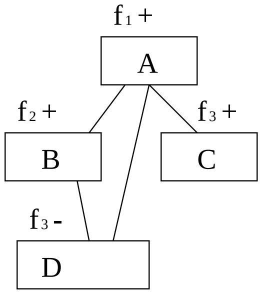

Figure 5.2: A Feature-Theoretic Account of Phrases A, B, C and D

5.1.3 Sanders *et al*'s Study: Independent Evidence for a Feature-Theoretic Approach

To sum up: the taxonomy seems to sanction a conception of relations as decomposing into a number of orthogonal features. It is interesting to note that Sanders *et al*'s (1992, 1993) studies come to the same conclusion—particularly so since these studies are also based on a psychological conception of relations. Note, however, that Sanders *et al*'s evidence for a set of independent features is of a quite different kind from that being presented here. For Sanders *et al*, the decomposition is initially advanced as a psychologically plausible hypothesis. This hypothesis is subsequently supported by an analysis of the disagreements amongst writers about which cue phrases should be used in a given linguistic context (see Section 3.4.1); it was found that where there was disagreement, it tended to be over the value of just one feature. In the present work, the feature-theoretic account is not supported by an analysis of disagreements or by a priori psychological arguments; it simply emerges as a useful way of describing the taxonomy of cue phrases. Nonetheless, both lines of evidence support the hypothesis that readers and writers are able to treat cue phrases, and the representations which underlie them, as composite constructs.

The question remains whether the *same* set of features will be motivated using the two methodologies. This question is currently being pursued in joint work (Knott and Sanders (1996)), which involves the construction of a taxonomy of Dutch cue phrases.

5.1.4 The Remaining Tasks

In this section, we have seen how the linguistic data held in the taxonomy of cue phrases can be given a theoretical interpretation. The distribution of cue phrases is now being used to come to decisions about the nature of the relations they signal: relations, unlike cue phrases, are theoretical constructs.

Note that the distribution of features in the taxonomy is already to a large extent predetermined by the substitutability relationships within it. It would now be possible to go through the whole taxonomy, labelling the categories with a set of features $f_1 \ldots f_n$; these features would be sufficient to distinguish between the different sets of cue phrases. However, the questions of how these anonymous features should be defined, and how to avoid redundancy in the description they afford, are as yet unaddressed. The following section outlines some guidelines to be followed in pursuing this task.

5.2 General Guidelines for Defining Features

Thus far in the methodology, the operations have been fairly replicable. The test for cue phrases and the test for substitutability can be systematically applied, and it is likely that different analysts will come up with very similar taxonomies for a single given language. Likewise, a description of the taxonomy in terms of a set of unknown features $f_1 \dots f_n$ can be achieved automatically. However, the final step of creating *definitions* for these features is less tightly constrained. When deciding how to distinguish between two exclusive cue phrases, for instance, several different features may seem appropriate. In this section, a number of guiding principles concerning this decision are outlined.

5.2.1 Some Substitutability Relationships to be Ignored

'Swap-Substitutable' Phrases

Consider the extract in Figure 5.3. *Since* and *so* are rightly represented as exclusive:

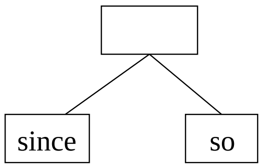

Figure 5.3: Two 'Swap-Substitutable' Phrases

clearly they can never be substituted for one another. In both of the following two examples, the transition from one phrase to another involves a clear change of meaning:

$$\begin{array}{r}\text{Jim had a lot of money on} \\ \text{him that day,}\end{array} \left\{ \begin{array}{l} \textit{so} \\ \#\ \textit{since} \end{array} \right\} \text{he went shopping.} \tag{5.6}$$

$$\text{Jim went shopping,} \left\{ \begin{array}{l} \textit{since} \\ \#\ \textit{so} \end{array} \right\} \begin{array}{l}\text{he had a lot of money on} \\ \text{him that day.}\end{array} \tag{5.7}$$

However, as the examples show, the two phrases can be used in the same context: they are only exclusive because *since* must be attached to one clause, and *so* to the other. When it comes to motivating a set of coherence relations, such differences should be ignored: they do not signal two distinct coherence relations, but merely two distinct ways of *marking* a single relation.

We can refer to such pairs of phrases as **swap-substitutable**. The original substitutability test specified that the candidate phrase had to be inserted into the same clause as the original phrase. But if we relax the substitutability test a little more, and allow the candidate phrase to be inserted into *either* of the related clauses, then phrases like *since* and *so* can be successfully substituted for one another.

There are many other examples of this kind of substitutablility: for instance, *on one hand* and *on the other hand* are swap-substitutable; so are *firstly*, *secondly* and *thirdly*. It will not be necessary to give a feature-theoretic interpretation of the exclusivity relationships between these phrases. Rather, we shall assume that the phrases signal exactly the same feature combinations.

'Disjunctive' Phrases

Up to now, we have been thinking of hypernymic cue phrases as phrases which isolate certain features of a relation. If a cue phrase is ambiguous, then it signals *some* of the features of relations associated with its hyponyms; and if it is general, then it isolates *all* of the features of some more abstract relation. However, there is one other possibility: some ambiguous cue phrases do not pick out a single subset of features, but a *disjunction* of sets of features. Consider the phrases in Figure 5.4: *Since*'s hyponyms

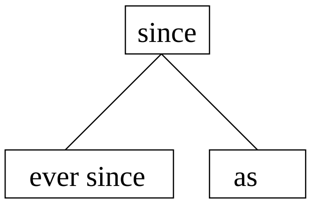

Figure 5.4: A Disjunctive Cue Phrase and its Hyponyms

are phrases with completely different characteristics: they seem to share nothing in common. *Ever since* is used to situate some state of affairs with respect to a previous event:

$$\text{I have mistrusted you} \left\{ \begin{array}{l} \textit{ever since} \\ \surd\, \textit{since} \\ \#\ \textit{because} \end{array} \right\} \text{we first met.} \tag{5.8}$$

Because, on the other hand, is used to explain the cause or justification for some eventuality:

$$\text{I accepted the gift,} \left\{ \begin{array}{l} \textit{because} \\ \surd\, \textit{since} \\ \#\ \textit{ever since} \end{array} \right\} \begin{array}{l} \text{I knew that my host would} \\ \text{lose face if I refused it.} \end{array} \tag{5.9}$$

Phrases like *since* are best thought of simply as having multiple possible meanings, just as 'traditionally' homonymous words like *bank* (= edge of river or financial institution) or *pen* (= writing instrument or animal cage). The alternative meanings of such words do not have anything in common; therefore they are of no great interest in motivating a set of relations.

Since there is nothing interesting to be learned from disjunctive cue phrases, most of them have been assigned subscripts according to their different meanings, and each meaning appears separately in the taxonomy. Thus the taxonomy treats the phrases *as(1)*, *as(2)*, and *as(3)* as completely different phrases.

Of course, deciding whether or not a cue phrase is disjunctive is a tricky question. In creating the original taxonomy, a measure of circularity is likely to be present—the decision calls for a theoretical assessment of the phrase, but the *justification* of this assessment is supposed to make reference to the finished taxonomy. The tactic adopted here has been to identify phrases as disjunctive only when they have two very clearly distinct and unrelated meanings, as is the case with phrases like *bank* and *pen*. *Since* seems to fall into this category; the other main examples are *as* (= because or while) and *while* (= as or whereas).

5.2.2 Necessary and Sufficient Conditions for the Use of a Cue Phrase

The extract in Figure 5.5 shows two exclusive phrases, *while* and *afterwards*.

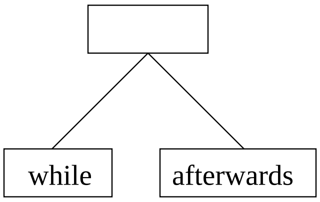

Figure 5.5: Exclusive Phrases

There is no question of swap-substitutability here; so we need to find a feature with respect to which these phrases differ. One likely candidate would make reference to the *temporal* relationship between the linked clauses: for *while*, the two clauses have to describe temporally overlapping events, whereas for *afterwards*, the two clauses have to describe events which happen in succession.

These do indeed seem to be necessary requirements for the two phrases. However, they are not sufficient. For instance, the events described in Example 5.10 are temporally simultaneous, but *while* is not appropriate:

$$\begin{array}{r} \text{I don't know where Jill is.} \\ \text{She was out} \end{array} \left\{ \begin{array}{l} \textit{when} \\ \#\ \textit{while} \end{array} \right\} \text{I called by.} \tag{5.10}$$

If we are using cue phrases to motivate relation definitions, then each cue phrase must be described using features which indicate both the necessary *and* the sufficient conditions for its use.

5.2.3 Constraints on the Range of Feature Values: Exclusivity and Exhaustivity

Each feature can take a number of alternative values. In this section, two important constraints on the range of values are outlined.

Firstly, the different values must be genuine alternatives: they must express inconsistent predicates. In other words, the different possible values of a feature must be mutually **exclusive**. If this were not the case, it would be possible to find an object which could be described by two or more values of the same feature: such an object could not be properly represented by the feature formalism, in which each feature can only take a single value.

Secondly, the different possible values of a feature must be **exhaustive**: between them, they should be able to represent every object in the set to be described. For instance, a feature COLOUR must have sufficient alternative values to represent the colour of all the objects in the set to be described. Of course, there are many types of object for which a given feature is *undefined*: thus to represent the category 'car', the feature COLOUR will not be assigned a value. However, a value for COLOUR could be assigned for any token car: it only violates the requirement of exhaustivity if there is some token car which cannot be described by one of its values.

5.2.4 The Requirement of Productivity

There is one final, extremely important principle governing the choice of features in the taxonomy. This time it is global, relating to the whole set of features eventually motivated. It is stipulated that the values of each feature must be defined so as to make sense in conjunction with every other combination of feature values. In other words, there must be no contradictions possible between the value of one feature and the value of any other feature: the only contradictions that are permitted are between alternative values of the same feature. This will be referred to as the requirement of **productivity**.

The requirement of productivity is an important departure from the system of feature structures found in ALE (Carpenter and Penn (1994)). Objects in ALE are classified within a hierarchy of **types**, and there is a function specifying for each type the features which are **appropriate** for that type. (For instance, the feature 'gender', taking alternative values 'male' and 'female', is appropriate for objects of type 'person', but not for objects of type 'furniture'.) In the present model, there is no such appropriateness function. Or rather, it is specified that each feature must be appropriate for *every* object in the set to be described.[4]

The need for typed feature hierarchies in lexical semantics is not in doubt (see e.g. Boguraev and Pustejovsky (1990), Evans and Gazdar (1989), Briscoe *et al* (1990)). Clearly, to represent the huge range of concepts that words describe, such formalisms are necessary: it makes sense, for instance, to have different sets of features for describing people and for describing furniture. So why should the requirement of productivity be imposed in representing cue phrases? The reason stems from the fact that cue phrases are closed-class words. The set of cue phrases is very small in comparison to the set of nouns or the set of verbs; moreover, while new nouns and verbs are continually being invented, the set of cue phrases is relatively static, evolving at a much

[4] The idea of 'an inheritance hierarchy for features' is thus used quite differently in ALE than it is in the current work. In ALE's type hierarchy, it is the *appropriateness* of a feature which is inherited from a type to its subtypes. In the taxonomy of cue phrases, on the other hand, it is the *values* of the features associated with a node that are inherited by its subordinate nodes. There is no notion of different sets of features being appropriate for different nodes. (Indeed, all features are appropriate for every node.) The resemblance between the taxonomy and an ALE-style type hierarchy is thus quite superficial.

slower pace. The same can be said for other closed-class words, such as quantifiers or prepositions. It can thus be argued that the studies of open-class and closed-class words are quite separate branches of lexical semantics; and that there is no requirement to adopt the same formalism in the two cases.

The requirement of productivity amounts to an assumption that the set of cue phrases is more homogeneous, and simpler to describe, than sets of open-class words. A typed feature hierarchy could clearly be used to represent cue phrases, but if it is possible to represent them without the additional power such a formalism provides, then it is preferable to do so, because to do otherwise would be to miss some generalisations.

To illustrate this argument, consider an existing non-productive classification of cue phrases—that developed by Martin (1992). As already noted in Section 2.4.2, Martin's relations are represented using a systemic network: an illustrative extract is given in Figure 5.6. Of course, it is possible to represent a productive set of features in

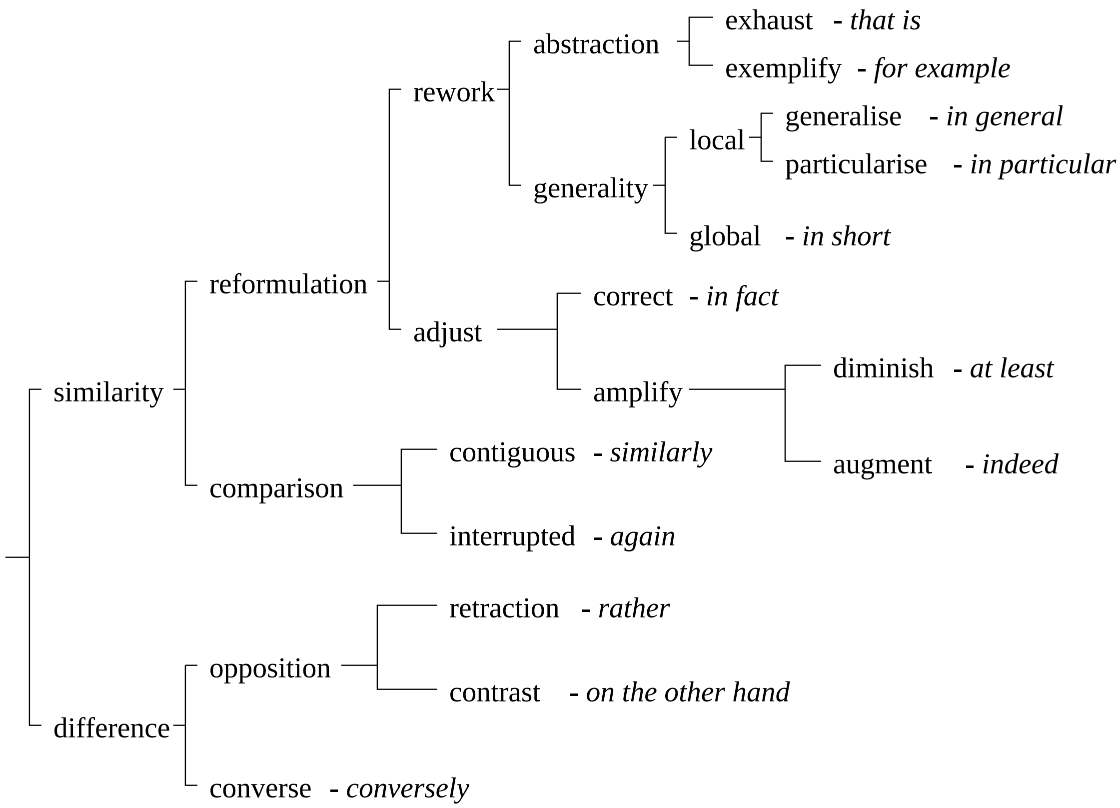

Figure 5.6: An Extract from Martin's Systemic Networks (Martin (1992) p.217)

a system network. But it is also possible to represent a non-productive one; and Martin takes advantage of this extra facility. Consider the choice between DIMINISH and AUGMENT in Figure 5.6. Martin describes this as being 'to do with whether the reformulation is toned up or down'. The distinction is expressed in terms of the value of a previously-decided feature, REFORMULATION. It cannot be expressed for relations which are not REFORMULATIONS (for instance, COMPARISONS): in other words, it is not a productive distinction. The non-productivity is achieved by imposing an order on the decisions about feature values, and only asking about the value of a given feature in circumstances where the values of other relevant features are appropriate.

There are problems with this system for classifying cue phrases. In particular, it does not allow each pair of phrases to be compared on the same terms. For example, the distinction between *at least* and *indeed* is expressed in terms of a feature with alternative values DIMINISH and AUGMENT; while the distinction between *that is* and *for example* is expressed in terms of a feature with alternative values EXHAUST and EXEMPLIFY. How are we to know that there is not something in common between these two distinctions, which is not being factored out? It is impossible to compare them directly, as they are specialised for different values of several other features. Using a productive set of features, each pair of phrases can be compared on the same terms: each feature makes sense for all possible combinations of all other feature values, so each cue phrase can be given a value for every feature in the set.

The main problem with non-productive formalisms such as this one is that deciding on features is almost too easy. Since a new specialised feature can be used to describe each new distinction, there is no necessity to ensure that the new feature does not overlap with other features in the types of variation it captures. There is thus a possibility for redundancy in the final set of features. Consider for instance, Martin's distinction between EXHAUST and EXEMPLIFY, used to represent the difference between *that is* and *for example*. The choice between these two relations is presented as a refinement of several other choices, namely REFORMULATION, REWORK and ABSTRACTION. And yet from the informal definitions Martin provides, it is hard to see exactly how these earlier choices end up contributing to the final concepts of EXHAUST and EXEMPLIFY. The final question 'shall I use an example?' could simply have been asked at the outset; it recapitulates all of the questions previously asked. So it is hard to see it as identifying a dimension of variation separate from those identified by previous portions of the network.

It is considerations such as these that prompt a search for a more independent set of parameters. Of course, it may turn out that no completely productive set of parameters can be found for describing cue phrases, in which case something like a systemic network would have to be adopted. However, the requirement of productivity is useful at least in forcing the theorist to begin by looking for as general an account as possible.

A Note about the Uneven Distribution of Cue Phrases

Although productivity is demanded of the set of features used to describe cue phrases, this is *not* to say that there must be a cue phrase for each different combination of features. In fact, as we shall see, this is far from being the case: there are many combinations of features for which no cue phrase exists, and many leaf-level cue phrases which are undefined for particular features.

This is not a violation of the productivity requirement. Productivity is a requirement of the *representation system* for cue phrases, not of cue phrases themselves. It is perfectly possible to represent the uneven distribution of cue phrases at the leaves of the taxonomy by showing, for instance, that some leaf-level phrases are undefined for more features than others. In fact, it is very useful to be able to represent those combinations of feature values for which no cue phrase exists, in order to frame questions about *why* no cue phrase exists to signal these features, or whether cue phrases exist in other languages for signalling them. If the representation system only captures the space of actual phrases, these lines of questioning cannot be pursued.

However, because some cue phrases are undefined for some features, it may actually make sense to use a systemic network when it comes to *deciding* on a phrase to use in a given situation. To make this decision, we have to ask a number of questions, to fix the values of the different relevant features. If the value of one feature determines whether or not some other feature is relevant in deciding between alternative phrases, it makes sense to find out the value of the former parameter first, just on grounds of efficiency. But again, it should be emphasised that the additional power of the network is only being used for *choosing* phrases, not for representing them.

5.3 The Structure of Feature Definitions

The final preliminary to creating feature definitions is to note some general requirements which they should meet, and to give an indication about the primitive concepts in terms of which they will be expressed.

5.3.1 Relations as Planning Operators

An important requirement for the feature definitions is that they form the basis for an implementable set of relations. In particular, we would like the relations we motivate to play a part in the process of text planning. Relations have already been usefully adapted to this task, as described in Section 2.5; and one of the primary aims of the thesis is to develop a new set of relations which builds on those which have thus far been implemented.

The notion of relations as **planning operators** is thus one which we will want to adopt from the outset. We need to describe relations in terms of the effects they achieve on the reader, and in terms of the circumstances in which they can be used; in other words, in terms of their **preconditions** and **postconditions**. We will thus be looking for definitions broadly along the lines of those developed by Hovy *et al* (1988, 1993) and Moore and Paris (1989, 1993).

A conception of relations as planning operators is not only useful from the standpoint of implementation, however. Since the advent of speech act theory (Austin (1962), Searle (1969)), an active research programme has grown up around plan-based approaches to pragmatics—see for instance Cohen and Perrault (1979), Cohen and Levesque (1990), Allen (1995). And if utterances are best thought of as actions intended to achieve particular effects, then it makes sense to think of pairs of adjacent utterances in the same way.

5.3.2 The Primitives to be Used in Feature Definitions

We now turn to the question of the primitive concepts which should be used to represent the preconditions and postconditions of relations. At the highest level, relations (and the features into which they decompose) will be expressed in terms of the writer's desires and beliefs. We can think of the writer as having a **believed world model**, containing her representation of what is true in the world, both with regard to particular circumstances and to generalisations of different kinds. Likewise, the writer will have a **desired world model**, representing the particular and general conditions which she would like to be the case. Actions of different kinds, including linguistic ones, are triggered when the appropriate correspondences between believed and desired world models arise.

For linguistic actions, the writer's model of the reader plays a particularly important role. The reader's beliefs and desires will both be important, and representations of each will figure in the writer's believed and desired world models. The world models at the highest level of structure are thus nested as in the diagram in Figure 5.7.

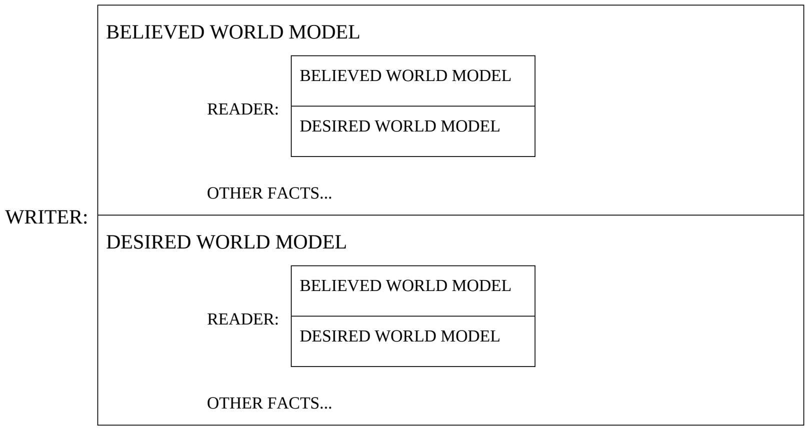

Figure 5.7: Primitives for Feature Definitions: The Top Level of Structure

The notion of an infinite regression of nested beliefs ('Writer believes [Reader believes [Writer believes...]]') emerges easily from consideration of this diagram. A problem to beset many plan-based approaches to natural language processing (see e.g. Cohen and Perrault (1979)) is that such a regression seems an essential component of the postconditions of linguistic actions. This problem will not be addressed directly here; for the purposes of the definitions developed in the next chapter, two levels of nesting should prove sufficient.

5.4 Summary

This chapter contains a general discussion of how the linguistic data in the taxonomy of cue phrases should be exploited from a theoretical point of view. The most important conclusion is that the taxonomy lends itself to description in terms of a set of independent features—in other words, that cue phrases (and the relations they signal) should be thought of as composite constructs. The task of defining the set of relations now reduces to the task of defining the different features from which relations are composed.

A number of requirements for the features to be motivated from the taxonomy have been set out in advance. Firstly, they must describe both the necessary and the sufficient conditions for the presence of a cue phrase. Secondly, they must be productive, by which is meant that the values of each feature should be defined for all combinations of all other feature values. Finally, they should conform to a conception of relations as planning operators, for reasons both implementational and philosophical.

In the next chapter, the different features proposed to analyse the taxonomy are presented one by one. To what degree they meet the requirements set out here will be an important matter for discussion.

Chapter 6

Using the Taxonomy to Create Relation Definitions

6.1 Introduction

In this chapter, the linguistic data assembled in the taxonomy is finally given a theoretical interpretation. Until now, the investigation has focused on surface structures in text; actual words and phrases, and how they can be manipulated by writers. However, as Scott and Paris (1995) note, at some point it is necessary to go 'beyond the text', to produce some description of it which is independent of its surface linguistic structures, in terms of which these structures can be defined. This is the concern of the present chapter.

A great deal has already been written about the semantics and pragmatics of cue phrases, and it should be no surprise that many of the ideas in this chapter draw on or expand on existing work. There are, however, two novel elements in the study, which provide some interesting new perspectives. Firstly is the fact that the set of cue phrases under investigation is much larger than usual. Often, theorists concentrate on a small set of cue phrases, or even on a single phrase—for instance *but* (Spooren (1989), von Klopp (1993)) or *when* (Moens and Steedman (1988)). The present study, as a result of the requirement of productivity, is much broader in scope: we will be looking for parameters which are valid right across the space of cue phrases. The reasoning behind this approach is that much can be learned about the semantics of a given cue phrase by comparing it to a number of other quite different phrases. To investigate the semantics of *because*, for instance, it would be instructive to be able to point up the similarities and differences between it and phrases as diverse as *when*, *then*, *even if*, *or*, *to*, *on the other hand* and *but*.

A second difference in the present work is that each theoretical construct introduced will be motivated in exactly the same way; by examining a portion of the taxonomy of cue phrases, and noting a demand for features to represent the patterns of substitutability it contains.

It should be stressed that the aim of this chapter is not to come up with a complete and watertight set of relation definitions. Producing a complete feature-theoretic account of the taxonomy is a huge task, and well beyond the scope of this thesis. The present aim is rather twofold:

- to demonstrate the utility of the substitution methodology, by noting some of the interesting theoretical constructs which emerge when it is pursued;
- to motivate a core set of features needed to describe the taxonomy, to serve as the basis for further investigations.

In Section 6.2, the set of feature definitions is presented. Each feature is motivated separately, using appropriate extracts from the taxonomy. Section 6.3 provides a brief account of how the individual features interact with each other. Section 6.4 takes another look at the mapping between the cue phrases for which definitions have been formulated and the coherence relations which they signal. A summary and discussion is given in Section 6.5.

6.2 Features Motivated by the Taxonomy

6.2.1 Semantic and Pragmatic Relations

The first extract from the taxonomy to be considered is given in Figure 6.1.[1] Motivating examples are given in Texts 6.1 and 6.2:

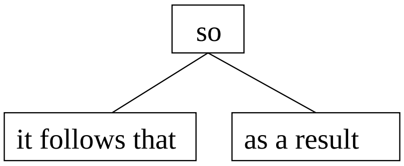

Figure 6.1: Semantic and Pragmatic Phrases

$$\text{The footprints are deep and well-defined.} \left\{ \begin{array}{l} \textit{It follows that} \\ \checkmark\, \textit{So} \\ \#\ \textit{As a result,} \end{array} \right\} \text{the thief was a heavy man.} \qquad (6.1)$$

$$\text{I had a puncture on the M25 on my way back from work.} \left\{ \begin{array}{l} \textit{As a result,} \\ \checkmark\, \textit{So} \\ \#\ \textit{It follows that} \end{array} \right\} \text{I missed most of the first half.} \qquad (6.2)$$

As a result seems strange in the context of Example 6.1, because it suggests that the thief's heaviness is *caused* by the footprints being deep. Conversely, *it follows that* is odd in Example 6.2, because it suggests that the writer is *deducing* the fact that she missed most of the first half—while in fact she is reporting from her own experience. Note that *so* is acceptable in both cases.

[1] The extracts from the taxonomy given in this section will not be labelled with the features they serve to motivate. A labelled version of the taxonomy is given in Section 6.3; the reader might find it useful to refer forward to this section to see how feature values are assigned to cue phrases.

Both examples are commonly analysed as involving a causal/inferential relation of some kind, the difference between them being to do with what this relation holds between. Traditionally, the relation in examples like 6.2 is taken to be between the events in the world described by the two clauses: the puncture causes the missed first half. In examples like 6.1, the relation is taken to involve linguistic events themselves, not just the events they represent. For Martin (1992), Sanders *et al* (1992) and others, the writer's *statement* that the thief was heavy is caused by her belief that the footprints are deep. Commentators have used a variety of terms to represent these distinctions. Martin (1992) and Halliday and Hasan (1976) use the labels EXTERNAL and INTERNAL to refer to examples like 6.2 and 6.1 respectively; Redeker (1990) uses the labels IDEATIONAL and PRAGMATIC; Van Dijk (1979) and Sanders *et al* (1992) talk of SEMANTIC and PRAGMATIC relations.

A useful modification of the notion of PRAGMATIC relations is introduced by Sweetser (1990). For her, the relation in examples like 6.1 primarily describes the cause of the writer's *conclusion* that the thief must have been heavy, and only indirectly describes the cause of her *statement* to this effect. The important relation in the example is the logical one, between two of the writer's beliefs. Sweetser calls relations involving the writer's beliefs EPISTEMIC, and defines a further category of SPEECH ACT relations which make reference to actual writer utterances. Prototypical of SPEECH ACT relations are examples like the following:

(6.3) What are you doing tonight, because there's a good movie on.

Here it is certainly appropriate to analyse the relation as describing the cause of the writer's utterance *What are you doing tonight?*.

Sweetser's definition of EPISTEMIC relations is an improvement on the previous definitions. However, it still leaves something to be desired. Consider again Sweetser's EPISTEMIC analysis of Example 6.1, as a statement about the writer's conclusions and how they were reached. The text, according to the new analysis, is still fundamentally descriptive; instead of describing the external world, it now contains a description of the writer's own thought processes. What is missing is an account of how an argumentative text like this one achieves a rhetorical *effect* on the reader—how it *persuades* the reader that the thief was heavy, where a simple statement like *The thief was heavy* might not have sufficed. Of course, in offering the reader a trace of the writer's reasoning, the text suggests how the reader might come to the same conclusion. But while it is vital for the writer's purposes that the reader take this last step, it is not represented in Sweetser's analysis.

Thinking about utterances in terms of their intended effects on the reader suggests an alternative definition for EPISTEMIC relations. The new definition expresses the intended effect of a text containing two related utterances *as a relation between the intended effect of each individual utterance*. The suggestion is, for instance, that in a text containing an EPISTEMIC *so*, the writer's intended effect is not that the reader *believe* a statement about the causes of the writer's beliefs, but rather that a causal relation *actually does* hold, in the real world, between the intended effects of the two related utterances—in other words, between two reader beliefs.

Consider how this definition works in the case of Example 6.1. The text is presented again below, and the intended effects of its two clauses are shown in italics:

(6.4)	The footprints are deep.	*So*	the thief was a heavy man.
	R believes the footprints are deep.	*So*	*R believes the thief was a heavy man.*

On this interpretation, the intended effect of the text is that the reader's belief that the footprints are deep *causes* the reader to believe that the thief was a heavy man, where otherwise the reader would not have been inclined to believe this latter statement. It is because of *this* cause that the relation has persuasive force.

Note that the proposed new definition of EPISTEMIC relations actually extends to some relations Sweetser considers as SPEECH ACT. For Sweetser, "if an utterance is imperative (...) in form, then it cannot reasonably be causally conjoined to another utterance except at the speech act level" (p78). Thus Sweetser would interpret an example like 6.5 as SPEECH ACT:

(6.5) Hurry up, because we haven't much time!

But here again, there are advantages in expressing the relation in terms of intended effects. The writer's main intention in such a case is not to inform the reader about the cause of her utterance; but rather that the realisation that they haven't much time should motivate the reader to hurry up. Using the new definition, this is just what is expressed. The intended effect of the imperative *Hurry up* is that the reader hurry up; the intended effect of the statement *we haven't much time* is that the reader believe they haven't much time; and the intended effect of the whole utterance is that this belief causes the reader to hurry up.

Because the new definition encompasses examples such as this one, we have decided to revert to the label PRAGMATIC to refer to the relations it describes; and consequently to return to the label SEMANTIC for what Sweetser calls CONTENT relations. The definition for a feature with alternative values SEMANTIC and PRAGMATIC can now be given. After Sanders *et al*, we can call this feature SOURCE OF COHERENCE.[2]

SOURCE OF COHERENCE

> SEMANTIC: the intended effect of the text containing the relation is that the reader believes some relation holds between two propositions A and C. A and C are the propositional contents of the two related text spans S_A and S_C.
>
> PRAGMATIC: the intended effect of the text containing the relation is that some relation actually holds between two propositions A and C. A and C are the intended effects of the two related text spans S_A and S_C.

[2] It should be noted at this point that, as new features are motivated, changes are sometimes required to the definitions of features already introduced. In consequence, the feature definitions presented one by one in this chapter should not be regarded as final, but only as sufficient to account for the data so far introduced.

Note that the new notion of PRAGMATIC by no means covers all of Sweetser's SPEECH ACT relations. (For example, Text 6.3 is still much better analysed as describing the causes of the writer's speech act.) However, the class of SPEECH ACT relations has not yet been motivated from the taxonomy, as it is hard to find cue phrases which are specific to this class.

6.2.2 POSITIVE and NEGATIVE POLARITY Relations

A second portion of the taxonomy is given in Figure 6.2. Some motivating examples are provided in Texts 6.6 and 6.7.

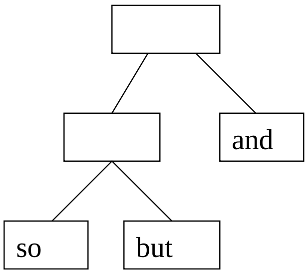

Figure 6.2: POSITIVE and NEGATIVE POLARITY Phrases

$$\begin{array}{r}\text{Jim had just washed his}\\ \text{car,}\end{array} \left\{\begin{array}{l} \textit{so} \\ \checkmark\, \textit{and} \\ \#\ \textit{but} \end{array}\right\} \begin{array}{l}\text{he wasn't keen on lending it}\\ \text{to us.}\end{array} \tag{6.6}$$

$$\begin{array}{r}\text{It was odd. Bob shouted}\\ \text{very loudly,}\end{array} \left\{\begin{array}{l} \textit{but} \\ \checkmark\, \textit{and} \\ \#\ \textit{so} \end{array}\right\} \text{nobody heard him.} \tag{6.7}$$

The fact that the phrases *but* and *so* can never be substituted for one another is clear from consideration of examples like these. But the examples also show that the phrase *and* is contingently substitutable both for *but* and (in other contexts, of course) for *so*. In feature-theoretic terms, we can conclude that *but* and *so* are defined for different values of some feature; and that *and*, being contingently substitutable for both, is undefined for this feature. It remains now to decide what the feature is.[3]

Many different suggestions have been made as to the similarities and differences between phrases like *but* and *so*. It is uncontroversial (as far as it goes) that *A, so C* signals some kind of implication or cause, with A as the antecedent/cause and C as the consequent/result. And it is likewise uncontroversial to say that *A, but C* signals (or at least can signal) a violation of the type of relation signalled by *so*. To illustrate with reference to the above examples: in Text 6.6, *so* signals that it follows from the fact that Jim had just washed his car that he was unwilling to lend it to us. In Text 6.7, *but* signals that it *normally* follows from the fact that Bob shouts loudly that people

[3] Note that in order to motivate a feature which distinguishes between *so* and *but*, we are not obliged to find a common hypernym of the two phrases, as we did in the previous section. The fact that *and* can sometimes be substituted for both *but* and *so* is sufficient to show that it *cannot* be defined for any feature which takes alternative values for these two phrases. If it were, there would have to be an exclusive relationship between it and one or other of the phrases.

hear him, but in this case no-one does. Both phrases can thus be thought of as having a consequential component: for *so*, the consequence relation is specified as succeeding; while for *but*, an expected consequence is not forthcoming. With *and*, it is simply not specified whether or not the consequence relation succeeds—the information is left to be inferred by the reader.

The important question is how to express the above ideas more precisely. It has been common to begin formalising the difference between relations signalled by *so* and those signalled by *but* by making reference to a 'statement of implication' $P \rightarrow Q$ which underlies both types of relation. The difference between the two relations is then expressed in terms of the relationship between P and Q and the propositions in the related spans of text. For *so*, P relates to the proposition in the first span and Q to that in the second span. For *but*, P relates to the proposition in the first span and Q to the *negation* of that in the second span. This story is roughly that given by Longacre (1983) in distinguishing between 'consequence' and 'frustrated consequence' relations. Sanders *et al* (1992) give a similar story to distinguish between POSITIVE and NEGATIVE POLARITY relations; I shall use these latter terms in what follows.

Central to the distinction between *so* and *but* is the notion that causal or consequential rules can be **defeated**. A number of recent accounts of concessive relations have employed the notion of defeasible rules; in particular Oversteegen (1995) and Grote, Lenke and Stede (1995). Defeasible rules provide a useful method for representing the kind of common-sense generalisations which people rely on in order to make up for their partial knowledge of the world. They are becoming increasingly popular in computational linguistics, as a tool for modelling the influence of the reader's world knowledge on the resolution of ambiguities. For instance, Lascarides and Asher (1991), Lascarides, Asher and Oberlander (1992) use a system of defeasible rules to develop a framework for deciding which coherence relation is present at a particular point in a text when this is not signalled explicitly. Hobbs *et al* (1993) use defeasible rules to model a range of processes in text interpretation, including the resolution of anaphora, lexical ambiguities and compound nominals. The use of defeasible rules in the present context is somewhat different, however. They are not being proposed as a way of *deciding* about the interpretation of some part of a text, but as a part of the interpretation *itself*—the defeasible rules used by a reader and writer to model the world are actually implicit in the semantics of phrases like *but*. As an initial model, then, we might propose that the phrases *so* and *but* are each associated with a defeasible rule, which in the case of *so* succeeds and in the case of *but* is defeated.

A number of questions still remain, however. Most importantly, what is the communicative status of the defeasible rule? Is it something which the reader must already know as a precondition to understanding the text, or is it something which the reader is told in the text? Along with Oversteegen (1995), it is here proposed that the existence of the defeasible rule should be seen as a precondition. It is problematic to suggest that the rule itself is part of the information communicated to the reader by the writer. For one thing, the reader is only given one instance of the rule—it would then be necessary to abstract away from this to the rule itself; a process which is very underconstrained. Moreover, it is questionable whether causal or inferential *rules* constitute the kind of information that a reader will accept 'on authority' from a writer in any circumstance. Consider again the statement in Text 6.6: *Jim had just washed his* *car, so he wasn't keen on lending it to us.* This may certainly provide new information, but it is implausible to suggest that the writer is *informing* the reader that 'if a person with temperament T has just washed his car, he normally doesn't like to lend it to others'. It is more plausible to suggest that a rule along these lines is already known by the reader, and what is being communicated is the fact that the rule succeeds in this instance.

Note that the information conveyed by such a statement might be more than the bare assertion that 'there is nothing unusual about the situation being described'. Knowing that the rule in question is *triggered* would allow the reader to infer that Jim is of type T, for instance, if this was not already known. But it is much easier to imagine the reader adding such facts to his database than whole causal rules.

To sum up: we can hypothesise a feature called POLARITY, with alternative values NEGATIVE and POSITIVE. It is assumed that each relation presupposes the presence of a defeasible rule $P \rightarrow Q$.[4] The relationship between P and Q and the propositions A and C (defined in the SOURCE OF COHERENCE feature) is determined by the different values of the POLARITY feature, as follows:

POLARITY
POSITIVE: $A = P$; $C = Q$. The rule is specified to succeed.
NEGATIVE: $A = P$; C is inconsistent with Q. The rule is specified to fail.

Conditional NEGATIVE **and** POSITIVE POLARITY

Another portion of the taxonomy which can be used to motivate the POLARITY parameter is given in Figure 6.3. Motivating examples are given in Texts 6.8 and 6.9:

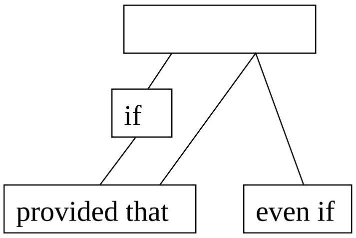

Figure 6.3: Conditional POSITIVE and NEGATIVE POLARITY Phrases

$$\text{You can sit in the front seat,} \left\{ \begin{array}{l} \textit{provided that} \\ \checkmark\ \textit{if} \\ \#\ \textit{even if} \end{array} \right\} \text{you put your seatbelt on.} \tag{6.8}$$

$$\text{I wouldn't vote for Major} \left\{ \begin{array}{l} \textit{even if} \\ \checkmark\ \textit{if} \\ \#\ \textit{provided that} \end{array} \right\} \text{you gave me a thousand pounds.} \tag{6.9}$$

[4] We will not at this point buy into any particular formalism for representing defeasible rules. However, some of the requirements for the formalism eventually to be adopted will emerge from the discussion in following sections.

The phrases in Figure 6.3 can be compared to those in Figure 6.2, a crucial difference being the fact that the latter group of phrases relate hypothetical eventualities while the former phrases relate actual ones. Motivation for a feature representing this dimension of variation will be provided below, in Section 6.2.8; for now the important thing to note is the variation in polarity exhibited by the phrases. For each phrase, an underlying defeasible rule $P \rightarrow Q$ can be identified. For *C, provided that A*, A and C map onto P and Q respectively and the rule is represented as succeeding. For *C, even if A*, A and C map onto P and $\neg Q$ respectively, and the rule is represented as failing. Thus in Example 6.8, the rule that putting a seatbelt on causes being allowed to sit in the front seat is asserted to succeed, while in Example 6.9 the rule that giving people lots of money causes them to vote against their will is asserted to fail in the case of the writer.

Semantic and Pragmatic Negative Polarity Relations

The distinction between POSITIVE and NEGATIVE POLARITY also cuts across the SEMANTIC/PRAGMATIC distinction. Consider Figure 6.4, for which motivating examples are given below:

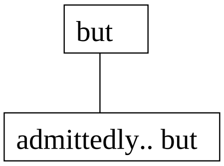

Figure 6.4: Semantic and Pragmatic Negative Polarity Phrases

$$\begin{array}{r}\text{United have some key}\\ \text{players injured;}\end{array}\left\{\begin{array}{l}\textit{admittedly}\ldots\textit{but}\\ \checkmark\ \textit{but}\end{array}\right\}\text{ they're still bound to win.}\tag{6.10}$$

$$\begin{array}{r}\text{Mary was behaving oddly.}\\ \text{She ordered a pizza,}\end{array}\left\{\begin{array}{l}\textit{but}\\ \#\ \textit{admittedly}\ldots\textit{but}\end{array}\right\}\text{ she didn't eat any of it.}\tag{6.11}$$

(The construction *admittedly... but* in these examples is to be read as distributed between the two clauses in the relation. The first example should thus read '*Admittedly,* United have some key players; *but...*', and the second example should read '*Admittedly,* she ordered a pizza, *but...*'.)

The point is that *Admittedly... but* signals the breaking of a defeasible rule just as *but* does; yet it has a specifically argumentative flavour. In Text 6.10, *admittedly* introduces a proposition which suggests one conclusion, and the negation of that conclusion is then asserted. *But* on its own can also be used in the absence of any argument, as in Text 6.11: here, the writer is simply informing the reader about an unusual state of affairs, and *admittedly* is quite out of place.

The SEMANTIC/PRAGMATIC distinction is useful in capturing the difference between these NEGATIVE POLARITY phrases. *Admittedly... but* can be defined as signalling the value PRAGMATIC, and *but* can be thought of as undefined for the feature. The difference between SEMANTIC and PRAGMATIC NEGATIVE POLARITY relations can be thought of as follows. In the SEMANTIC case, the writer's aim is to inform the reader that some general rule in the reader's model of the world is defeated in the situation being described. Thus in the above example, the defeated rule is that people who order food generally eat it. In the PRAGMATIC case, the writer's aim is that some general rule in the world itself *actually* fails in the present instance. This rule holds between two reader beliefs—the intended effects of the first and second clauses taken individually. In the above example, the intended effect of the first span is that the reader believe that United has some key players injured; that of the second span is that the reader believe that United will win. Normally, if the reader believes a team has several players injured, he will believe they will lose; but in this case, the writer's intention is that this conclusion is not drawn.

6.2.3 UNILATERAL and BILATERAL Relations

The next portion of taxonomy to be considered, given in Figure 6.5, also involves NEGATIVE POLARITY phrases. Motivating examples are given in Texts 6.12 and 6.13:

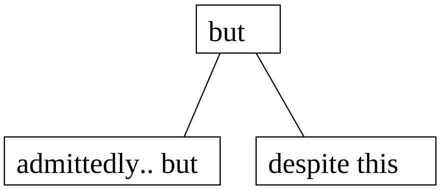

Figure 6.5: UNILATERAL and BILATERAL Phrases

$$\text{Bill lost the 400m last year.} \left\{ \begin{array}{l} \textit{Admittedly... but} \\ \checkmark\ \textit{But} \\ \checkmark\ \textit{Despite this,} \end{array} \right\} \text{He should win it this year.} \tag{6.12}$$

$$\begin{array}{r} \text{Bill should win the 400m.} \\ \text{He lost last year;} \end{array} \left\{ \begin{array}{l} \textit{Admittedly... but} \\ \checkmark\ \textit{but} \\ \#\ \textit{despite this,} \end{array} \right\} \begin{array}{l} \text{they're running at altitude} \\ \text{this time.} \end{array} \tag{6.13}$$

The relation in each of these examples can be signalled by *admittedly... but*, and can thus be considered as PRAGMATIC NEGATIVE POLARITY. However, the phrase *despite this* is only acceptable as a substitute in Text 6.12. In Text 6.13 it appears odd; it suggests that Bill's losing last year would normally lead to them not running at altitude this time—an implausible assumption.

The difference between the two examples appears to be to do with the status of the second span in the relation (the one introduced by *but*). In each case, the first span presents a premise P_1 which suggests a conclusion C. In Text 6.12, the second span presents the negation of this conclusion, $\neg C$. In Text 6.13, the second span presents another premise P_2, which is more telling than P_1, and suggests an alternative conclusion. This dimension of variation suggests another parameter, which we can call PATTERN OF INSTANTIATION.

To allow for multiple premises in the rule underlying the relation, we need to review the definition of POLARITY presented in the previous section. The assumption must now be that each relation presupposes a rule of the form $P_1 \wedge \ldots \wedge P_n \rightarrow Q$. We need to map elements from this rule onto the variables A and C, defined in the feature SOURCE OF COHERENCE. Let us assume that A is always on the left-hand side of the rule.[5] The POLARITY feature presented before assumed that C always mapped onto the conclusion of the rule (Q); but we must now abstract away from this idea. We must define a new variable—call it C'—whose relationship to C is determined by the value of the POLARITY feature. For symmetry, we will also introduce a variable A', which always equates directly with A.[6] The POLARITY feature can now be thought of as specifying a function from A and C to A' and C'. Its revised definition now looks like this:

> POLARITY (2nd definition)
>
> POSITIVE: $A' = A$; $C' = C$. The rule is specified to succeed.
>
> NEGATIVE: $A' = A$; C' is inconsistent with C. The rule is specified to fail.

The mapping between A' and C' and the rule $P_1 \wedge \ldots \wedge P_n \rightarrow Q$ is now given by the new feature:

> PATTERN OF INSTANTIATION
>
> UNILATERAL: C' is on the same side of the rule as A' (but not the same as A).
>
> BILATERAL: C' is on the opposite side of the rule to A'.

The phrase *despite this* can now be thought of as defined as BILATERAL, while both *but* and *admittedly... but* are undefined for the feature. Thus in Example 6.12, for instance, A' is the proposition 'Bill lost the 400m last year' and C' is the proposition 'It is *not* the case that Bill should win this year'. These two propositions can be thought of as premise and conclusion of a defeasible rule. The relation can thus be thought of as BILATERAL, (and thus *despite this* is appropriate). Note that since the relation is also NEGATIVE POLARITY, the point is that the rule is defeated.

Several examples of phrases defined as UNILATERAL will be given in the following sections.

6.2.4 CAUSAL and INDUCTIVE Relations

Consider next the extract of the taxonomy in Figure 6.6, again featuring NEGATIVE POLARITY phrases. Motivating examples are given below:

---

[5] This assumption itself changes in Section 6.2.5, for reasons developed in that section.

[6] This is another assumption which will be re-examined as further portions of the taxonomy are considered; see Section 6.2.6 for details.

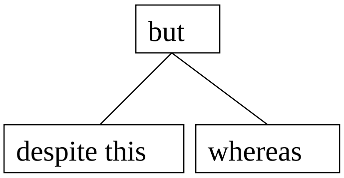

Figure 6.6: CAUSAL and INDUCTIVE Phrases

$$\text{Bill and Jill are like chalk and cheese. Bill lives for his books;} \left\{ \begin{array}{l} \textit{whereas} \\ \checkmark\ \textit{but} \\ \#\ \textit{despite this,} \end{array} \right\} \text{Jill is only interested in Tae Kwan Do.} \tag{6.14}$$

$$\text{Bob was out of training;} \left\{ \begin{array}{l} \textit{despite this} \\ \checkmark\ \textit{but} \\ \#\ \textit{whereas} \end{array} \right\} \text{he completed the marathon in record time.} \tag{6.15}$$

These phrases draw attention to another dimension of variation in the phrase *but*. Text 6.15 is of a kind we have already seen, where a causal or inferential rule is defeated: the rule in this case is that people who are out of training do not normally break records. But no such rule appears to underlie Text 6.14: knowing that Bill lives for his books gives no grounds for thinking that Jill is not just interested in Tae Kwan Do.

Whereas intuitively signals a contrast between two propositions. On a conventional analysis (see e.g. Spooren (1989)), we are able to say S_1, *whereas* S_2 if S_1 and S_2 allow the inference of $p(A)$ and $\neg p(B)$ respectively, for some predicate p and two entities A and B.[7] The question posed by the extract given in Figure 6.6 is: what does this have in common with the violated expectation analysis required for *despite this*? The feature-theoretic interpretation of the diagram requires us to find some feature or features of *whereas* that are shared by *despite this*; *but* is defined for this common component, and undefined for the feature(s) which distinguish the two phrases.

A point to note about $p(A)$ *whereas* $\neg p(B)$ is that A and B are required to belong in some sense to the same category of entities. Constrasts are not made between objects which have nothing in common at all. Thus Bill and Jill might be brother and sister, or friends, or two candidates for some job that needs doing. Put another way: relations signalled by *whereas* highlight an inability to *generalise* over the objects in a given class as regards some property p. This idea prompts the suggestion that a different type of rule underlies such relations; namely **inductive rules**.

An inductive rule is of the following general form: if property p is true of a certain finite number of elements from a particular class, then it follows that p is true of *all* the elements in that class. Clearly, as rules of inference, such statements are not sound. But inductive rules are nonetheless a mainstay of human reasoning: in the end, all our generalisations about the world are arrived at by inducing from particular instances. They can in fact be considered as another kind of *defeasible* rule, albeit quite different from those which we have so far been considering. Both kinds of rules

[7] In Text 6.14, for instance, we can infer from *Bill lives for his books* that *Bill is not only interested in Tae Kwan Do*, which is an explicit negation of the predicate in the second span of the relation.

are to be relied on when no information to the contrary is found, but are overruled if contrary information is forthcoming.

A proposition like $p(A)$ can thus be thought of as forming part of the left hand side of an inductive rule. If enough other objects in the same class as A have property p, then the inductive generalisation that $\forall X(X \in C \rightarrow p(X))$ will be triggered. Even the single instance of $p(A)$, in the absence of any other information, is presumably enough to trigger some very weak hypotheses about other similar objects. However, the inductive rule is defeated as soon as $\neg p(B)$ is presented.

In summary, what *whereas* and *despite this* seem to have in common is that they both presuppose a defeasible rule of some kind, and both signal its defeat. In one case the defeasible rule is causal and in the other case, inductive; *but* can then be seen as undefined with respect to the type of rule which is defeated. A new feature RULE TYPE is now motivated:

> RULE TYPE
>
> CAUSAL: the defeasible rule $P_1 \wedge \ldots \wedge P_n \rightarrow Q$ is a causal rule.
>
> INDUCTIVE: the defeasible rule $P_1 \wedge \ldots \wedge P_n \rightarrow Q$ is an inductive rule.

As might have been noticed, *whereas* and *despite this* differ not only as regards the feature RULE TYPE, but also as regards the feature PATTERN OF INSTANTIATION. *Despite this* is defined as BILATERAL, as we have already seen in Section 6.2.3. *Whereas* must be defined as UNILATERAL: its two spans present two propositions from which generalisations can be drawn, and the generalisation itself (or rather the lack of it) remains implicit. Consider Example 6.14 in more detail. A and C can be identified as 'Bill lives for his books' and 'Jill is only interested in Tae Kwan Do' respectively. The relation is NEGATIVE POLARITY, so while A' is identical to A, C' is inconsistent with C. *Whereas* is defined as UNILATERAL; so A' and C' are both on the left-hand side of some defeasible rule. The relation is INDUCTIVE, so each item on the left-hand side of the rule will be attributing the same predicate to a different member of a given class. In this case, then, C' must be defined as 'Jill lives for her books', or something to that effect. As this is inconsistent with the value of C, the inductive rule fails.

SEMANTIC and PRAGMATIC INDUCTIVE Phrases

The distinction between CAUSAL and INDUCTIVE relations cuts across that between SEMANTIC and PRAGMATIC relations. Consider the extract in Figure 6.7, motivated by the examples below:

Bill and Jill are like chalk and cheese. Bill lives for his books;	*whereas* √ *on the other hand,* √ *but* # *then again,* # *despite this,*	Jill is only interested in Tae Kwan Do.

(6.16)

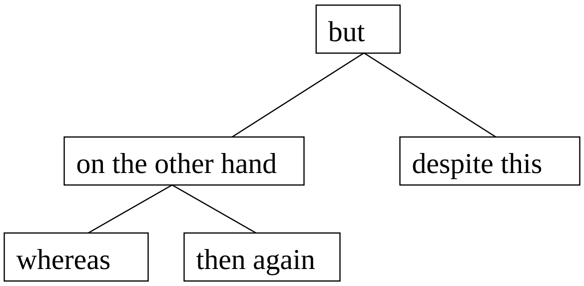

Figure 6.7: SEMANTIC and PRAGMATIC INDUCTIVE Phrases

$$\begin{array}{r}\text{I don't know where we}\\ \text{should eat tonight. The}\\ \text{Kalpna is great value for}\\ \text{money;}\end{array} \left\{\begin{array}{l} \textit{then again,}\\ \checkmark\ \textit{on the other hand,}\\ \checkmark\ \textit{but}\\ \#\ \textit{whereas}\\ \#\ \textit{despite this}\end{array}\right\} \begin{array}{l}\text{Sid isn't crazy about Indian}\\ \text{food.}\end{array} \tag{6.17}$$

Example 6.16 is the kind of contrast we have already considered, between objects in the world with inconsistent predicates. Example 6.17 can also be analysed as a contrast of sorts, but here the inconsistency is between the argumentative force of two propositions. The fact that the Kalpna is good value for money suggests that we should eat there. But the fact that Sid doesn't like Indian food suggests that we should not eat there.

The interesting relationship in this diagram is between *whereas*, *then again* and *on the other hand*. *Then again* seems quite wrong in the first example, and *whereas* is out of place in the second one. But note that *on the other hand* is acceptable in both cases. It will be suggested in the remainder of this section that all three phrases are signallers of NEGATIVE POLARITY INDUCTIVE relations; that the exclusivity between *whereas* and *then again* is due to the SEMANTIC/PRAGMATIC distinction; and that *on the other hand* is undefined for this latter feature and hence substitutable for both phrases.

The contrast signalled by *whereas* relates to the propositional content of the related spans, and the writer's objective in presenting it is to make the reader aware of the generalisation which fails. It can thus be thought of as a SEMANTIC contrast. But clearly, the contrast signalled by *then again* does not have to relate to the propositional content of the spans. No inconsistent predicates are present in the contents of the two spans in Example 6.17, for instance. In such cases, different kinds of objects are apparently being compared. A PRAGMATIC analysis of these cases will be suggested here: in this analysis, the objects are reader beliefs (rather than objects in the world), the predicates about the objects concern the different conclusions supported by different beliefs, and the classes into which beliefs fall concern the conclusions to which they are relevant.

Consider what happens in the process of 'reaching a conclusion by examining premises'. There must first be an attempt to delineate those premises which will be relevant; an exhaustive search will not be feasible in any system with a reasonably sized set of facts and rules. Even the set of relevant propositions is likely to be too large to be exhaustively searched, and further heuristics will need to be used to consider these selectively. It is reasonable to suppose that inductive principles play a part in these heuristics: if we consider a certain number of relevant premises, and each one supports the same conclusion, there will come a point at which we decide that the conclusion is true, and stop looking for additional premises. The notion of 'reaching a conclusion' can be thought of as the moment when an inductive generalisation is made, and we assume that *all* premises relevant to the conclusion in fact support the conclusion.

Now consider what happens with *then again*. Here two relevant premises are presented which support opposite conclusions. This has the effect of blocking the generalisation, making the reader unable to reach a decision. In Lascarides and Asher's (1991) terminology, the reader has encountered a 'Nixon Diamond', where two defeasible rules are triggered, and neither takes precedence. Note that as this is a PRAGMATIC relation, the effect is not simply that the reader *realises* that inconsistent premises have been presented, but that a Nixon Diamond *actually happens* in the reader's theorem proving system. However, we must also note that the system does not freeze up altogether as a result of this impasse: it is only as regards one particular conclusion that no decision can be reached. Inductive generalisations can still be made to reach other conclusions.

POSITIVE **and** NEGATIVE POLARITY INDUCTIVE **Phrases**

Finally, consider the diagram in Figure 6.8, motivated by Texts 6.18 and 6.19.

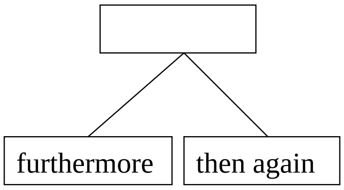

Figure 6.8: POSITIVE and NEGATIVE PRAGMATIC INDUCTIVE Phrases

United are bound to win. They have a great team; { *furthermore,* / # *then again,* } they're playing at home. (6.18)

I don't know whether United will win. They have a great team; { *then again,* / # *furthermore,* } they're playing away from home. (6.19)

If *then again* signals the defeat of a PRAGMATIC INDUCTIVE rule, then *furthermore* can be regarded as signalling the success of such a rule. In Example 6.18, a conclusion is reached: that United will win. Two relevant premises needed to be considered in order for the inductive rule to fire in this instance.

In other cases, more than two premises need to be advanced. Lists of premises signalled by phrases like *furthermore* can in principle be of any length. In the present model, these lists are analysed as nested applications of a binary relation, as in Figure 6.9 (i). (Premises are marked with a P; the conclusion with a C.) As regards the relations between premises, the topmost relation links the first premise with a complex span consisting of another relation between two further premises. It is easy to see how this pattern could be extended.

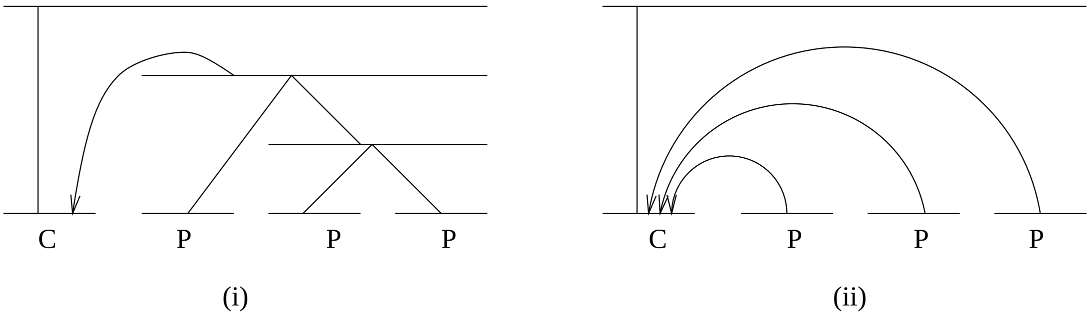

Figure 6.9: Alternative Structural Analyses of *Furthermore*

Figure 6.9 (ii) shows the structural analysis which RST would give for a text containing a sequence of premises. As already outlined in Section 2.4.3, the notion of **multiple schema applications** does service in such texts: the premises are not represented in relation to each other, but in relation to the conclusion they support. The conclusion span is thus related to several adjacent text spans. However, a disadvantage of this approach is that it can only be used if the conclusion is represented explicitly in the text; and this is far from always the case. If a conclusion is implicit, RST would have to analyse the premises using a much less informative relation, LIST. Maier and Hovy (1991) counter this problem by adding a separate level of 'textual' relations to the diagram in Figure 6.9 (ii), which link the adjacent premises (see Section 2.5.3). But in the present system, two levels of relations are not necessary: the relation defined by POSITIVE PRAGMATIC INDUCTIVE features is sufficiently abstract to capture both the relationship between two premises and that between the premises and the conclusion.[8]

6.2.5 Cause and Result-Driven Relations

A further extract from the NEGATIVE POLARITY portion of the taxonomy is given in Figure 6.10. Motivating examples appear below.

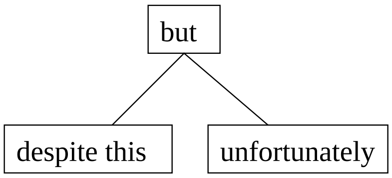

Figure 6.10: Cause-Driven and Result-Driven Phrases

$$\text{Bill took the lid off the pot. } \left\{ \begin{array}{l} \textit{Unfortunately,} \\ \checkmark \textit{ But} \\ \#\ \textit{Despite this,} \end{array} \right\} \text{ there was nothing inside.} \qquad (6.20)$$

[8] A question remains as to the nature of the relation which links the set of premises to the conclusion: it could either be PRAGMATIC POSITIVE CAUSAL or PRAGMATIC POSITIVE INDUCTIVE. I favour the former suggestion; note, for instance that *it follows that* can be used to introduce a conclusion supported by several premises.

$$\text{Sue had been up all night; } \left\{ \begin{array}{l} \textit{despite this,} \\ \checkmark \textit{ but} \\ \# \textit{ unfortunately,} \end{array} \right\} \begin{array}{l} \text{she was looking fresh as a} \\ \text{daisy.} \end{array} \quad (6.21)$$

The text in Example 6.20 can be thought of as presenting an unsatisfied desire. Let us say Bill wants to eat something; one way of achieving this goal would be if the pot contained food and its lid were removed. He lifts the lid, but there is nothing inside. *But* is certainly appropriate to describe the circumstance which blocks fulfilment of the goal; as is *unfortunately* (understood as 'unfortunately *for Bill*'). But *despite this* is quite wrong—it suggests that lifting the lid is expected to cause or entail that the pot is not empty. In Example 6.21, a text with a more conventional violated expectation is presented. Here *but* and *despite this* are acceptable, but *unfortunately* is inadmissible as a substitute.

These two different uses of *but* have been noted by others;[9] the following two examples of the unsatisfied-desire type are respectively from Longacre (1983) and Spooren (1989):

(6.22) I intended to go, but we had visitors that night.

(6.23) I went to the church, but the vicar was not there.

Both of these commentators attempt an explanation of such texts by proposing more abstract ways in which expectations are violated. For Longacre,

> something is presupposed here like the Newtonian assumption (inertia) that a body in motion in a given direction will keep moving in that direction unless some force deflects or stops it...

For Spooren, the expectation arises as a result of implicatures that follow from the statement of intention:

> part of our world knowledge is that going to church probably means that the vicar is in the church.

Neither of these explanations is very convincing. It is far from the case that intentions are normally achieved. It is certainly possible to set up contexts where an intention has no chance of being achieved and yet *but* is still appropriate. Imagine Jim is in a prison cell from which he cannot escape. We could still say that

(6.24) Jim looked around for food, but there was none to be found.

No amount of wanting or looking for food is going to satisfy Jim's intention, so the 'inertia' explanation is ruled out. His looking for food does not probably mean that there is food to be found, so the introduction of implicatures is similarly inadmissible. Yet the phrase *but* does not seem out of place.

[9] However, no-one to my knowledge has suggested using *unfortunately* as a diagnostic for the unsatisfied-desire use.

A preferable explanation—and one that is motivated by the pattern of substitutability in Figure 6.11—is that *but* is undefined for a further feature, for which *despite this* and *unfortunately* signal different values. The values of the feature relate to the manner in which the rule $P_1 \wedge \ldots \wedge P_n \rightarrow Q$ is used: are we *predicting* Q from our knowledge of $P_1 \wedge \ldots \wedge P_n$, or are we seeking to *achieve* Q, and thus investigating whether $P_1 \wedge \ldots \wedge P_n$ are true or themselves achieveable? We can call this feature ANCHOR, to reflect whether the 'certainty' relates to the knowledge of the premises, or the desirability of the conclusion. In each case, the 'certain' thing will be identified as A (for **anchor**). The feature can be defined as follows:

> ANCHOR
>
> CAUSE-DRIVEN: $A \in P_1 \ldots P_n$; $P_1 \wedge \ldots \wedge P_n$ is true.
>
> RESULT-DRIVEN: A corresponds to Q; and A is desired by the protagonist.

To take an example, consider again Text 6.22: *I intended to go, but we had visitors that night.* This is a RESULT-DRIVEN relation: A, the first clause, which presents the writer's intention to go, corresponds to the right-hand side of the rule $P_1 \wedge \ldots \wedge P_n \rightarrow Q$. C is the second clause, presenting the fact which prevents the intention being achieved. The relation is BILATERAL, since C relates to a fact on the left-hand side of the rule $P_1 \wedge \ldots \wedge P_n \rightarrow Q$. It is NEGATIVE POLARITY, since the relevant fact in the rule (C') is inconsistent with C.

In one respect, the above definition of the ANCHOR feature is slightly fudged. Consider the original example of a RESULT-DRIVEN relation: *Bill took the lid off the pot; but there was nothing inside.* The first clause must still be considered A, but note that it does not *itself* present the intention. Rather, it presents an action performed to achieve the intention. There will be more to say about such cases in Section 6.2.7, which deals with presupposition. For the moment, note that the intention behind the action in the current example can be expressed in a subordinate clause:

(6.25) Bill took the lid off the pot *to* get some food; but there was nothing inside.[10]

Note that the revised definitions of POSITIVE and NEGATIVE POLARITY presented in Section 6.2.3 are still serviceable for both CAUSE-DRIVEN and RESULT-DRIVEN relations. In POSITIVE POLARITY relations, nothing is negated, and so no problems arise. And in NEGATIVE POLARITY relations, it is always C which is negated. For a CAUSE-DRIVEN NEGATIVE POLARITY relation, the anchor A is on the left-hand side of the rule, and C is the negation of the expected conclusion. For a RESULT-DRIVEN NEGATIVE POLARITY relation, the right-hand side of the rule is desired by the protagonist, and C is the negation of one of the conditions necessary for this desire to be brought about.

It should also be noted that the definition of PATTERN OF INSTANTIATION in Section 6.2.3 does not need amendment. This definition determines whether C is on the

[10] Strictly speaking, *to* is not a cue phrase, as it does not pass the test for relational phrases. But, as noted in Section 4.2, it is similar enough to a cue phrase to warrant attention.

same side of the rule as A (UNILATERAL), or whether they are on opposite sides (BILATERAL). In combination with the two values of the ANCHOR feature, four possible patterns of instantiation can now be expressed: A and C can both be on the left of the rule; or they can both be on the right; or A can be on the left and C on the right; or C can be on the left and A on the right.

Figure 6.11 provides some additions to the diagram in Figure 6.10. Again, motivating examples are provided:

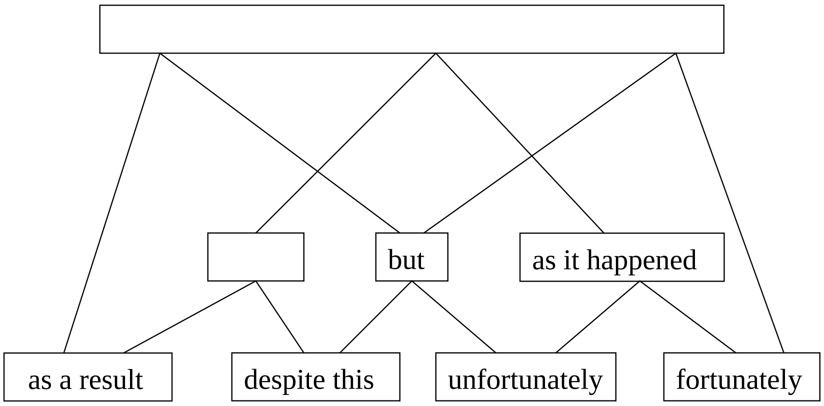

Figure 6.11: Additional CAUSE-DRIVEN and RESULT-DRIVEN Phrases

$$\text{Bill took the lid off the pot.} \left\{ \begin{array}{l} \textit{Unfortunately,} \\ \checkmark\ \textit{But} \\ \checkmark\ \textit{As it happened,} \\ \#\ \textit{Fortunately,} \\ \#\ \textit{Despite this,} \\ \#\ \textit{As a result,} \end{array} \right\} \text{there was nothing inside.} \tag{6.26}$$

$$\text{Bill took the lid off the pot.} \left\{ \begin{array}{l} \textit{Fortunately,} \\ \checkmark\ \textit{As it happened,} \\ \#\ \textit{Unfortunately,} \\ \#\ \textit{But} \\ \#\ \textit{Despite this,} \\ \#\ \textit{As a result,} \end{array} \right\} \text{there was something inside.} \tag{6.27}$$

Note that the exclusivity of *fortunately* and *unfortunately* depends crucially on the fact that the two phrases set up different implicatures about what Bill wants. Both of the above texts presuppose that Bill's plan requires there to be something in the pot. It might also have been that the plan required the pot to be empty—in which case the appropriateness of *fortunately* and *unfortunately* would be reversed. The two phrases cannot be substituted for one another, as to do so requires a change in the assumptions about what Bill's plan involves.

The most significant fact about the diagram in Figure 6.11 is that *as it happened* is substitutable for both *fortunately* and *unfortunately*, while these two latter phrases are exclusive. All three phrases are exclusive with CAUSE-DRIVEN phrases like *as a result* and *despite this*, and it is plausible that they should all be labelled as RESULT-DRIVEN. The variation within the phrases can then be traced to the POLARITY feature. *Fortunately* has POSITIVE POLARITY: the intention behind the first clause is achieved if the second clause is true. *Unfortunately* has NEGATIVE POLARITY: the intention behind the first clause is achieved if the *negation* of the second clause is true. And *as it happened* is undefined with respect to the POLARITY feature.

Conditional Variants of CAUSE-DRIVEN and RESULT-DRIVEN Phrases

Evidence for the distinction between *cause-driven* and *result-driven* relations also shows up in the conditional phrases in the taxonomy. Consider the two examples below:

(6.28) If you feel like cooking tonight, make something spicy.

(6.29) If you feel like cooking, there's a kitchen on the first floor.

Both of these examples are of POSITIVE POLARITY PRAGMATIC relations. In Example 6.28, the rule which the writer wants to succeed is that people who feel like cooking and who have a certain disposition make something spicy. (The writer's aim is thus that the reader takes on this disposition.) But we cannot envisage a similar rule for Example 6.29. The kitchen is on the first floor whether or not the reader feels like cooking. The point is rather that it is only *relevant* for the reader to know where the kitchen is if he feels like cooking. Treating the *if* in this text as RESULT-DRIVEN thus provides a better analysis. We assume there is a goal underlying the first clause; namely that the reader cooks. This will happen if the reader feels like cooking, *and* knows where the kitchen is.[11]

A similar story can be told for NEGATIVE POLARITY relations.

(6.30) Even if you manage to break out of the prison, you'll never make it home.

(6.31) You'll never make it home. Even if you manage to break out of the prison, the jungle round here is impenetrable.

In Example 6.30, the presupposed rule is that if one breaks out of prison, one can normally get home. The rule is defeated in the present case. In Example 6.31, however, there is no rule stating that breaking out of prison normally entails the jungle not being impenetrable. Rather, we must assume a goal behind the breaking out of prison, and a rule stating that the goal will be achieved if the outbreak occurs and the jungle is *not* impenetrable.

[11] An alternative analysis for this kind of text is given by Sweetser (1990). She considers such a text to be an example of a SPEECH ACT conditional, to be read as 'if you feel like cooking, then (let us consider that) I *inform* you that there's a kitchen on the first floor'. According to this analysis, the speech act of informing is only to be understood as having occurred if the information about the kitchen is considered relevant. However, it is odd to suggest that the speech act simply disappears if its content is not relevant. The information about the kitchen is conveyed to the hearer under any circumstances, even if it is not relevant; and it is hard to see what there is to an informative speech act beyond the deliberate conveying of information. The problem is exacerbated if the analysis is extended to cover examples such as *Whenever you feel like cooking, there's a kitchen on the first floor*. If we interpret this as a SPEECH ACT conditional, we must envisage a whole series of informative speech acts, one for each time the hearer feels like cooking. At this point, we are clearly stretching the notion of a 'speech act' beyond its normal use.

The distinction between CAUSE-DRIVEN and RESULT-DRIVEN conditional relations does not emerge systematically from the taxonomy. There are not always phrases which differ only with regard to this feature (although see Section 6.2.8 for a discussion of the phrase *in case*). The only systematic evidence for the feature in hypothetical relations is thus that the phrases *if* and *even if* can take either of its values. However, indirect evidence can be obtained by converting 'hypothetical' texts to 'actual' ones, and observing which cue phrases are now appropriate. The following two texts are 'actual' versions of Texts 6.31 and 6.29; the patterns of substitutability for the phrases *fortunately*, *unfortunately*, *despite this* and *so* are what we would expect for CAUSE-DRIVEN and RESULT-DRIVEN relations.

$$\begin{array}{r}\text{Bill managed to escape}\\ \text{from prison.}\end{array} \left\{ \begin{array}{l} \textit{Unfortunately,} \\ \checkmark\ \textit{but} \\ \#\ \textit{despite this,} \end{array} \right\} \begin{array}{l}\text{the jungle was}\\ \text{impenetrable.}\end{array} \qquad (6.32)$$

$$\text{Bill felt like cooking.} \left\{ \begin{array}{l} \textit{Fortunately,} \\ \#\ \textit{So} \end{array} \right\} \begin{array}{l}\text{there was a kitchen on the}\\ \text{first floor.}\end{array} \qquad (6.33)$$

6.2.6 ANCHOR-BASED and COUNTERPART-BASED Relations

Another portion of the taxonomy is given in Figure 6.12. The motivating examples are as follows:

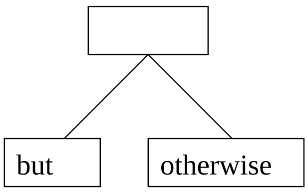

Figure 6.12: ANCHOR-BASED and COUNTERPART-BASED Phrases

$$\text{Bob put his hands up,} \left\{ \begin{array}{l} \textit{otherwise} \\ \#\ \textit{but} \end{array} \right\} \text{Jill would have shot him.} \qquad (6.34)$$

$$\begin{array}{r}\text{Bob kept his hands by his}\\ \text{sides,}\end{array} \left\{ \begin{array}{l} \textit{but} \\ \#\ \textit{otherwise} \end{array} \right\} \text{Jill didn't shoot him.} \qquad (6.35)$$

It seems that both of these texts involve a rule along the following lines:

> If Bob doesn't put his hands up, Jill will shoot him.

For both texts, the anchor relates to the left-hand side of this rule, and the counterpart to the right-hand side. The texts are similar, in that the conclusion of the rule is avoided in each case. However, the reason for this is different in the two cases. In Example 6.35, the premise of the rule is true, but the rule is defeated: some stronger conflicting rule must therefore be supposed to have taken precedence. In Example 6.34, the premise of the rule does not even occur: the protagonist takes action to avoid a conclusion which is inconsistent with his goals.

In order to link the propositions related by the cue phrases onto the premise and conclusion of the relevant rule, another dimension of variation must be introduced. At present, in NEGATIVE POLARITY relations, it is always the *counterpart* span which needs to be negated to map back onto the rule $P_1 \wedge \ldots \wedge P_n \rightarrow Q$. This is still the case for Example 6.35; the counterpart is the second span, which is the negation of the expected conclusion. But for Example 6.34, it is rather the anchor (*Bill put his hands up*) which must be negated to map onto the rule, and the counterpart (*Jill would have shot him*) which maps onto the rule without being negated.

Examples such as this suggest the need for a new feature, which we can call FOCUS OF POLARITY, to specify whether the POLARITY transformation (if there is one) operates on the **anchor** or on the **counterpart**.

The definition of FOCUS OF POLARITY will require another modification to the definition of the POLARITY feature. In the current definition in Section 6.2.3, it is always the counterpart of the rule (C) which is a candidate for negation: the anchor A is never negated. We now need to express the definition so as to allow the candidate for negation to be determined by FOCUS OF POLARITY. To this end, we must introduce some new variables: the **focus of polarity** (F), which is the candidate for negation, and the **invariant** (I), which is never negated. The variables F' and I' will be used to represent F and I after the polarity transformation has taken place. The new definition for POLARITY now looks like this:

POLARITY (3nd definition)

POSITIVE: $F = F'$; $I = I'$;

NEGATIVE: $F = \neg F'$; $I = I'$.

The FOCUS OF POLARITY feature now identifies F and F' with A and A' (and I and I' with C and C') or F and F' with C and C' (and I and I' with A and A'.

FOCUS OF POLARITY

ANCHOR-BASED: $F = A$; $F' = A'$; $I = C$; $I' = C'$.

COUNTERPART-BASED: $F = C$; $F' = C'$; $I = A$; $I' = A'$.

Consider how these new definitions work with the phrases *but* and *otherwise*. For *Bob kept his hands by his sides, but Jill didn't shoot him*, the first clause is A and the second clause C. It is BILATERAL CAUSE-DRIVEN, so A' is part of the left-hand side of $P_1 \wedge \ldots \wedge P_n \rightarrow Q$, and C' is Q. The relation is COUNTERPART-BASED, so F is C and F' is C'. It is NEGATIVE POLARITY, so F is $\neg F'$; which means that C' is $\neg C$. A is the invariant, and hence maps straight onto A'. The effect is a violated expectation.

For *Bob put his hands up; otherwise Jill would have shot him*, the first clause is again A and the second clause C. The relation is BILATERAL CAUSE-DRIVEN, so A' is part of the left-hand side of $P_1 \wedge \ldots \wedge P_n \rightarrow Q$, and C' is Q. This time the relation is ANCHOR-BASED, so F is A and F' is A'. It is NEGATIVE POLARITY, so F is $\neg F'$; which means that A' is $\neg A$. This means that the rule does not trigger, and the right-hand side of the rule (C) does not occur.[12]

Semantic and Pragmatic Anchor-Based Phrases

As it is expressed in terms of A and C, the definition of FOCUS OF POLARITY serves equally well for SEMANTIC relations and for PRAGMATIC ones. Consider the following two cases:

(6.36) Bob put his hands up, *otherwise* Jill would have shot him.

(6.37) Put your hands up, *otherwise* I'll shoot you.

Example 6.36 requires that the reader believe that a protagonist in the world being described (Bill) does not want some eventuality (being shot) to occur. It is thus a SEMANTIC relation, holding between the propositional contents of the related spans. However, for Example 6.37 it is a precondition that someone in the real world (namely the reader) does not want to be shot. The intended effect of the relation is that the reader actually put his hands up in order to avoid this eventuality. (Note that it is not specified whether the eventuality is *in fact* avoided; or even whether it is true that the writer would have shot the reader if he had not obeyed her instruction.)

Distinctions Amongst Anchor-Based Relations

Otherwise is only one of a number of ANCHOR-BASED phrases. There are many others that seem to fall into this category: for example, *or*, *unless*, *until*, and *before* (see Section 6.2.7 for a discussion about this latter phrase). There is not room to talk about them all here, but the dimensions of variation between them look likely to correspond to those identified by other features. Deciding whether this is indeed the case is a matter for further research.

6.2.7 Presupposed and Non-Presupposed Relations

Another informative extract from the taxonomy is given in Figure 6.13. Motivating examples are given in Texts 6.38 and 6.39:

$$\begin{array}{r}\text{I haven't always been unfit.}\\ \text{I played a lot of rugby}\end{array} \left\{\begin{array}{l} \textit{while} \\ \checkmark\ \textit{when} \\ \#\ \textit{meanwhile} \end{array}\right\} \text{I was at college.} \tag{6.38}$$

$$\begin{array}{r}\text{They set about preparing}\\ \text{the meal. Bill marinated}\\ \text{the meat;}\end{array} \left\{\begin{array}{l} \textit{meanwhile,} \\ \checkmark\ \textit{while} \\ \#\ \textit{when} \end{array}\right\} \text{Bob lit the barbecue.} \tag{6.39}$$

All three of the phrases in the diagram convey information about temporal simultaneity, among other things. However, there is a syntactic difference between *when*, which

[12] In fact, the story is likely to be more complicated than this. The reason why C does not occur is *because it is not desired by Bob*, and he takes action to avoid it. There thus seems likely to be a result-driven component to the relation which has not so far been captured.

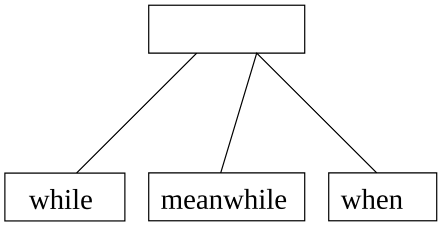

Figure 6.13: PRESUPPOSED and NON-PRESUPPOSED Phrases

is a temporal subordinator, and *meanwhile*, which is a sentential adverb. *While* is appropriate as a substitute for either phrase, as the examples show.

The clauses introduced by temporal subordinators are conventionally thought of as **presupposed** (Karttunen (1973), Keenan (1971), Lascarides and Oberlander (1993)). They describe an eventuality with respect to which the material in the main clause is temporally situated: this eventuality must either be already known to the reader, or must be **accommodated** prior to the addition of the temporal relation. The asymmetry of these sentences can be demonstrated by swapping the main and subordinate clauses. This typically results in incoherence, although the relation of 'temporal overlap' between the clauses is presumably unchanged:

(6.40) I haven't always been unfit. ?? I was at college *when* I played a lot of rugby.

On the other hand, no harm is done by changing the clause introduced by *meanwhile*:

(6.41) They set about preparing the meal. Bob lit the barbecue; *meanwhile* Bill marinated the meat.

The question of when accommodation is possible is addressed by Lascarides and Oberlander (1993). They propose two different mechanisms, which apply in different situations. If the subordinate clause functions simply as a temporal adverbial, as in Example 6.38, then there is no need to find a coherence relation between it and the preceding context; the important relation ('BACKGROUND', in this case) is between the main clause and the preceding context. But in other cases, as in the following example, the subordinate clause has an important narrative function:

(6.42) The backbenchers were in revolt. They were pacified after Major launched a charm offensive.

Here, a coherence relation between the subordinate clause and the preceding context must be found to allow accommodation, and only after the subordinate clause has been attached is the main clause considered. The explanation of the asymmetry introduced by the subordinator thus turns on the *order of attachment* of the two clauses.

A similar explanation of temporal subordinators must be sought for the present theory. Here, however, it must emerge from the definitions of the relations marked by these phrases, rather than from an account of the algorithm used to interpret them. We therefore need to think about how to express the preconditions for a relation between two spans in terms of how these link to the immediately preceding context. We need a feature which takes one value for subordinating phrases like *when*, and another for non-subordinating phrases.

A feature with alternative values PRESUPPOSED and NON-PRESUPPOSED is used here to capture this difference. The feature introduces another defeasible rule, $X_1 \wedge \ldots \wedge X_n \rightarrow Y$, similar in structure to the one which the feature definitions have so far referred to. This rule makes a link between the span in the preceding context (which we will call *Precond*) and the anchor A. We can define the new feature as follows:

> PRESUPPOSITIONALITY
>
> PRESUPPOSED: *Precond* is part of the left-hand side of the rule $X_1 \wedge \ldots \wedge X_n \rightarrow Y$, and A is Y.
>
> NON-PRESUPPOSED: *Precond* is A itself.

We can now give an account of what happens in Example 6.42. The first clause *the backbenchers were in revolt* is *Precond*. The rule $X_1 \wedge \ldots \wedge X_n \rightarrow Y$ has this clause as part of its left-hand side, and anchor clause *Major launched a charm offensive* as its right-hand side. When *Precond* occurs, the rule is thus triggered.

A similar story can be told for other temporal subordinators. Consider the following example:

(6.43) Bob heated the water. When it boiled, he stirred in the sugar.

The first clause *Bob heated the water* is *Precond* here. It can be seen as triggering a causal rule whose right-hand side is *the water boiled*, which is the anchor of the relation signalled by *when*.

We can now consider what happens in NON-PRESUPPOSED relations. An example of such a relation appears in the text below.

(6.44) Jill was curious, so she pulled the lever. Instantly, an alarm went off.

In fact, there appear to be two overlapping relations in this example: one between the first and second clauses (signalled by *so*); the other between the second and third clauses (signalled by *instantly*). It seems quite plausible that the counterpart of the first relation actually *identifies with* the anchor of the second relation. This is what is captured in the definition of NON-PRESUPPOSED: an identity is specified to hold between A and *Precond*.

RESULT-DRIVEN PRESUPPOSED **Relations**

It was noted above, in connection with Example 6.40, that swapping the main and subordinate clauses in a PRESUPPOSED relation leads to incoherence. However, it is interesting that where the subordinate clause has a narrative function (rather than just acting as a temporal adverbial), coherent texts can be created by swapping the main and subordinate clause, provided that the right subordinating phrase is chosen. Consider the following variation on Example 6.42:

(6.45) The backbenchers were in revolt. Major launched a charm offensive to pacify them.

The anchor for the presuppositional relation is now 'The backbenchers are pacified', and this is what must be attached first to the preceding context. However, it is no longer the pacification *itself* which is caused by the context. Rather it is the *intention* that the backbenchers be pacified. We have now set up a context where *Major launched a charm offensive* can act as the anchor for a RESULT-DRIVEN rule of the kind discussed in Section 6.2.5.

ANCHOR-BASED PRESUPPOSED **Phrases**

Finally, it is interesting to note that the rule $X_1 \wedge \ldots \wedge X_n \rightarrow Y$ is defeasible, just like $P_1 \wedge \ldots \wedge P_n \rightarrow Q$. Consider this example:

(6.46) Tidy your room, before I lose my temper.

The relation signalled by the subordinator *before* is PRESUPPOSED; in other words, the subordinate clause 'W loses her temper' is A, and the main clause 'Reader tidies his room' is C. Since it is PRESUPPOSED, there is a rule running from the preconditions of the text (i.e. the situation that is currently true) to A. In other words, A is predicted to happen as things stand. However, the relation is also RESULT-DRIVEN, NEGATIVE POLARITY, and ANCHOR-BASED. These parameters describe the relationship between A and C. They specify that there is a rule $P_1 \wedge \ldots \wedge P_n \rightarrow Q$, whose left hand side includes C, the action which the writer intends the reader to perform, and whose right-hand side is the *negation* of A. 'Not A' is a desire of the reader's: he wants it not to be the case that the writer loses her temper. We thus have two conflicting rules: one leading to A and one leading to $\neg A$. The left-hand side of the former rule is currently true. The left-hand side of the latter rule contains what is currently true *plus* the desired reader action. The latter rule is the one which is intended to fire, and thus the rule leading to A is defeated.

The above account of counterfactual *before* still needs to be worked out in detail. However, it is interesting that the features being developed here seem well-suited for handling such cases.

6.2.8 HYPOTHETICAL and ACTUAL Relations

A final extract from the taxonomy is given in Figure 6.14. It is motivated by the following examples:

$$\text{We had a strict upbringing. } \left\{ \begin{array}{l} \textit{If} \\ \surd\ \textit{When} \end{array} \right\} \begin{array}{l} \text{we were naughty, we were} \\ \text{sent to bed with no supper.} \end{array} \qquad (6.47)$$

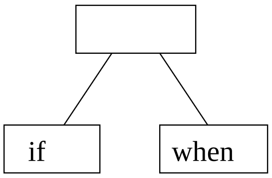

Figure 6.14: ACTUAL and HYPOTHETICAL Phrases

$$\text{You can stay up} \left\{ \begin{array}{l} \checkmark\ \textit{if} \\ \#\ \textit{when} \end{array} \right\} \text{you don't squabble.} \tag{6.48}$$

$$\left\{ \begin{array}{l} \textit{When} \\ \#\ \textit{If} \end{array} \right\} \text{Mary gets home, ask her to call me.} \tag{6.49}$$

In all of these examples, the suitability of the different cue phrases seems determined by whether the subordinate span (A) is known or unknown. A feature called MODAL STATUS can thus be proposed, with alternative values ACTUAL and HYPOTHETICAL.

The MODAL STATUS feature interacts productively with a number of other features. We have already talked about 'conditional' phrases in a number of other places; for instance in connection with the POLARITY feature (*if* versus *even if*) and the ANCHOR feature (for which *if* is undefined). The question is now how to define it.

A simple idea would be to identify the contexts where the anchor A is known as ACTUAL, and those where it is not known as HYPOTHETICAL. However, there are a number of problems with this approach. Consider the case of Text 6.49. It cannot be that the writer actually *knows* that Mary gets home in this example, as it is an event in the future. So what is it which makes *if* and *when* different in this case? It is plausible to suggest that *when* is sanctioned by the writer's ability to *predict* Mary's return before it has happened. It would thus be preferable to define the feature in terms of knowledge (or lack of knowledge) of the *cause* of A rather than of A itself. This being the case, we can thus make use of the variable *Precond* introduced by the PRESUPPOSITIONALITY feature in Section 6.2.7, which for PRESUPPOSED relations such as the above, represents the cause of A. The definition of MODAL STATUS would then be as follows:

MODAL STATUS
ACTUAL: *Precond* is known by the protagonist/writer.
HYPOTHETICAL: *Precond* is not known by the protagonist/writer.

It should also be possible to talk about the HYPOTHETICAL/ACTUAL distinction for NON-PRESUPPOSED phrases. Consider the following examples:

$$\begin{array}{r} \text{Bob piled up the boxes} \\ \text{underneath the bananas.} \end{array} \left\{ \begin{array}{l} \textit{Then} \\ \#\ \textit{In that case,} \end{array} \right\} \begin{array}{l} \text{he was able to reach them} \\ \text{easily.} \end{array} \tag{6.50}$$

$$\begin{array}{r} \text{Bob might try piling up the} \\ \text{boxes underneath the} \\ \text{bananas.} \end{array} \left\{ \begin{array}{l} \textit{Then} \\ \checkmark\ \textit{In that case,} \end{array} \right\} \begin{array}{l} \text{he'll be able to reach them} \\ \text{easily.} \end{array} \tag{6.51}$$

Both *then* and *in that case* are NON-PRESUPPOSED in the above contexts, so *Precond* is identified with the anchor clause, which is the first clause in each case. *Then* seems to be undefined for MODAL STATUS, being appropriate in both contexts; but *in that case* seems to require an anchor which is unknown.

RESULT-DRIVEN HYPOTHETICAL Relations

A possible instance of a HYPOTHETICAL RESULT-DRIVEN phrase is *in case*. Consider the following text:

$$\text{Bill tidied the house, } \left\{ \begin{array}{l} \textit{in case} \\ \#\ \textit{because} \end{array} \right\} \begin{array}{l} \text{his parents came home} \\ \text{early.} \end{array} \qquad (6.52)$$

This text requires an inference about one of Bill's goals; namely that he does not want his parents to come home early and find that the house is not tidy. It is not certain that his parents will come home early—hence the relation is HYPOTHETICAL—but it is sufficiently likely to make Bill tidy the house. We can therefore assume that the two related spans *Bill tidied the house* and *his parents came home early* are both part of the left-hand side of some rule whose right-hand side is a state of affairs desired by Bill, such as 'Bill's parents are not angry'. Note that the ACTUAL phrase *because* is inappropriate for signalling this relation.

6.2.9 SEMANTIC and PRAGMATIC Relations Revisited

This final section addresses a number of issues and problems that arise in connection with the SEMANTIC/PRAGMATIC distinction. These will first be outlined, and then some suggestions for a solution will be put forward.

A first observation is that the present definitions of SEMANTIC and PRAGMATIC bring together two quite different ideas: on the one hand, the issue of whether A and C represent the propositional content of the related clauses (SEMANTIC) or their intended effects (PRAGMATIC); and on the other, that of whether the intended effect of the whole relation is that the reader *believe* a relation between two propositions (SEMANTIC) or that a relation between two propositions is *actually the case* (PRAGMATIC). The latter distinction seems to define whether the relation is part of a description or narrative, which the reader accepts without question, or whether it takes place in the real world, where the writer's goals are not just communicative. There seems no reason a priori why these two dimensions should be related.

Indeed, it is not hard to find examples of prototypically PRAGMATIC cue phrases in purely narrative discourse. For instance, *furthermore* and *then again* can both feature in 'free indirect speech', where an agent's thought processes are being described:

(6.53) Sally couldn't decide who would win the match that evening. Spurs were at home; *furthermore*, they were on good form. *Then again*, they were playing the league champions...

But the intended effect of these relations is surely still descriptive rather than persuasive.

Conversely, going by the current definitions, it is hard to find any phrases which are purely SEMANTIC. A temporal phrase like *after this* is a plausible candidate, but such phrases can be used to link imperative clauses, which currently count as PRAGMATIC:

(6.54) Sweep the floors. *After this*, tidy the cupboards.

At the same time, *furthermore* and *after this* are exclusive phrases, as was noted in the very first extract to be presented from the taxonomy in Section 4.6.2. The motivating examples are reproduced below:

$$\begin{array}{r}\text{Television is bad for us. It}\\ \text{kills creativity;}\end{array}\left\{\begin{array}{l}\textit{furthermore,}\\ \#\ \textit{after this}\end{array}\right\}\begin{array}{l}\text{it promotes an unhealthy}\\ \text{kind of `crowd mentality'.}\end{array} \qquad (6.55)$$

$$\begin{array}{r}\text{Bob set about cleaning the}\\ \text{house. He swept the floors;}\end{array}\left\{\begin{array}{l}\textit{after this,}\\ \#\ \textit{furthermore,}\end{array}\right\}\text{he tidied the cupboards.} \qquad (6.56)$$

We need to find a feature to account for this exclusivity. Clearly, the present definitions of SEMANTIC and PRAGMATIC are not able to.

A suggestion for remedying the above problems comes from considering another of the distinctions motivated from the taxonomy; that between CAUSE-DRIVEN and RESULT-DRIVEN phrases (see Section 6.2.5). The point is that some of the work being done by the current SEMANTIC/PRAGMATIC distinction appears to overlap with work being done by this feature. The definition of PRAGMATIC talks about the 'intended effects' of utterances S_A and S_C; the definition of RESULT-DRIVEN also talks about an agent's intention, and an action or actions which are caused by this intention. It is interesting to speculate that the writer's utterances could be represented as goal-driven actions, just as are the actions of the agents which the writer talks about. To take just one piece of evidence for this line of reasoning: the notion of RESULT-DRIVEN NEGATIVE POLARITY relations seems to find useful application in analysing the kind of *but* which occurs in dialogues, between two speakers. Consider the following exchange:

(6.57) A: *Go to bed.*
B: *But I haven't done my homework yet...*

Just as with other RESULT-DRIVEN relations, it is odd to analyse the second span as 'violating an expectation' set up by the first span, and preferable to think of it as defeating a *goal* underlying the first span. In this case, however, the goal is that of the first speaker, rather than that of a protagonist being described in the text. This idea will not be pursued here, as the present work is not concerned with inter-speaker relations. But it is an interesting thought that such relations might eventually be netted in by the theory.

6.3 Summary of Features Motivated

Now that a number of features have been individually motivated, we can begin to put them together to build up the complex definitions required for cue phrases and relations. Until now, feature definitions have been presented individually. The complete

Name of Feature	Possible Values	
SOURCE OF COHERENCE	SEMANTIC	PRAGMATIC
ANCHOR	CAUSE-DRIVEN	RESULT-DRIVEN
PATTERN OF INSTANTIATION	UNILATERAL	BILATERAL
FOCUS OF POLARITY	ANCHOR-BASED	COUNTERPART-BASED
POLARITY	NEGATIVE	POSITIVE
PRESUPPOSITIONALITY	PRESUPPOSED	NON-PRESUPPOSED
MODAL STATUS	HYPOTHETICAL	ACTUAL
RULE TYPE	CAUSAL	INDUCTIVE

Figure 6.15: The Features So Far Motivated, and Their Alternative Values

set of definitions for the features so far motivated can be found in Appendix C; here, the interactions between the definitions can be more easily appreciated. For the moment, a summary of the eight features and their possible values is given in Figure 6.15.

Figure 6.16 presents a larger extract from the taxonomy, which draws together a selection of the phrases which have so far been discussed, labelled with the feature values which have so far been established. The feature-theoretic interpretation of the taxonomy is clearly illustrated here: exclusive phrases are defined for alternative values of at least one feature; hyponyms inherit all of the feature values associated with their hypernyms and are defined for other features in addition, and so on.

The diagram in Figure 6.16 is complex: it documents many of the substitutability relationships shown in the smaller diagrams in Section 6.2, as well as many relationships between phrases which appeared in different diagrams. The figure divides roughly into four exclusive groups of phrases:

- POSITIVE POLARITY CAUSAL ACTUAL relations (dominated by the phrase *so*);
- NEGATIVE POLARITY ACTUAL relations (dominated by the phrase *but*);
- HYPOTHETICAL relations (involving *if* and *even if*);
- POSITIVE POLARITY INDUCTIVE relations (the single phrase *furthermore*).

Much of the complexity in the diagram is due to high-level phrases such as *and* and *while*, which cut across these divisions.[13]

The features with which the phrases are labelled are unlikely yet to be sufficient as definitions, as many additional features have still to be motivated from the taxonomy. Even in this diagram—still just a small portion of the overall taxonomy—there remain relationships which are not yet explained by the features provided. (For instance, the contingent substitutability between *while* and *whereas* remains unexplained. So does

[13] The diagram in Figure 6.16 is already quite difficult to read—clearly, extending it to encompass *all* the phrases in the corpus would soon lead to problems. It is for this reason that the complete taxonomy in Appendix B is divided into a number of separate diagrams when it is presented; see Section 4.6.4.

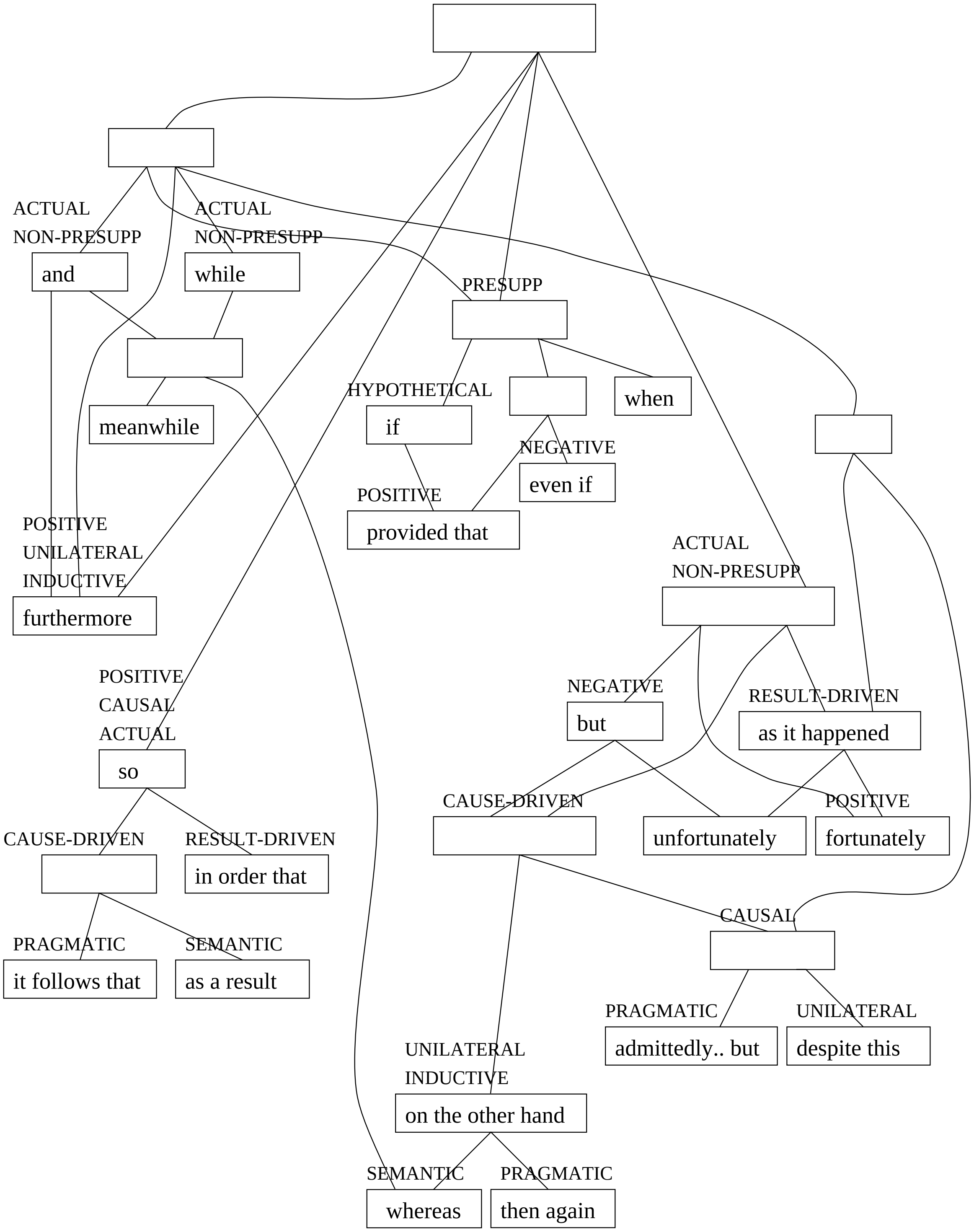

Figure 6.16: Summary of Motivated Features: A Labelled Extract from the Taxonomy

the exclusivity between *and* and *it follows that*.) However, the current set of features already give good approximate definitions in most cases, and at very least serve to indicate the flavour of the definitions which will eventually be reached.

A larger set of composite definitions is given in Appendix D, this time in tabular form. Again, the definitions are not always complete, but they already begin to provide a reasonable account of the variations between the phrases.

6.4 The Mapping between Cue Phrases and Relations

It is useful to sum up what has happened so far in the theoretical interpretation of the taxonomy. In Chapter 5, it was noted that the taxonomy lends itself well to a conception of relations as collections of independent features, and some general principles governing the motivation of features were put forward. In the first part of this chapter, a number of features were systematically motivated by considering small excerpts from the taxonomy one by one. In this section, the final step in the methodology is considered: how the features should be combined to give a set of coherence relation definitions.

Until now, the assumption has been that there will be a one-to-one correspondence between relations and the cue phrases at the *leaves* of the taxonomy. Cue phrases higher up in the taxonomy do not correspond to 'more general' relations, but are used to signal some components of a relation in circumstances where the reader is able to infer the others from context and background knowledge. We begin, therefore, by looking at the cue phrases at the leaves of the taxonomy.

6.4.1 An Uneven Distribution at the Leaves of the Taxonomy

It should be clear by now that even the most specific 'leaf-level' cue phrases in the taxonomy may still be undefined with regard to certain features. For instance, as noted in Section 6.2.5, *if* is undefined for the ANCHOR feature; but there are no phrases below *if* to distinguish between the alternative values of this feature (CAUSE-DRIVEN and RESULT-DRIVEN.) In fact, it may be that when all the features necessary to describe the taxonomy are found, few if any leaf-level phrases will be defined for every one of them.

It might also be that when the possible combinations of feature values are investigated more extensively, some combinations will be found which are not signalled by *any* cue phrase, even a general one which requires feature values to be inferred. Again, this remains an open question.

In short, the phrases at the leaves of the taxonomy are unevenly distributed over the space of possible feature value combinations. Some phrases, being undefined for various features, can be used for a wide range of possible combinations; there may also be possible combinations which are not signalled by any cue phrase. The mapping between leaf-level phrases and possible feature value combinations is thus many-to-one, and incomplete.

This is in no way a disadvantage of the theory. There is no reason to expect a perfect mapping between leaf-level phrases and possible value combinations, as was stressed in Section 5.2.4. And relations are all abstractions, so the idea of some features being undefined in a given relation is quite understandable.

6.4.2 Relations at Different Levels of Abstraction

A trickier problem is raised by the patterns of substitutability illustrated in Figure 6.17. Consider first the extract in 6.17 (i). In this case, the hypernym is *after*, and the

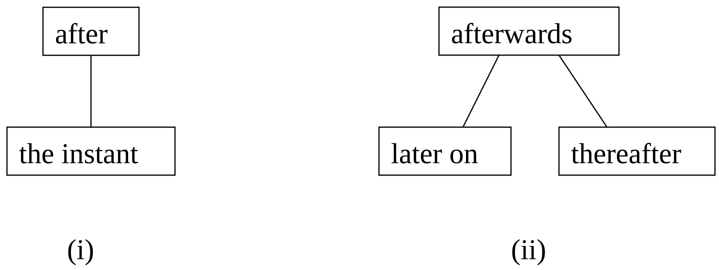

Figure 6.17: Some 'Problem' Extracts from the Taxonomy

hyponym is *the instant*, which in the example we can take to be a leaf node. According to the current hypothesis, *the instant*, being a leaf node, signals all the features of some particular relation: *after* is used to signal this same relation if the extra information carried by *the instant* is easily inferrable from context. The odd thing is, that there should *never* be any need to use *the instant* to signal the relation in question. *After* can only ever mark one possible relation; that is the relation marked by *the instant*. So why does the phrase *the instant* exist at all? If *after* could signal two different relations, then *the instant*'s existence would be understandable: it would be used whenever the information to distinguish between these two relations was not inferrable (in which case *after* would not identify the relation). But there is just one relation in the present case: we know this because *the instant* has no sister phrases, and the whole rationale of the taxonomy is that a cue phrase will exist for *each relation*.

So we have a problem in trying to explain this type of pattern in the taxonomy. In fact, we have exactly the same type of problem with the pattern in 6.17 (ii), where the leaf nodes *later on* and *thereafter* are contingently intersubstitutable, and have a common hypernym *afterwards*. It might be thought that in this case there are two different relations, picked out uniquely by *later on* and *thereafter* respectively. In this case, *afterwards* is able to signal either relation, provided that the information needed to distinguish between them is inferrable from context. However, the phrases *later on* and *thereafter* could not be used to pick out their respective relations uniquely: they are contingently intersubstitutable, which means that in some contexts, they can be substituted for one another. In these contexts, there are no cue phrases for identifying either relation exclusively; again, this goes against the rationale of the whole taxonomy. An alternative suggestion is that *later on* and *thereafter* signal the same relation but identify different subsets of its features, making them suitable for use in different contexts, when different features will be inferrable. On this hypothesis, *afterwards* is to be used to signal this same relation, in contexts where *all* those features unspecified by either *later on* or *thereafter* are safely inferrable. However, this makes the situation identical to that in 6.17 (i): if there is just a single relation, there is no need for the phrases *later on* and *thereafter* at all.

Patterns such as those illustrated in Figure 6.17 are common in the taxonomy, and some explanation is called for. One explanation is that the relations which cue phrases signal are not all associated with the cue phrases at the leaves of the taxonomy: some are associated with cue phrases at higher levels. In other words, *some relations are more abstract than others*. This idea permits an explanation of patterns like those in Figure 6.17: in each case, the hypernymic phrases signal relations at a higher level of abstraction, subsuming those relations signalled by the hyponymic phrases. Our general assumption that 'for every relation there exists a cue phrase which uniquely identifies it' now actually *requires* such hypernymic phrases to exist.

To give an example: in Figure 6.17 (i) we can posit two relations; one signalled by the leaf node phrase *the instant*, and another (this one more abstract) signalled by the hyponymic phrase *after*. *The instant* uniquely identifies one particular relation. *After*, at another level of abstraction, picks out another relation: there is no other cue phrase which can pick out all the occurrences of this relation.

It might be thought that this scenario permits a different kind of ambiguity: if a writer uses the hypernym *after*, in a situation where the relation associated with *the instant* can be inferred from context, how would the reader know whether the more abstract or the less abstract relation is intended? Both would appear to be possible. However, more careful consideration about what an 'abstract relation' is will show that this kind of 'ambiguity' presents no real problem for the reader.

Levels of Abstraction in Text Processing

In Chapter 3, when the conception of relations as cognitive constructs was being outlined, Rosch's notion of the **basic level** of categorisation—the level of abstraction at which the trade-off between usefulness and generality is optimised—was invoked. Rosch suggested that humans would operate at this level when performing certain tasks with a heavy information-processing load. I argued that tasks like reading and writing are likely to be mediated by constructs at a similar level of abstraction; and proposed to think of coherence relations as modelling such constructs.

The hypothesis now under consideration, that 'relations exist at different levels of abstraction', is quite consistent with this conception of relations. A writer's information-processing load can vary considerably, reflecting factors such as time pressure or the difficulty of the task at hand. It is likely that as the information processing load varies, so too does the optimal level of abstraction at which the writer should operate: the higher the load, the greater the degree of abstraction. To take an example: if the writer has to produce a text in a hurry, or if the elements of the text plan are hard to manipulate, then maybe the writer's planning operators should be less tightly defined. Likewise, if the writer's time constraints are relaxed, then more features of a relation can be taken into account.

Many researchers have postulated relations at different levels of abstraction (eg Maier and Hovy (1991), Hovy *et al* (1992), Mann and Thompson (1988)). But no-one has yet come up with a clear reason for doing this. However, if we accept the argument on which the present project is founded—that for any relation that people actually use, there will exist a cue phrase which uniquely marks it—then we have a genuine motivation for postulating the existence of relations at different levels of abstraction. And by drawing on the psychological conception of relations developed in Chapter 3, we have a way to understand why different levels of abstraction should ever be used.

Motivating Relations from High-Level Cue Phrases

The question of *when* a high-level cue phrase motivates a 'more abstract' relation is still to some extent an open one. Given patterns like those in Figure 6.17, we have definite evidence for different levels of abstraction. But consider a pattern such as that in Figure 6.18, where two sister leaf nodes are exclusive. Here, two possibilities can

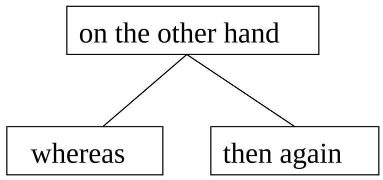

Figure 6.18: Exclusive Sister Leaf Nodes

be discerned. The original story was that *on the other hand* can signal two different relations, which are identified uniquely by *whereas* and *then again*. But, in the light of the current discussion, it might also be that *on the other hand* corresponds directly to a 'more abstract' relation. The answer to this question is a matter for further research. For the moment, I will assume that abstract relations are only motivated where patterns such as those in Figure 6.17 are found. In cases such as that in Figure 6.18, no abstract relation will be postulated.

6.5 Summary

This chapter has presented the beginnings of a feature-theoretical description of the taxonomy of cue phrases. Eight two-valued features have been proposed to account for various dimensions of variation amongst the phrases in the taxonomy. Definitions of the features are summarised in Appendix C, and a preliminary table of relation definitions, expressed in terms of these features, is given in Appendix D.

There remains much work to be done, of course. The taxonomy contains a vast amount of substitutability data, and only a small portion has been looked at. The features so far motivated only cover some portions of the taxonomy. Moreover, feature definitions have been expressed at a relatively informal level: more precise definitions would certainly be preferable, both from a theoretical point of view and as a precursor to implementation. Finally, a great deal more effort is needed to investigate all the possible combinations of the features presented here. Again, this is an area where much further work is required.

All the same, the preliminary conclusions reported here will be useful as the foundation for a more complete account of the phrases in the taxonomy. At very least, they serve as a convincing demonstration of the utility of the methodology being proposed. In particular, they provide good support for the decision taken in Chapter 5, to look for a completely productive set of features. The directions in which further progress can be made are thus quite clearly indicated by the ideas developed in this chapter.

Chapter 7

An Evaluation of the Substitution Methodology

This chapter contains a discussion of some of the potential problems with the substitution methodology proposed in Chapters 3 to 6.

A first problem relates to the substitutability test itself: there are some cases where it seems unable to provide the kind of clearcut data necessary to motivate a feature-based account of relations. This problem will be discussed in Section 7.1; as a solution, an empirical test is suggested which promises a clearer picture about these cases.

Two other problems concern the set of relations eventually motivated by the substitutability test—there are many respects in which the set of relations seems unable to provide a complete coverage of texts. For one thing, it is not hard to find pairs of sentences or clauses in a coherent text for which no cue phrase at all seems appropriate. Since our set of relations is based on the set of cue phrases, such contexts are beyond the scope of the theory. This problem is considered in Section 7.2; a solution is proposed by appealing to the concept of **focus** as better suited for an explanation of these contexts.

Finally, in Section 7.3, the question of relations between large sections of text is raised. It might be thought that cue phrases are only suitable for signalling relations at a low level of hierarchy in a text. However, it is argued that the presence of anaphoric elements in many cue phrases allows them in practice to signal relations between very large sections of text. A discussion of such 'high-level' cue phrases proves to be of relevance to the issue of the interaction of theories of relations and of focus.

7.1 Limitations of the Substitutability Test: The Case of Presentational Sequences

Some relationships in the taxonomy of cue phrases seem more clearcut than others. No-one, for instance, would claim that the phrases *nevertheless* and *on the grounds that* are anything other than exclusive. However, in other cases, our intuitions seem to be less certain. This is particularly so for decisions about phrases signalling what some researchers have called 'presentational sequence'—phrases such as *moreover*, *for another thing* and *furthermore*. For instance, consider the following examples:

The Tories have done a terrible job in government over the last decade. Their early policies led to huge unemployment. Their increasing isolationism over Europe wasted a precious opportunity to gain international influence.	$\left\{ \begin{array}{l} \textit{Furthermore,} \\ \textit{? Later,} \end{array} \right\}$	they reduced Britain to a laughing stock with their 'back to basics' campaign.	(7.1)

Let us review the sequence of governmental blunders during the past decade. Their early policies led to huge unemployment. Their increasing isolationism over Europe wasted a precious opportunity to gain international influence.	$\left\{ \begin{array}{l} \textit{Later,} \\ \textit{? Furthermore,} \end{array} \right\}$	they reduced Britain to a laughing stock with their 'back to basics' campaign.	(7.2)

The taxonomy represents *furthermore* and *later* as exclusive, based on examples such as these. The claim is that in Example 7.1, replacing *furthermore* with *later* changes the text from an argument to a temporal sequence; and that in Example 7.2, replacing *later* with *furthermore* changes the text from a temporal sequence into an argument. Indeed, there does seem to be some kind of difference between the texts. But nevertheless, it is still possible to imagine a writer replacing one phrase by the other. Certainly, it is easier to imagine this than to imagine replacing *nevertheless* with *on the grounds that*.

Another problematic case is given in Example 7.3; here, it is unclear whether or not *whereas* can be replaced by *furthermore*.

It's crazy to keep Bill and Bob in their present positions: we should swap them. Bill is a better player in attack;	$\left\{ \begin{array}{l} \textit{whereas} \\ \textit{? furthermore,} \end{array} \right\}$	Bob is a better player in defence.	(7.3)

The original phrase seems to imply that there is only one reason why Bill and Bob should be swapped, namely that they are better suited to each other's positions. *Furthermore*, on the other hand, suggests that there are two independent reasons why Bill and Bob should be swapped. Again, the difference seems to be one of emphasis—ultimately, the same information is derivable from both texts—but there is nonetheless a difference. Should this difference be treated as theoretically significant or not?

The problem is that in order to answer this question, the tester has to stop being a normal reader/writer, and start being a 'discourse analyst'. Should the alternative versions of the text be given the same analysis, or different ones? Such questions require the kind of 'post-theoretical' intuitions which were called into question at the beginning of the thesis in Section 1.3.2—we cannot be sure if they are reliable, and there seems to be no way of resolving any disagreements which might occur. Ideally, therefore, the test for substitutability should not be relied on in such cases.

It is not surprising that disagreements do in fact occur as to how to analyse texts like those above. The disagreements turn on alternative theories of **span structure**—see Section 2.2.3. Two alternative positions can be identified; these are set out in the following two sections.

7.1.1 The Simultaneous Representation Hypothesis

One line of thought is that *later* and *furthermore* do not signal contradictory information; they simply make different unconnected features of the text explicit. On this story, the two phrases would be contingently substitutable. The relation between the two spans in Examples 7.1 and 7.2 contains both temporal information and information about the structure of an argument, and the phrases are used to signal one or other of these two types of information. Likewise, the relation between the two spans in Example 7.3 contains argumentative information as well as information about a contrast between the two premises; so *whereas* and *furthermore* should also be considered as contingently substitutable. The analysis for these texts would look something like that given in Figure 7.1.

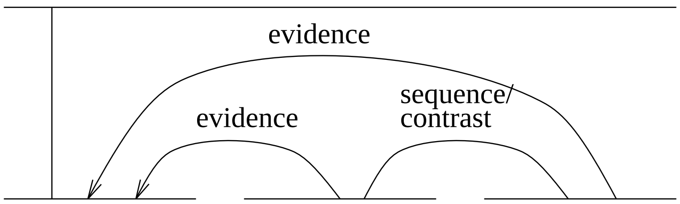

Figure 7.1: The Simultaneous Representation Hypothesis for Texts 7.1–7.3

Such an analysis seems to be espoused by Halliday and Hasan (1976). The following example is referred to:

(7.4) Next, previously to this he had already offered to resign.

The apparent contradiction in this sentence is explained by noting that two different kinds of sequence are being signalled; one INTERNAL (argumentative) and one EXTERNAL (temporal). Martin (1992) also suggests that INTERNAL and EXTERNAL relations can apply simultaneously. Maier (1993) perhaps comes closest to espousing the view outlined above, giving analyses very similar to that in Figure 7.1. For Maier, *furthermore* signals a TEXTUAL relation, and *whereas* and *later* signal *ideational* relations, and these two types of relation can co-occur in a text.

7.1.2 The Dominant Representation Hypothesis

The alternative to the simultaneous representation hypothesis can be referred to as the **dominant representation hypothesis**. According to this hypothesis, there is a significant difference between a text marked with *furthermore* and one marked with *whereas* or *later*, which makes both these latter phrases exclusive with the former.

According to this view, the texts with *whereas* or *later* should be analysed as in the diagram on the left of Figure 7.2, as containing a single complex premise, while the texts with *furthermore* should be analysed as in the diagram on the right, as containing two independent premises.

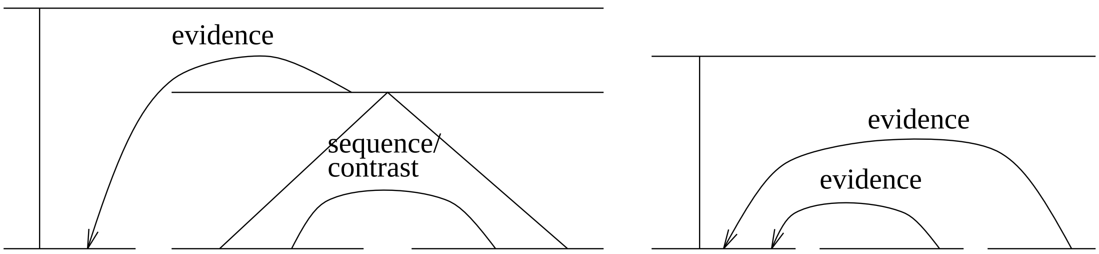

Figure 7.2: The Dominant Representation Hypothesis for Texts 7.1–7.3

This is the view taken by Mann and Thompson (1988): their schemas do not permit the type of diagram shown above. It is also the view taken in this thesis. It will be argued here that although the above texts *can* be analysed using simultaneous relations, thinking of relations as modelling psychological constructs makes this analysis implausible. It seems likely that for a reader or a writer, one of the relations will always be more relevant than the other, and will dominate a representation of the text. This is not to say that when one reading dominates, the alternative reading is *not available at all* to the reader. Clearly, both the argumentative information and the temporal (or contrastive) information can be inferred from the text if it is considered with any serious attention. It is only suggested that this does not typically happen in what we might call 'normal' reading or writing.

This hypothesis can be supported on grounds of cognitive economy. Consider the case of a writer planning the text in Example 7.1. The writer's intention is to persuade the reader that the Tories have done a terrible job over the past decade. Her tactic for achieving this goal is to state a series of facts, each of which will individually push the reader towards this conclusion. The writer's task is then to determine which facts fall into this category: in other words, she needs to find a collection of facts which stand in the relationship of PRAGMATIC ADDITIVITY with each other. Once she has found such facts, there is no need for her to ascertain anything else in order to pursue her strategy. In particular, it is not important to ask about the temporal ordering of the facts: communicating this information will not contribute towards the overall persuasive goal.

Now consider Example 7.2. Here, the overall goal (as set out in the first sentence) is to present to the reader a sequence of events. In order for the writer to achieve this goal, she clearly needs to verify the temporal order of the events. However, it will not be necessary to consider whether or not the events act as premises for some common conclusion: this question is just not relevant to the writer's goals.

Given these facts, it is plausible to suggest that writers concentrate on either the temporal relation or the argumentative relation when constructing text. The reader of a text can be expected to have a similar bias one way or the other, given that he is really trying to make a coherent representation of the text, which ultimately includes a representation of the writer's intentions.

It might be objected to this account that the writer may not have a single overriding intention behind producing a piece of text. It is conceivable that she has two separate intentions; one to persuade the reader of a particular fact, and one to tell the reader about a sequence of events. It thus makes sense to have a strategy for killing the two birds with one stone in a single piece of text: in order to analyse this text, it will be necessary to represent both argumentative and temporal relations between its spans.

Of course, it *is* possible to have two separate intentions simultaneously. To take an unrelated example: I can have an intention to buy an ice cream, and at the same time one to write a letter to my parents. However, there is no reason to expect these two intentions to co-occur with particular regularity; and therefore no reason to expect that any special strategy will have evolved to deal with them both in the same action. Likewise, there is no reason to expect a regular co-occurrence between the intention to convey a temporal sequence of events and the intention to argue for a given conclusion. So, if we are interested in modelling the specialised strategies a writer can make use of for conveying intentions in a text, we are unlikely to need to represent a special mechanism whereby both these intentions can be achieved simultaneously.

7.1.3 An Experimental Design for Testing the two Hypotheses

Whatever the arguments put forward for the alternative hypotheses, it is unlikely that either will be conclusive by itself. However, other empirical means for deciding between them may be more promising. In this section, an experiment is outlined which forces the two hypotheses to make different measurable predictions.

The Form of the Experiment

We will begin by considering the case of Texts 7.1 and 7.2. The dominant representation hypothesis suggests that readers and writers concentrate on the argumentative relation *or* on the temporal relation. The simultaneous representation hypothesis suggests that readers and writers should be able to concentrate on both relations at the same time.

To decide between these alternatives, a experiment is proposed in which subjects read a text containing both temporal and argumentative information, and then answer questions which relate specifically to one type of information or the other. Different reading conditions can be created by varying the cue phrase used in the text: *furthermore* emphasises the argumentative information, while *later* emphasises the temporal information; finally, an unmarked version of the text is neutral between the two.

The text used in the earlier examples can be adapted for this purpose. A neutral introductory sentence is used, so that both *furthermore* and *later* result in coherent texts:

The sequence of governmental blunders during the past decade makes for interesting reading. Their early policies led to huge unemployment. Their increasing isolationism over Europe wasted a precious opportunity to gain international influence.	$\left\{ \begin{array}{l} \checkmark \textit{Later,} \\ \checkmark \textit{Furthermore,} \end{array} \right\}$	they reduced Britain to a laughing stock with their 'back to basics' campaign.	(7.5)

When the readers have read the text, they are asked to verify a number of statements, including the following pair:

S1 The Tories were isolationists over Europe before 'back to basics' was launched.

S2 The Tories can be criticised on several grounds.

The subjects' reaction times will be monitored, to give an indication of how easy they find it to decide about these statements. We can then make comparisons between the subjects' performance after reading one of the alternative marked texts, and their performance after reading the unmarked text.

Predictions of the Two Hypotheses

It seems reasonable to predict that the text using *later* will facilitate the verification of S1 compared to the unmarked text; given that it makes explicit the temporal relation between the two propositions in question. Likewise, it seems reasonable to predict that the text with *furthermore* will help readers verify S2 more than the unmarked text. The interesting question is how the *later* text influences the decision about S2, and how the *furthermore* text influences the decision about S1. The simultaneous representations model in these cases predicts the null hypothesis; namely that

- verifying S2 will be just as easy after reading the *later* text as after reading the unmarked text;
- verifying S1 will be just as easy after reading the *furthermore* text as after reading the unmarked text.

These predictions follow from the claim that both the temporal and the argumentative relations in the original text should be represented: they are independent aspects of its meaning. By marking just one aspect, we should in no way be inhibiting the other, which should be just as clear as in the unmarked case.

Different predictions are made by the dominant relation hypothesis: according to this, the explicit signalling of the temporal information will *inhibit* the representation of the argumentative information, and vice versa. Thus

- verifying S2 will be *harder* after reading the *later* text than after reading the unmarked text;

- verifying S1 will be *harder* after reading the *furthermore* text than after reading the unmarked text.

If these predictions were borne out, there would be a good case for arguing that temporal and argumentative relations are thought of by the reader as real alternatives to each other, and that they are not expected to co-occur. If the reader finds an explicit temporal marker, it actually prejudices him against looking for an argumentative reading; while if he finds an argumentative marker, it prejudices him against looking for a temporal reading. For this reason, it seems legitimate to represent the two markers as exclusive in the taxonomy: they really provide the reader with contradictory information.

Testing the Hypotheses for *Furthermore* and *Whereas*

A similar experiment can be set up to test the predictions made by the two hypotheses about Example 7.3. This text can be given in three conditions; with *whereas*, with *furthermore*, and with the neutral null cue phrase:

$$\begin{array}{r}\text{It's crazy to keep Bill and Bob in their}\\ \text{present positions: we should swap them.}\\ \text{Bill is a better player in attack;}\end{array} \left\{\begin{array}{l} ?\ \textit{whereas} \\ ?\ \textit{furthermore,} \\ ? \end{array}\right\} \begin{array}{l}\text{Bob is a}\\ \text{better player}\\ \text{in defence.}\end{array} \qquad (7.6)$$

The statements to be verified in the subsequent decision task are now as follows:

S1 Bill and Bob are different.

S2 There are two reasons why we should swap Bill and Bob.

Again, the dominant representation hypothesis would predict that the text with *whereas* would slow down the subjects' response to S2 as compared with the null marker, while the text with *furthermore* would slow down the response to S1. The simultaneous representation hypothesis would not predict slowed down responses.

A Control Condition: Contingently Substitutable Phrases

In the case of two phrases which are 'genuinely' contingently substitutable, we would not expect to find one phrase inhibiting the reading signalled by the other. We could thus run a control condition, using clearly contingently substitutable phrases like *once* and *as soon as* (see Figure 7.3).

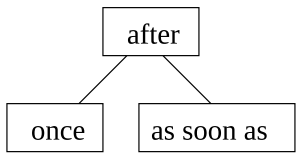

Figure 7.3: Contingently Substitutable Phrases

Both phrases signal a temporal succession between the two related spans, but *once* signals in addition that they are related as cause and effect, while *as soon as* signals in addition that the temporal succession is very rapid. As the following two examples show, *as soon as* can be used in non-causal contexts, and *once* can also be used where the temporal succession is not immediate:

$$\begin{array}{r}\text{We got the fright of our}\\ \text{lives last night coming back}\\ \text{from Sue's party.}\end{array} \left\{ \begin{array}{l} \textit{As soon as} \\ \#\ \textit{Once} \\ \checkmark\ \textit{After} \end{array} \right\} \begin{array}{l}\text{we had stepped out of the}\\ \text{house, we heard a huge clap}\\ \text{of thunder.}\end{array} \tag{7.7}$$

$$\left\{ \begin{array}{l} \textit{Once} \\ \#\ \textit{As soon as} \\ \checkmark\ \textit{After} \end{array} \right\} \text{the war ended, the country's economy gradually improved.} \tag{7.8}$$

Note that the phrase *after*, being a common hyponym, can be used in both contexts: it signals nothing in addition to the temporal succession relationship.

In a context describing an immediate causal succession, a writer could use any one of the three phrases:

$$\begin{array}{r}\text{The thieves didn't have}\\ \text{long to search Jones'}\\ \text{apartment.}\end{array} \left\{ \begin{array}{l} \textit{After} \\ \checkmark\ \textit{As soon as} \\ \checkmark\ \textit{Once} \end{array} \right\} \begin{array}{l}\text{he had fallen asleep, they}\\ \text{set to work as fast as they}\\ \text{could.}\end{array} \tag{7.9}$$

In such a case, there are no niggling suspicions that by substituting *as soon as* for *once* (or vice versa) we are somehow contradicting a presupposition set up in the original text. It seems legitimate to talk about 'causality' and 'immediacy' as independent features of the reader's model of the text, which can both be represented simultaneously. If this is indeed the case, we would not expect that *once* actively inhibits the information about immediacy, or that *as soon as* inhibits the information about causality. Again, we could create decision tasks for subjects to test this prediction; for example about the following alternative statements:

S1 The thieves had to wait for Jones to fall asleep before they could search his apartment.

S2 The thieves searched Jones' apartment immediately after he fell asleep.

Here, our predictions conform to the null hypothesis—that verifying S2 will be no harder after reading the *once* text than after reading the neutral *after* text; and that verifying S1 will be no harder after reading the *as soon as* text than after reading the neutral text.

7.1.4 A Revised Role for The Substitutability Test

This section has presented an empirical means of examining the substitutability relationship between two cue phrases. In some cases, this seems to provide a better means of investigating the relationship than the test for substitutability; it will be interesting to see what results are produced in these cases.

However, the test for substitutability cannot be dispensed with in favour of this new method. There are still many cases where the test for substitutability provides clear information. More importantly, the experiment only works where the alternative cue phrases both lead to coherent texts—most of the time, this is not the case. For instance, it would be impossible to find a context neutral between the phrases *because* and *although*, where the relations marked by both phrases are inferrable from the same context.

7.2 Issues of Descriptive Adequacy: The Problems of ELABORATION and BACKGROUND

Two further problems with the substitution methodology concern the descriptive coverage of the set of relations it eventually produces. One will be discussed in this section, and the other in Section 7.3.

The first problem is simple—many coherent pairs of clauses or sentences can be found for which no cue phrase at all seems appropriate. Given that all the relations in the new set have their basis in the phrases that can signal them, this indicates that the new relations do not by themselves provide a descriptively adequate account of text.

A few examples can be given of contexts where no cue phrase can be used. Consider the following cases:

$$\begin{array}{r}\text{Dow Associates is one of}\\ \text{Britain's largest companies.}\end{array} \left\{\begin{array}{l} \#\ \textit{Indeed,} \\ \#\ \textit{Specifically,} \\ \#\ \textit{Furthermore,} \\ \#\ \textit{Incidentally,} \end{array}\right\} \begin{array}{l}\text{its head office is in}\\ \text{Kensington, where Dow}\\ \text{himself presides.}\end{array} \tag{7.10}$$

$$\begin{array}{r}\text{The University is playing}\\ \text{host to the 4th VM}\\ \text{conference on geochemistry}\\ \text{in August.}\end{array} \left\{\begin{array}{l} \#\ \textit{Indeed,} \\ \#\ \textit{Specifically,} \\ \#\ \textit{Furthermore,} \\ \#\ \textit{Incidentally,} \end{array}\right\} \begin{array}{l}\text{Geochemistry helps in the}\\ \text{search for minerals by}\\ \text{looking at the origins and}\\ \text{natural associations of}\\ \text{chemical elements and}\\ \text{compounds.}\end{array} \tag{7.11}$$

In these texts, it seems impossible to find an appropriate cue phrase to fit between the two sentences. In each case, the best approach is simply to present the sentences one after the other with no explicit cue. No doubt it is texts such as these which led Mann and Thompson to their claim that 'some types of rhetorical relations have *no* corresponding conjunctive signals'.

The lack of a prototypical cue phrase is characteristic of two RST relations in particular: ELABORATION and BACKGROUND. Text 7.10 illustrates the ELABORATION relation. The second sentence provides additional details about Dow Associates, and thus elaborates on the first. Text 7.11 illustrates BACKGROUND. In order properly to understand the first sentence in this case, it is necessary to know what geochemistry is, so this additional information is provided in the second sentence. While it is sometimes hard to distinguish between ELABORATION and BACKGROUND, it seems that neither of them is associated with any cue phrases at all.[1]

[1] Note that RST's ELABORATION can be signalled by relative clauses (Scott and de Souza (1990)):

7.2.1 Re-Assessing the Cue Phrase Hypothesis

The finding that there are some coherent contexts where no cue phrase is appropriate has important consequences for the working hypothesis in this thesis, that a connection can be made between the set of cue phrases and the set of coherence relations. We are forced to choose between the following two possibilities:

- If coherence relations are required to provide a full account of text coherence on their own, then the hypothesis that we can use the set of cue phrases as evidence for the set of relations is false.
- If we wish to maintain the hypothesis, then we have to introduce some other theoretical mechanism alongside relations in order to account for contexts where no cue phrase is appropriate.

The question we need to ask now, of course, is whether there are any principled reasons for proposing different mechanisms for explaining the coherence of texts such as those given above. In the remainder of this section, I will argue that there are such reasons, re-iterating that relations were only intended to model *some aspects* of the phenomenon of coherence, and suggesting that the coherence of texts such as 7.10 and 7.11 is better explained in terms of the metaphor of **focus**. I will argue that the lack of descriptive adequacy shown by the new set of relations may actually be an advantage: it makes for a less redundant account of the interaction of coherence relations and focusing phenomena in text.

7.2.2 The Concept of Focus Revisited

The concept of focus is also of central importance to a theory of coherence. To recap from Section 2.3: the notion of focus (eg Sidner (1983), Grosz, Joshi and Weinstein (1983), Brennan *et al* (1987), McCoy and Cheng (1991)) is used to model those entities in a text's domain of reference which are uppermost in a reader's mind as the text is being read. At any particular moment, certain entities are said to be 'in focus'; focus theories are concerned to chart the constraints on how the focused entities can change from one portion of text to another.

The concept of a **potential focus list** (Sidner (1983)) is commonly invoked in such theories. Every portion of a text is associated with a potential focus list, which specifies all the items to which the focus can legally shift in the next portion. Consider the following three texts:

(7.12) Dow Associates today launched a surprise bid for the beleaguered Harris group. It started buying shares as soon as the market opened, and continued buying all day.

Dow Associates, *whose head office is in Kensington,* is one of Britain's largest companies.

However, relative clauses and cue phrases are two very different kinds of syntactic resource. Spans linked by relative clauses are embedded one within another; those linked by cue phrases are presented consecutively. Furthermore, relative clauses can only operate at low levels of hierarchy within a text; whereas cue phrases such as *this is because* can link quite large text spans.

(7.13) Dow Associates today launched a surprise bid for the beleaguered Harris group. Harris reacted speedily, holding an upbeat press conference.

(7.14) Dow Associates today launched a surprise bid for the beleaguered Harris group. Beans, which contain lots of protein, are good for you.

In Text 7.12, the focused item in the first sentence is *Dow Associates*, and this focus is preserved in the second sentence. Text 7.13 shifts its focus from *Dow Associates* to *Harris*; since *Harris* is in the potential focus list associated with the first sentence, this text is also coherent. However, the potential focus list does not contain the item *beans*—so Text 7.14 is incoherent.

The concept of focus has proved useful primarily for modelling the pattern of anaphora in a text. For instance, it can explain how it is that the focused item in Text 7.12 can be pronominalised, but those in Texts 7.13 and 7.14 cannot be—essentially, an item can only be pronominalised if it is already in focus. Focus theories have been used successfully for resolving anaphora (Sidner, 1983) and for generating it (Dale, 1988, 1988).

7.2.3 Relations and Focus: Two Overlapping Metaphors

In the present context, the important point to note is that an account of focus seems to touch on phenomena which a theory of coherence relations might also be called on to explain. For instance, a theory of focus might equally well be used to explain the coherence of texts such as 7.10 and 7.11, which are presently explained respectively in terms of the ELABORATION and BACKGROUND relations. Conversely, the ELABORATION relation seems appropriate for analysing Texts 7.12 and 7.13, while no relation seems appropriate for the incoherent Text 7.14.

Much of the overlap between theories of relations and of focus can be traced to the relations of ELABORATION and BACKGROUND. The RST definitions for these relations make explicit reference to 'objects' being elaborated, and 'elements' for which further background is required; in this respect, they are unlike any other RST relations. It is thus clear how such relations might be re-described in terms of the focus metaphor.

In fact, the overlap between the focus metaphor and the ELABORATION relation is only partial. The definition for ELABORATION identifies several sub-types of relation, not all of which are elaborations of objects. The sub-relations WHOLE-PART, OBJECT-ATTRIBUTE and SET-MEMBER seem clearly to involve a notion of an entity being elaborated on. However, the sub-relations ABSTRACT-INSTANCE, PROCESS-STEP and GENERALISATION-SPECIFIC elaborate not on entities but on propositions. It should be noted that for these latter species of ELABORATION, cue phrases can readily be found: ABSTRACT-INSTANCE can be signalled by *for instance*; PROCESS-STEP can be signalled by *to do this* or *by*; and GENERALISATION-SPECIFIC can be signalled by *specifically* or *to be precise*.

A clear difference seems to be emerging between those relations which can be marked by cue phrases—which hold between propositions—and those which cannot be marked by cue phrases—which could just as well be thought of as focusing phenomena. The current cue-phrase-based conception of relations thus seems promising as the basis for a principled distinction between relations and focusing phenomena in text.

7.2.4 Why do we need both Relations and Focus?

The previous section has noted the potential for overlap between the metaphors of relations and focus, and consequently for redundancy in any theory in which they both figure. Is it not therefore possible to frame a theory of text solely in terms of one metaphor or the other? This section presents some arguments against such a proposal.

Problems with a Purely Relational Account

Several commentators (e.g. Hovy and McKoy (1989)) have noted that relations by themselves do not provide tight enough constraints on coherence. Consider the text in Figure 7.4. This text can be successfully analysed using relations, but is nonetheless

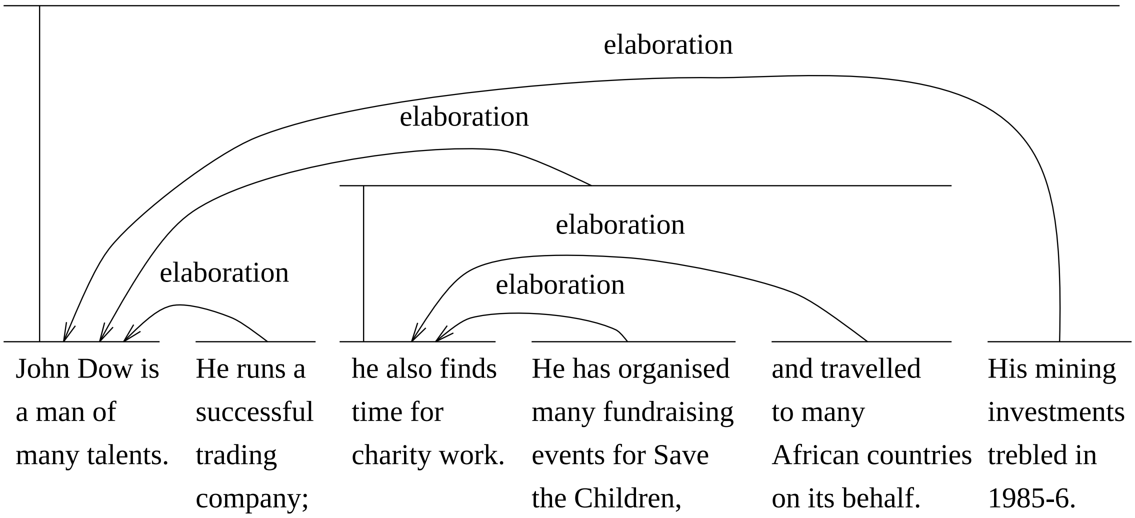

Figure 7.4: A Poorly Structured Text

poorly structured. The problem is with the last sentence, which returns to a topic which had apparently been closed. (It would be more appropriate to include the material from this sentence next to the material about Dow's trading company.) The sentence in its present position can nevertheless be described as an ELABORATION on the first clause—it gives additional information about Dow, which is the only requirement specified by the relation.

It thus seems that the ELABORATION relation is overly permissive. Note, however, that if we are only using relations to analyse texts, then leaving out ELABORATION results in a theory which is overly restrictive. For instance, how could we describe Text 7.15 in relational terms, except by saying that the second sentence elaborates on the first?

(7.15) Dow Associates is one of Britain's largest companies. Its head office is in Kensington.

Given that ELABORATION is the most commonly used relation in text analyses (cf. Mann (1984), p372), we must conclude that there are important aspects of text coherence which do not lend themselves well to description in terms of relations.

Problems with a purely focus-based account.

There are also problems with the focus metaphor when used by itself to model text coherence. Consider Text 7.16:

(7.16) Last year must have been hard for Dow. Oil prices fell by 20%.

This text is clearly coherent, but it seems to contain a sharp focus shift: from *Dow* and *last year* to *oil prices*. In order to account for shifts such as these, focus theorists are forced to abstract away from entities referred to explicitly in a text when creating the potential focus list, and add references to properties or features associated with these items. For instance, Grosz, Joshi and Weinstein allow that entities which are 'elements of the situation' described by an utterance are members of the potential focus list. This would permit them to include *the fall of oil prices* in the list associated with the first sentence of Text 7.16.

However, such abstractions seem to make the focus shifting mechanism itself too permissive: clearly, it will not always be the case that a text's focus can move from *last year* to *oil prices*. It seems more appropriate to explain coherence in a case like Text 7.16 in relational terms, for instance by identifying the second sentence as a JUSTIFICATION for the first. Coherence in this text is due to the rhetorical force of the second sentence, not to any general tendency to shift from one topic to another.

Summary

To sum up thus far: there are some aspects of coherence that seem best described in terms of rhetorical relations, and other aspects that seem best described in terms of focus. However, there is also a significant overlap between the phenomena described by the two metaphors. This leads to what we can term the **redundancy problem**: in a full account of coherence, it is likely that much information is expressed twice.

The redundancy problem is symptomatic of a second, more fundamental problem with relations and focus, which we can call the **grounding problem**. The problem with the two metaphors is that they are too expressive: theorists are free to use them to model as much or as little as they want. What is needed, therefore, are *empirically-grounded* conceptions of relations and focus which allow us to specify *in advance* what is to count as a relation or a focused entity.

7.2.5 Recent Attempts to Link Relations and Focus

Many discourse theorists have appreciated the need to integrate relational and focus-based approaches for a full account of coherence. Two strategies which have been suggested are discussed in this section.

Grosz and Sidner's (1986) theory presents one influential hypothesis—that the relational structure of a text has a role in *determining* shifts in focus. The DOMINANCE relations in a text, which model the hierarchy of the writer's goals, are used during its interpretation to determine pushes and pops of a **focus stack**. The concept of a 'potential focus list' does not feature in this model; its work is done by the intentional structure of the text. Abrupt changes of focus such as that illustrated in Text 7.16 therefore present no problems. However, as has already been noted in Section 2.4.3, the level of detail of this theory does present problems, in particular when it is used as the basis for automatic text processing applications.

Of the recent text planning systems to have incorporated both relational and focusing devices, the most notable are those of McKeown (1985), Hovy and McCoy (1989), and Hovy *et al* (1992). These are all characterised by their use of relations and focus as **multiple simultaneous constraints** on coherence. As a text is built, each new element is fitted into an appropriate rhetorical structure, and is also checked with adjacent elements to ensure that a legal focusing move is made. While this strategy certainly guarantees text conforming to both relational and focusing constraints, it is not guaranteed to be efficient: the large overlap between relations and focus means that, in all likelihood, many constraints are being checked twice.

7.2.6 A New Proposal about the Interaction of Relations and Focus

In order to overcome the redundancy and grounding problems, we need to develop independent conceptions of relations and focus, which model clearly separable aspects of discourse coherence, and which can be empirically motivated.

The new set of relations is promising as the basis for such a model. We have already seen how ELABORATION and BACKGROUND are responsible for much of the overlap between relations and focus, and how ELABORATION is overly permissive in its own right. The new set of relations, in which the exclusion of ELABORATION and BACKGROUND are motivated on independent grounds, seems a good starting point for a joint model of relations and focus.

As was mentioned in Section 2.3, a theory of relations goes hand in hand with a theory of **span structure**, which specifies whereabouts in a coherent text relations are expected to be found. The theory of span structure determines those points in a text where coherence is to be attributed to relations, and consequently, those parts of a text where coherence is to be accounted for by other devices. As was argued in Section 7.2.4, there are good reasons for adopting a theory of span structure which does not force relations to do all the explanatory work.

Based on the present conception of relations, a new suggestion about the interaction of relations and focus can be made. The principal claim is that text coherence where no cue phrase can be used is better explained with the focus metaphor than with the relational metaphor. If one clause elaborates on another, the focus is likely to remain the same, or to shift to some other item that has been explicitly mentioned in the text. A specific hypothesis can be advanced:

H1 Where there is no coherence relation between two text units—i.e. where no cue phrase appears and none can be inserted—coherence is ensured by **focusing constraints**.

It should now be possible to give the term **focusing constraints** a much more concrete interpretation; one that can be grounded in surface textual phenomena. (A simplistic hypothesis could be, for instance, that when no cue phrase is appropriate between two spans, the entity in focus in the second span must appear as a lexical item in the first span. Of course this hypothesis is likely to need refinement, for example to account for phenomena such as bridging references.) We can also propose a hypothesis about how coherence is preserved between text units which are linked by relations:

H2 Focusing constraints are *not* needed to ensure coherence between text units linked by relations.

The main new idea in these hypotheses is that relations and focusing constraints account not just for different aspects of a text, but *different portions* of it. To ensure a text is coherent, it is not necessary that there be a relation between every pair of adjacent text units; but where there is no relation, separate constraints on focus must be met. Conversely, where there is a relation, focusing constraints are not needed to ensure coherence. This idea differs from existing models of coherence, in which relational and focusing constraints apply simultaneously at every point in a text. By using relations and focus to model different portions of text, the new model presents an appealing solution to the redundancy problem. Moreover, since it permits a more concrete conception of focusing constraints, it addresses the grounding problem as well.

There thus seem to be good reasons for proposing an account where relations and focus interact in the way proposed above. However, it should be borne in mind that these are only suggestions, and that much further research is required to explore them in detail. Meanwhile, we should reiterate that the hypothesis proposed in this thesis, of a connection between cue phrases and coherence relations, is contingent on these suggestions proving fruitful.

7.3 Relations at Different Levels of Hierarchy

A final potential problem with the substitution methodology concerns the issue of relations between large sections of text. The idea that relations can apply between text units of any size was noted as one of their attractive features. In many theories of relations, structures as large as entire paragraphs are linked together by relations. The hierarchical, recursive analyses which result from this conception of relations make it attractive from a computational point of view, and also from the standpoint of theoretical parsimony.

It may seem that the decision to associate relations with cue phrases threatens to tie them to text spans of a particular size. For instance, the cue phrase *because* is best suited for joining two clauses within a single sentence:

(7.17) Mary was in a good mood, because she had passed her exam.

While it is possible to use *because* in more complex sentences, it is often inappropriate to use the phrase to link units larger than clauses. Consider Text 7.18:

(7.18) Mary was in a good mood. Because she had passed her exam.

While some such texts may border on acceptability, *because* is not commonly used to link whole sentences, and would probably lead to incoherence if used to link units of several sentences.

Other cue phrases such as *nevertheless* are better able to link whole sentences. However, the problem with such phrases is essentially the same: there is always a limit to the size of spans which they are suitable for joining. For instance, we would not expect to find the word *nevertheless* linking whole chapters of a book, or even sections of an article. How can the current theory of relations deal with relations between large units of this kind?

An answer is suggested here which draws on the fact that cue phrases can contain anaphoric elements.

7.3.1 Cue Phrases and Propositional Anaphora

The test for cue phrases as outlined in Section 4.2 permits phrases which contain propositional anaphora, such as *this is because*, or *following this*. This decision is justified in two ways. Firstly, though the test for cue phrases calls for the replacement of all anaphoric elements by their antecedents, in the case of these phrases the replacement renders the outcome of the test a foregone conclusion—when the antecedent is used, the two clauses originally related are effectively re-expressed in a single clause, which can therefore stand on its own. For instance, consider what happens when the test is applied to the following isolated clause:

(7.19) Because of *this*, Mary was in a good mood.

Replacing *this* by a non-anaphoric NP would result in something of the following form:

(7.20) Because of *the fact that she had passed her exam*, Mary was in a good mood.

Text 7.20, unlike Text 7.19, does not need additional context to be interpreted.

The second reason for allowing anaphora in cue phrases is that many simple connectives have evolved from more complex expressions containing anaphora. As has been noted by Halliday and Hasan (1976), the etymology of words like *therefore* and *thereby* shows that they derive from phrases containing anaphora.

If simple propositional anaphora such as *this* and *that* are used, the resulting cue phrases are still inappropriate for signalling very large relations; readers tend not to look for antecedents beyond the previous sentence. However, if more specific anaphoric expressions are used, antecedents can be further away. Consider the following text:

(7.21) Although the developers had scheduled construction to begin in January, an important archaeological find was made while the foundations were being dug. A dig has quickly been organised to make the most of the find; heading the investigation is Professor W Percus, an authority on Roman Britain.

Because of this discovery, the building has been delayed several months. It is now unlikely to be finished on time...

In this example, the cue phrase *because of this discovery* can be seen as signalling a relation between two whole paragraphs. It should be clear how such expressions could be used to link even larger sections of text.

7.3.2 Different Relations at Different Levels?

The substitutability test deliberately ignores the issue of span size when comparing two phrases. It is possible, for instance, that two phrases are classed as synonymous even though they are suitable for linking spans of different sizes. So it remains an open question whether the set of relations used at low levels of hierarchy is the same as that used for higher levels. For the most part, the examples in this thesis have been of relations between single sentences or single clauses; whether the same set of relations emerges with larger texts is a matter for further investigation.

Nevertheless, some observations can already be made. As was noted in Section 4.3.2, many cue phrases can be systematically modified to include anaphora—for instance, *in contrast* becomes *in contrast to this*; *instead* becomes *instead of this*; *as a result* becomes *as a result of this*.

An interesting link can also be noted between the devices used to signal relations at high levels and those which perform the same function within single clauses. Consider the following texts:

(7.22) Because of this problem, the experiment failed.

(7.23) Because of a small leak in the hydraulic system, the experiment failed.

In Text 7.22, *this problem* refers back to some previously described proposition, and a large relation appears to be signalled. However, in Text 7.23, the problem is being described for the first time—in effect, a causal relation is being signalled inside a single clause. Note that the only difference between the two clauses is whether or not the noun phrase is interpreted as anaphoric.

7.3.3 Relations, Focus and Nominalisation

The use of nominalisation to refer to previous propositions places texts such as 7.22 within the scope of a theory of focus. Are we to interpret this clause as one span of a high-level CAUSE relation, or should we consider it simply as an additional statement about a topic introduced earlier? The fact that an anaphoric expression can be produced, and correctly interpreted, might be taken as support for the focus-based account. However, propositional anaphora are rather different from nominal or verbal anaphora. They introduce a meta-level of description: the propositions in which they appear are not about objects in the domain of reference, but about propositions about these objects.

Nonetheless, it is implausible that all predications about propositions should be interpreted as 'indicators of text structure'. Predications do not have to provide the sort of information given by relations. Consider the following text:

(7.24) Due to a leak in the hydraulic system, the experiment failed. The problem was spotted by a junior lab assistant.

We would not want to suggest that this is an instance of a relation called SPOTTED-BY.

In many cases, therefore, it remains unclear how a joint theory of relations and focus would handle texts containing nominalisations. Perhaps it would be best to leave a decision about such cases until a detailed theory of the interaction of relations and focus has been established for more concrete texts. However, if this theory did not lead to a clear distinction between relational and focusing phenomena in these more abstract cases, this would certainly tell against it.

7.4 Summary

This chapter has discussed in some detail three potential problems with the substitution methodology. Firstly, a limitation with the substitutability test was noted—in some cases it appears to draw on post-theoretical intuitions. For these cases, an alternative method for investigating substitutability relationships was suggested, based on an empirical study of subjects' response times. Secondly, contexts were presented where no cue phrase is appropriate, and hence in the current model no relation is present. It was suggested that these contexts are better explained in terms of the focus metaphor, and the outlines of a joint theory of relations and focus were sketched. Finally, the question of relations between large sections of text was raised. It was partially answered by noting the existence of cue phrases containing propositional anaphora, which are able to link quite large sections of text. The possibility of focusing mechanisms applying at these high levels was also discussed.

Clearly, all of these issues call for further investigation. However, none of the objections raised should yet be seen as fatal to the theory being proposed in this thesis. On the contrary, they all suggest interesting new avenues of research.

Chapter 8

Conclusions

8.1 A Summary of the Thesis

This thesis addresses an important problem for theories of discourse coherence: the lack of a standard, well-motivated set of discourse relations. While the general idea of coherence relations has proved extremely useful in many theories and many text processing applications, no solid conception has emerged about what it is that relations actually represent, or how to go about defining them.

The solution proposed in this thesis rests on a conception of relations as modelling psychological constructs operative in writers and readers when they are processing text. This conception is fleshed out with reference to Rosch's notion of the **basic level** of categorisation, and to the notion of **action schemata** as used in theories of skilled task performance.

It is argued that evidence for these constructs can be sought in a study of the connective **cue phrases** found in a given language. If relations are actually communicated between the writer of a text and its reader, then it makes sense for there to be resources in the language for signalling relations explicitly. While relations can sometimes be inferred by the reader from context and background knowledge, there is no reason to suppose that any relation exists which will *always* be thus inferrable.

Based on these arguments, a methodology is proposed for justifying a set of relations. The methodology centres around two simple linguistic tests. The first is a **test for cue phrases**, which is used to gather a corpus of some 200 cue phrases from a sample of naturally-occurring text. The second is a **test for substitutability**, which is used to organise the corpus of cue phrases into a hierarchical taxonomy. This taxonomy is created without buying into any particular theory of discourse, but it provides an extremely rich source of information from which such a theory can be motivated.

The taxonomy lends itself to a conception of relations as **feature-based constructs**. The substitutability relationships in the taxonomy each have a natural feature-theoretic interpretation, which means that extracts from the taxonomy can be used separately to motivate the individual features which are the components of relations. Some guidelines for creating feature definitions are laid down, and a number of specific definitions are then motivated one by one. The features eventually motivated prove useful in analysing a wide range of cue phrases.

The set of relation definitions is conceived as an overlay on the set of cue phrase definitions. There remains some work to be done in deciding exactly how to map one set onto the other; but in any case, most of the 'hard work' will have been done once the cue phrase definitions are completed. At this point, all of the important dimensions of variation for relations will have been established, and it will just be a question of deciding whether or not *each* cue phrase is in one-to-one correspondence with a relation. Some guidelines for answering this final question have been set out.

8.2 The Contributions of the Thesis

The thesis delivers on two levels. Its most tangible contribution is the table of cue phrase definitions set out in Appendix D. These definitions can be taken to form the basis for a new set of relations to be implemented in text processing applications, or used by discourse theorists as the basis for further research. The new definitions differ significantly from existing sets, and contain some novel ideas. While several studies suggest that relations are composite constructs, this is the first to propose a set of *independent* features to analyse them, in which the order of the decisions about feature values is not constrained. Moreover, the feature definitions are expressed in relatively simple terms—representing alternative feature values is often just a matter of choosing different bindings between variables—and they should eventually lend themselves well to implementation. Some of the feature definitions are innovative extensions of existing work; in particular the new SEMANTIC/PRAGMATIC distinction and the new distinction between POSITIVE and NEGATIVE POLARITY. Finally, the use of defeasible CAUSAL and INDUCTIVE rules in the definitions is promising as part of an account of how relations are grounded in more general principles of cognitive economy. The new definitions are thus in tune with the suggestion in Hobbs' opening quote, that coherence relations should be thought of as 'instantiations in discourse comprehension of more general principles of coherence that we apply in attempting to make sense of the world we find ourselves in'.

However, the main purpose of the thesis is not to put forward a completely worked-out set of relation definitions. It is rather to consider the question of how a set of relations should be justified in the first place. The primary contribution of the thesis is the methodology proposed for motivating a set of relation definitions; the definitions themselves are intended first and foremost as an implementation of this methodology.

As such, they can certainly be contested. The corpus of cue phrases can likely be enlarged with new phrases. Exceptions might be found to some of the relationships in the current taxonomy. The relationships in the taxonomy might be better modelled by altering the set of features. Such changes, made within the framework of the tests for cue phrases and for substitutability, do not damage the substance of the thesis. Indeed, the tests provide a useful forum for discussion about the set of relations. If the substitution methodology is accepted, then disagreements about relation definitions become disagreements about the interpretation of concrete linguistic examples, rather than about intangible first principles.

8.3 The Substitution Methodology: A Balanced Verdict

The thesis therefore stands or falls on an assessment of the substitution methodology. Such an assessment can be made from several different standpoints, which will be considered below in turn.

8.3.1 The Arguments for the Methodology

Firstly, the arguments on which the methodology rests can be evaluated. The main argument is that the cue phrases in a language will tell us about the mechanisms used by readers and writers of that language. As noted in Section 3.6.1, it is unusual to propose an investigation of psychological mechanisms by any means other than psychoplogical experiments. However, experiments which investigate psychologically real relations must make hypotheses about the behavioural indicators of relations, and these hypotheses are themselves contentious. It is notoriously difficult to interpret the behaviour of, say, a writer, in terms of a theory of writing. Faced with this difficulty, it is surely permissible to look at the *medium* in which writing and reading occur—namely language—to find out something about these tasks. Language is not adopted arbitrarily by those who read and write; its evolution is inseparable from that of the tasks themselves. It would be surprising if it did not contain valuable information about how the tasks are performed.

This is not to say that the substitution methodology obviates the need for empirical experiments on readers and writers. On the contrary, the two approaches complement each other. The arguments used to support the substitution methodology also act to legitimise the use of cue phrases as an experimental window on relations. And theorising about the substitutability relationships in the taxonomy can be thought of as a systematic way of generating experimental hypotheses to test.

8.3.2 The Practicability of the Methodology

The practicability of the substitution methodology should also be considered. Are the tests easy to use? Do different testers come to the same conclusions? The approach in this thesis has been to assume that they will—in keeping with much work in linguistics, the examples of the operation of the tests are presented on the assumption that readers will agree with them. However, it would be a good idea to put this assumption to the test in an experiment comparing the decisions of several testers in particular cases. In the case of the substitutability test, a fairly high degree of consensus is predicted. Nevertheless, in cases where there is disagreement, we can resort to the kind of experiment described in Section 7.1.3, which examines the influence of alternative cue phrases on the time taken to answer different questions about a text.

8.3.3 The Results of the Methodology

Finally, the substitution methodology can be assessed by looking at the set of relations which it serves to justify.

A key result of the methodology is the motivation of a feature-theoretic approach to relation definitions. Most previous classifications of relations have taken one particular dimension of variation to be dominant—the 'top level' of classification varies from theory to theory, as was shown in Chapter 2, but successive classifications do not seem to improve on one another.

The taxonomy of cue phrases prompts a different categorisation of relations. Several dimensions of variation are identified, but no single dimension is taken to be dominant. A conception of relations as unstructured collections of features falls quite naturally out of the taxonomy of cue phrases, where nearly all of the variation between phrases is found at the lowest levels. This feature-based classification promises to provide a much better fit for the data. What is more, the only other study starting from a conception of relations as cognitive constructs (that of Sanders *et al*) also suggests their decomposition into orthogonal features, and does so on independent grounds.

The job of deciding on a set of features which make sense for *all* the various kinds of phrase in the taxonomy is, of course, a very difficult one. Some interesting generalisations have been noted in this thesis, but much work still remains to be done before a satisfactory set of features is produced. However, we have tried to avoid the temptation of concentrating on small subsets of phrases, and analysing these in isolation. While useful results can be obtained in this way, it is hoped that the findings in this thesis testify to the benefits of seeking a theory with much broader coverage.

The set of relations must also eventually be assessed in terms of its descriptive adequacy in describing coherent text. It is perhaps here that the main question mark about the methodology lies. As was noted in Section 7.2, the descriptive adequacy of the set of relations cannot be determined until an accompanying theory of focus is provided. It was argued that a theory of relations should not be saddled with all the explanatory work in a theory of discourse coherence; a theory of focus should also be involved to account for pairs of text segments whose relationship cannot be made explicit by any cue phrase. It remains to be seen whether such a theory can be devised, and how well the theories of relations and focus would interact. Clearly, this is a matter for further research.

8.4 Towards a Complete Account of Discourse Coherence

The present theory of relations is thus only a partial theory of discourse coherence. However, it provides a solid foundation on which a complete account can be based. Specific hypotheses have been proposed about the segments of a discourse which will be explained by relations—those which can be linked by cue phrases—and hence, about the parts of a discourse which some other metaphor will have to explain.

The emphasis has been on producing a compact theory whose constructs are systematically justifiable. Every relation is justified in precisely the same way: it is composed of a number of features, each of which is motivated in the same way from the substitutability relationships in the taxonomy of cue phrases. These relationships are in turn based on the pre-theoretical judgements of ordinary readers and writers.

The analysis of texts in terms of the theory is similarly grounded in pre-theoretical decisions. The decision about whether a relation applies between two spans of text is based on the decision about whether a particular cue phrase is appropriate to link them; it does not require the intuitions of a discourse analyst. Once the decision has been made, however, the complex definition of the relation provides a wealth of theoretical information about the discourse at that point, representing the intentions which underlie it, and the states in the world and the reader's world model which are requisite for their accomplishment.

The theory of relations is based on an analysis of a well-delineated class of textual phenomena. By examining how these phenomena are used by ordinary readers and writers, an extremely informative account of text is produced. It remains to be seen whether a theory of focus can be developed to similar standards.

Bibliography

Allen, J. (1995). *Natural Language Understanding.* Benjamin/Cummings, Redwood City, CA.

Austin, J. L. (1962). *How to Do Things with Words.* Oxford: Clarendon Press.

Ballard, D., Conrad, R., and Longacre, R. (1971). The deep and surface grammar of interclausal relations. *Foundations of Language*, **4**, 70–118.

Bateman, J. and Rondhuis, K. (1994). Coherence relations: Analysis and specification. Technical Report R1.1.2: a,b, The DANDELION Consortium, Darmstadt, Germany.

Bateman, J., Degand, L., and Teich, E. (1993). Towards multilingual textuality: Some experiences from multilingual text generation. In *Proceedings of the Fourth European Workshop on Natural Language Generation*, pages 5–18, Pisa, Italy.

Beaugrande, R. d. (1984). *Text Production: Toward a Science of Composition.* Ablex Publishing Corporation, Norwood, NJ.

Beckwith, R., Fellbaum, C., Gross, D., and Miller, G. (1990). WordNet: A lexical database organised on psycholinguistic principles. Technical Report CSL-42, Cognitive Science Laboratory, Princeton University.

Bereiter, C. and Scardamalia, M. (1987). *The Psychology of Written Composition.* Lawrence Erlbaum Associates, Hillsdale, NJ.

Berry, D. and Broadbent, D. (1984). On the relationship between task performance and verbalizable knowledge. *Quarterly Journal of Experimental Psychology*, **36A**, 209–231.

Boguraev, B. and Pustejovsky, J. (1990). Lexical ambiguity and the role of knowledge representation in lexicon design. *Proceedings of 14th International Conference on Computational Linguistics (*COLING90*)*, pages 36–42.

Brennan, S. E., Friedman, M. W., and Pollard, C. (1987). A centering approach to pronouns. In *Proceedings of the 25th ACL Conference*, pages 155–162, Stanford.

Briscoe, E., Copestake, A., and Boguraev, B. (1990). Enjoy the paper: Lexical semantics via lexicology. *Proceedings of 14th International Conference on Computational Linguistics (*COLING90*)*, pages 42–47.

Butterworth, B. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research (4)*, pages 75–87.

Carpenter, B. and Penn, G. (1994). ALE: The attribute logic engine user's guide, version 2.0.1. Technical Manual, Carnegie-Mellon University.

Chomsky, N. (1964). *Current Issues in Linguistic Theory*. Mouton, The Hague.

Cohen, P. and Levesque, H. (1990). Rational interaction as the basis for communication. In P. Cohen, J. Morgan, and M. Pollack, editors, *Intentions in Communication*, pages 221–256. MIT Press, Cambridge, MA.

Cohen, P. and Perrault, C. (1979). Elements of a plan-based theory of speech acts. *Cognitive Science*, **3**, 177–212.

Cohen, R. (1984). A computational theory of the function of clue words in argument understanding. In *Proceedings of the 10th International Conference on Computational Linguistics*, pages 251–258.

Dale, R. (1988a). *Generating Referring Expressions in a Domain of Objects and Processes*. Ph.D. thesis, Centre for Cognitive Science, University of Edinburgh.

Dale, R. (1988b). The generation of subsequent referring expressions in structured discourses. In M. Zock and S. G, editors, *Advances in Natural Language Generation*, pages 58–75. Pinter Publishers Ltd, London.

Dale, R. and Reiter, E. (1992). A fast algorithm for the generation of referring expressions. *Proceedings of COLING*.

Dale, R., Mellish, C., and Zock, M. (1990). *Current Research in Natural Language Generation*. Academic Press, London.

Dale, R., Hovy, E., Rösner, D., and Stock, O. (1992). *Aspects of Automated Natural Language Generation*. Springer-Verlag, Berlin.

Delin, J. and Oberlander, J. (1992). Aspect-switching and subordination: the role of *it*-clefts in discourse. *Proceedings of 14th International Conference on Computational Linguistics (*COLING92*)*, pages 281–287.

Delin, J., Hartley, A., Paris, C., Scott, D., and Vander Linden, K. (1994). Expressing procedural relationships in multilingual instructions. In *Proceedings of the 7th International Workshop on Natural Language Generation*, pages 61–70, Kennebunkport, Maine.

Dornan, E. and Dawe, C. (1984). *The Brief English Handbook*. Little, Brown and Company, Boston.

Elhadad, M. and McKeown, K. R. (1990). Generating connectives. In *COLING-90*, pages 97–101.

Evans, R. and Gazdar, G. (1989). Inference in DATR. In *Proceedings of the 4th Conference of the European Chapter of the ACL*, pages 66–71, UMIST, Manchester, UK.

Flower, L. and Hayes, J. (1980). The dynamics of composing: Making plans and juggling constraints. In L. Gregg and E. Steinberg, editors, *Cognitive Processes in Writing*, pages 31–50. Lawrence Erlbaum Associates, Hillsdale, NJ.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics*, volume 3. Academic Press, New York.

Grimes, J. (1975). *The Thread of Discourse*. The Hague: Mouton.

Grosz, B. J. and Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, pages 175–203.

Grosz, B. J., Joshi, A. K., and Weinstein, S. (1983). Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st ACL Conference*, pages 44–50, Cambridge, MA.

Grote, B., Lenke, N., and Stede, M. (1995). Ma(r)king concessions in English and German. In *Proceedings of the Fifth European Workshop on Natural Language Generation*, Leiden, the Netherlands.

Haberlandt, K. (1982). Reader expectations in text comprehension. In J. F. Le Ny and W. Kintsch, editors, *Language and Language Comprehension*, pages 239–249. North-Holland, Amsterdam.

Halliday, M. and Hasan, R. (1976). *Cohesion in English*. Longman.

Hayes, J. and Flower, L. (1980). Identifying the organisation of the writing process. In L. Gregg and E. Steinberg, editors, *Cognitive Processes in Writing*, pages 3–30. Lawrence Erlbaum Associates, Hillsdale, NJ.

Hirschberg, J. and Litman, D. (1993). Empirical studies on the disambiguation of cue phrases. *Computational Linguistics*, **19**, 501–530.

Hobbs, J., Stickel, M., Appelt, D., and Martin, P. (1993). Interpretation as abduction. *Artificial Intelligence*, **63**.

Hobbs, J. A. (1985). On the coherence and structure of discourse. Technical Report CSLI-85-37, Center for the Study of Language and Information, Stanford University.

Hovy, E. (1993). Automated discourse generation using discourse structure relations. *Artificial Intelligence*, **63**, 341–385.

Hovy, E. H. (1988). Planning coherent multisentential text. In *Proceedings of the 26th ACL Conference*, pages 163–169, Buffalo.

Hovy, E. H. (1990). Parsimonious and profligate approaches to the question of discourse structure relations. In *Proceedings of the 5th International Workshop on Natural Language Generation*, Pittsburgh.

Hovy, E. H. and McCoy, K. (1989). Focusing your RST: A step toward generating coherent multisentential text. In *Proceedings of the 11th Cognitive Science Conference*, pages 667–674, Ann Arbor.

Hovy, E. H., Lavid, J., Maier, E., Mittal, V., and Paris, C. (1992). Employing knowledge resources in a new text planner architecture. In *Proceedings of the Sixth International Workshop on Natural Language Generation*, pages 57–72, Trento, Italy.

Karttunen, L. (1973). Presuppositions of compound sentences. *Linguistic Inquiry*, **4**, 169–193.

Keenan, E. (1971). Two kinds of presupposition in natural language. In C. Fillmore and D. Langendoen, editors, *Studies in Linguistic Semantics*. Holt, New York.

Knott, A. (1991). *New Strategies and Constraints in RST-Based Text Planning*. Master's thesis, Department of Artificial Intelligence, University of Edinburgh.

Knott, A. and Sanders, T. (1996). The classification of coherence relations and their linguistic markers: An exploration of two languages. In preparation.

Lascarides, A. and Asher, N. (1991). Discourse relations and defeasible knowledge. In *Proceedings of the 29th Conference of the Association for Computational Linguistics*, pages 55–63, Berkeley, CA.

Lascarides, A. and Asher, N. (1993). Temporal interpretation, discourse relations and common sense entailment. To appear in *Linguistics and Philosophy*.

Lascarides, A. and Oberlander, J. (1992). Abducing temporal discourse. In *Aspects of Automated Natural Language Generation*, pages 167–182. Springer-Verlag.

Lascarides, A. and Oberlander, J. (1993). Temporal connectives in a discourse context. In *Proceedings of the 6th Conference of the European Chapter of the Association for Computational Linguistics*, pages 260–268, Utrecht, The Netherlands.

Lascarides, A., Asher, N., and Oberlander, J. (1992). Inferring discourse relations in context. In *Proceedings of the 30th Conference of the Association for Computational Linguistics*, pages 1–8, Delaware.

Litman, D. and Allen, J. (1990). Discourse processing and commonsense plans. In P. Cohen, J. Morgan, and M. Pollack, editors, *Intentions in Communication*. MIT Press.

Longacre, R. (1983). *The Grammar of Discourse: Notional and Surface Structures*. Plenum Press, New York.

Maier, E. (1993). The extension of a text planner for the treatment of multiple links between text units. In *Proceedings of the Fourth European Workshop on Natural Language Generation*, pages 103–114, Pisa, Italy.

Maier, E. and Hovy, E. H. (1991). A metafunctionally motivated taxonomy for discourse structure relations. In *Proceedings of the 3rd European Workshop on Natural Language Generation*, pages 38–45, Judenstein, Austria.

Mann, W. C. (1984). Discourse structures for text generation. In *Proceedings of the 10th International Conference on Computational Linguistics and the 22nd Annual Meeting of the ACL*, pages 367–375, Stanford.

Mann, W. C. and Matthiessen, C. (1983). Nigel: a systemic grammar for text generation. Technical Report RR/83/105, USC Information Sciences Institute, Marina del Rey, CA.

Mann, W. C. and Thompson, S. A. (1988). Rhetorical structure theory: A theory of text organisation. *Text*, **8(3)**, 243–281. Also available as Tech Report RR-87-190, USC Information Sciences Institute, Marina del Rey, CA.

Mann, W. C., Moore, M. A., Levin, J. A., and Carlisle, J. H. (1975). Observation methods for human dialogue. Technical Report RR/75/33, Information Sciences Institute, Marina del Rey, CA.

Mann, W. C., Matthiessen, C. M., and Thompson, S. A. (1989). Rhetorical structure theory and text analysis. Technical Report RR/89/242, Information Sciences Institute, Marina del Rey, CA.

Martin, J. (1992). *English Text: System and Structure*. Benjamin, Amsterdam.

McCoy, K. F. and Cheng, J. (1991). Focus of attention: Constraining what can be said next. In C. L. Paris, W. R. Swartout, and W. C. Mann, editors, *Natural Language Generation in Artificial Intelligence and Linguistics*. Kluwer.

McDonald, D., Meteer, M., Fawcett, R., Hovy, E., and de Smedt, K. (1994). *Proceedings of the Seventh International Workshop on Natural Language Generation*. ACL, Maine.

McKeown, K., Moore, J., and Nirenburg, S. (1990). *Proceedings of the Fifth International Workshop on Natural Language Generation*. AAAI, Dawson, Pennssylvania.

McKeown, K. R. (1985). *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Cambridge University Press.

Meyer, B. and Freedle, R. (1984). Effects of discourse type on recall. *American Educational Research Journal*, **21**, 121–143.

Miller, G., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. (1990). Five papers on WordNet. Technical Report CSL-43, Cognitive Science Laboratory, Princeton University.

Moens, M. and Steedman, M. (1988). Temporal ontology and temporal reference. *Computational Linguistics*, **14**(2), 15–28.

Moore, J. and Pollack, M. (1992). A problem for RST: The need for multi-level discourse analysis. *Computational Linguistics*, **18**, 537–544.

Moore, J. D. (1989). *A Reactive Approach to Explanation in Expert and Advice-Giving Systems*. Ph.D. thesis, UCLA.

Moore, J. D. and Paris, C. L. (1989). Planning text for advisory dialogues. In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, pages 203–211.

Moore, J. D. and Paris, C. L. (1993). Planning text for advisory dialogues: Capturing intentional and rhetorical information. *Computational Linguistics*, **19**, 651–694.

Moser, M. and Moore, J. (1995). Using discourse analysis and automatic text generation to study discourse cue usage. In *Proceedings of the AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, pages 92–98, Stanford, CA.

Oberlander, J. and Lascarides, A. (1991). Discourse generation, temporal constraints, and defeasible reasoning. In *Proceedings of the AAAI Fall Symposium on Discourse Structure in Natural Language Understanding and Generation*, Asilomar, CA.

Oversteegen, E. (1995). Causal and contrastive connectives. Technical Report R1.1.3b, The DANDELION Consortium, CLS, Tilburg/Nijmegen.

Quirk, R., Greenbaum, S., Leech, G., and Svartik, J. (1972). *A grammar of contemporary English*. Longmans, London.

Rambow, O., editor (1993). *Intentionality and Structure in Discourse Relations: Proceedings of the ACL SIGGEN Workshop*, Columbus, Ohio.

Reason, J. (1979). Actions not as planned. In G. Underwood and R. Stevens, editors, *Aspects of Consciousness*. Academic Press, London.

Redeker, G. (1990). Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics*, **14**, 367–381.

Rosch, E. (1978). Principles of categorisation. In E. Rosch and B. Lloyd, editors, *Cognition and Categorisation*. Lawrence Erlbaum Associates, Hillsdale, N.J.

Rosch, E., Mervis, C., Gray, W., Johnson, D., and Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, **8**, 382–439.

Rösner, D. and Stede, M. (1992a). Customising RST for the automatic production of technical manuals. In *Aspects of Automated Natural Language Generation*, pages 199–214. Springer-Verlag.

Rösner, D. and Stede, M. (1992b). TECHDOC: A system for the automatic production of multilingual technical documents. In *Proceedings of KONVENS '92*. Springer-Verlag.

Rumelhart, D. and Norman, D. (1982). Simulating a skilled typist: A study of skilled perceptual motor performance. *Cognitive Science*, **6**, 1–36.

Sanders, T. J. M., Spooren, W. P. M., and Noordman, L. G. M. (1992). Towards a taxonomy of coherence relations. *Discourse Processes*, **15**, 1–35.

Sanders, T. J. M., Spooren, W. P. M., and Noordman, L. G. M. (1993). Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics*, **4–2**, 93–133.

Schiffrin, D. (1987). *Discourse Markers*. Cambridge University Press.

Scott, D. and Hovy, E., editors (1993). *Burning Issues in Discourse: Proceedings of the NATO Advanced Research Workshop*, Maratea, Italy.

Scott, D. and Paris, C. (1995). Identifying the mapping of semantics onto language: Going beyond the text. In *Proceedings of the AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, pages 139–142, Stanford, CA.

Scott, D. R. and de Souza, C. S. (1990). Getting the message across in RST-based text generation. In R. Dale, C. Mellish, and M. Zock, editors, *Current Research in Natural Language Generation*. Academic Press.

Searle, J. R. (1969). *Speech Acts*. Cambridge: Cambridge University Press.

Segal, E. M., Duchan, J. F., and Scott, P. J. (1991). The role of interclausal connectives in narrative structuring: Evidence from adults' interpretation of simple stories. *Discourse Processes*, **14**, 27–54.

Seligman, M. (1994). Discovery and format of input structures for tactical generation. In *Proceedings of the 7th International Workshop on Natural Language Generation*, pages 5–12, Kennebunkport, Maine.

Sidner, C. (1983a). Focusing in the comprehension of definite anaphora. In M. Bady and R. Berwick, editors, *Computational Models of Discourse*, pages 267–330. MIT Press, Cambridge, Mass.

Sidner, C. L. (1983b). What the speaker means: The recognition of speakers' plans in discourse. *International Journal of Computers and Mathematics, Special Issue in Computational Linguistics 9 (1)*, pages 71–82.

Spooren, W. (1989). *Some Aspects of the Form and Interpretation of Global Contrastive Coherence Relations*. Ph.D. thesis, Catholic University of Nijmegen, the Netherlands.

Sweetser, E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge University Press.

van Dijk, T. A. (1979). Pragmatic connectives. *Journal of Pragmatics*, **3**, 447–456.

Vander Linden, K., Cumming, S., and Martin, J. (1992). Using system networks to build rhetorical structures. In *Aspects of Automated Natural Language Generation*, pages 183–198. Springer-Verlag.

von Klopp, A. (1993). *Negation: Implications for Theories of Natural Language*. Ph.D. thesis, Centre for Cognitive Science, University of Edinburgh.

Zock, M., Adorni, G., and Ferrari, G. (1994). *Proceedings of the Fourth European Workshop on Natural Language Generation*. Springer-Verlag, Berlin.

Zuckerman, I. and Pearl, J. (1986). Comprehension-driven generation of meta-technical utterances in math tutoring. In *Proceedings of the 5th National Conference of the AAAI*, pages 606–611, Philadelphia, PA.

Appendix A

The Corpus of Cue Phrases

This appendix contains a list of the English cue phrases on which the classification of relations is to be based. Each phrase in the corpus passes the test for cue phrases set out in Section 4.2. The majority of the phrases were gathered systematically in the analysis of 'academic' discourse reported in Section 4.3; however, for the sake of completeness, several phrases which were discovered after the analysis have also been included (and identified as such).

The cue phrases are listed in alphabetical order in the table below; there are around 350 in all. The three columns in the table contain respectively a cue phrase, its syntactic category (though this is sometimes tricky to determine), and whether the phrase was found in the initial corpus analysis.

Phrase	Syntactic category	In corpus analysis
above all	Conj-adverb	Yes
accordingly	Conj-adverb	Yes
actually	Conj-adverb	Yes
admittedly	Conj-adverb	Yes
after	Conj-adverb	Yes
after	Subordinator	Yes
after all	Conj-adverb	Yes
after that	Conj-adverb	Yes
afterwards	Conj-adverb	Yes
again	Conj-adverb	Yes
all in all	Conj-adverb	Yes
all the same	Conj-adverb	Yes
also	Conj-adverb	Yes
alternatively	Conj-adverb	Yes
although	Subordinator	Yes
always assuming that	Subordinator	Yes
and	Coordinator	Yes

Phrase	Syntactic category	In corpus analysis
and/or	Coordinator	Yes
another time	Conj-adverb	No
anyway	Conj-adverb	Yes
apart from that	Prep Phrase	No
as	Subordinator	Yes
as a consequence	Conj-adverb	Yes
as a corollary	Conj-adverb	Yes
as a result	Conj-adverb	Yes
as it happened	Conj-adverb	No
as it is	Conj-adverb	No
as it turned out	Conj-adverb	No
as long as	Subordinator	Yes
as luck would have it	Conj-adverb	No
as soon as	Subordinator	Yes
as well	Conj-adverb	Yes
at any rate	Conj-adverb	Yes
at first	Conj-adverb	Yes
at first blush	Conj-adverb	Yes
at first sight	Conj-adverb	Yes
at first view	Conj-adverb	Yes
at last	Conj-adverb	No
at least	Conj-adverb	No
at once	Conj-adverb	No
at that	Prep Phrase	No
at the moment when	Subordinator	Yes
at the outset	Conj-adverb	Yes
at the same time	Conj-adverb	Yes
at which point	Prep Phrase	No
back	Adverb	No
because	Subordinator	Yes
before	Conj-adverb	Yes
before	Subordinator	Yes
before long	Conj-adverb	No
before then	Prep Phrase	No
before...ever	Subordinator	No
besides	Conj-adverb	No
but	Coordinator	Yes
but then	Coordinator	No
by all means	Conj-adverb	No
by and by	Conj-adverb	No
by comparison	Conj-adverb	Yes
by contrast	Conj-adverb	Yes
by the same token	Conj-adverb	Yes

Phrase	Syntactic category	In corpus analysis
by the time	Subordinator	No
by the way	Conj-adverb	Yes
by then	Prep Phrase	No
certainly	Conj-adverb	Yes
clearly	Conj-adverb	Yes
come to think of it	Conj-adverb	No
consequently	Conj-adverb	Yes
considering that	Subordinator	No
conversely	Conj-adverb	Yes
correspondingly	Conj-adverb	Yes
despite this	Conj-adverb	Yes
despite the fact that	Subordinator	Yes
each time	Subordinator	No
earlier	Conj-adverb	Yes
either	Conj-adverb	Yes
else	Coordinator	Yes
equally	Conj-adverb	Yes
especially because	Subordinator	No
especially if	Subordinator	No
especially when	Subordinator	No
essentially, then	Conj-adverb	Yes
even	Conj-adverb	Yes
even after	Subordinator	No
even before	Subordinator	No
even if	Subordinator	Yes
even so	Conj-adverb	Yes
even then	Conj-adverb	Yes
even though	Subordinator	Yes
even when	Subordinator	No
eventually	Conj-adverb	Yes
ever since	Conj-adverb	No
every time	Subordinator	Yes
everywhere	Subordinator	No
except	Conj-adverb	Yes
except after	Subordinator	No
except before	Subordinator	No
except if	Subordinator	No
except insofar as	Subordinator	Yes
except when	Subordinator	No
failing that	Conj-adverb	No
finally	Conj-adverb	Yes
first	Conj-adverb	Yes
first of all	Conj-adverb	Yes

Phrase	Syntactic category	In corpus analysis
firstly	Conj-adverb	Yes
following this	Conj-adverb	No
for	Subordinator	Yes
for a start	Conj-adverb	Yes
for another thing	Conj-adverb	Yes
for example	Conj-adverb	Yes
for fear that	Subordinator	No
for instance	Conj-adverb	Yes
for one thing	Conj-adverb	Yes
for one,	Conj-adverb	No
for that matter	Conj-adverb	No
for the simple reason	Subordinator	Yes
for this reason	Conj-adverb	Yes
fortunately	Conj-adverb	No
from then on	Prep Phrase	No
further	Conj-adverb	Yes
furthermore	Conj-adverb	Yes
given that	Subordinator	Yes
having said that	Conj-adverb	No
hence	Conj-adverb	Yes
however	Conj-adverb	Yes
however	Subordinator	No
I mean	Phr.w/scomp	No
if	Subordinator	Yes
if ever	Subordinator	Yes
if not	Conj-adverb	Yes
if only	Subordinator	Yes
if so	Conj-adverb	Yes
in a different vein	Conj-adverb	Yes
in actual fact	Conj-adverb	Yes
in addition	Conj-adverb	Yes
in any case	Conj-adverb	Yes
in case	Subordinator	Yes
in conclusion	Conj-adverb	Yes
in contrast	Conj-adverb	Yes
in doing this	Prep Phrase	No
in fact	Conj-adverb	Yes
in other respects	Prep Phrase	No
in other words	Conj-adverb	Yes
in particular	Conj-adverb	Yes
in short	Conj-adverb	Yes
in so doing	Prep Phrase	No
in spite of that	Conj-adverb	Yes

Phrase	Syntactic category	In corpus analysis
in sum	Conj-adverb	Yes
in that	Subordinator	No
in that case	Conj-adverb	Yes
in that respect	Prep Phrase	No
in the beginning	Conj-adverb	Yes
in the case of X	Conj-adverb	Yes
in the end	Conj-adverb	Yes
in the event	Conj-adverb	No
in the first place	Conj-adverb	Yes
in the hope that	Subordinator	No
in the meantime	Conj-adverb	Yes
in this way	Conj-adverb	Yes
in truth	Conj-adverb	No
in turn	Conj-adverb	Yes
in which case	Prep Phrase	No
inasmuch as	Subordinator	Yes
incidentally	Conj-adverb	Yes
indeed	Conj-adverb	Yes
initially	Conj-adverb	Yes
insofar as	Subordinator	No
instantly	Conj-adverb	No
instead	Conj-adverb	Yes
it follows that	Phr.w/scomp	Yes
it is because	Phr.w/scomp	No
it is only because	Phr.w/scomp	No
it might appear that	Phr.w/scomp	Yes
it might seem that	Phr.w/scomp	Yes
just	Conj-adverb	No
just as	Subordinator	Yes
just then	Conj-adverb	No
largely because	Subordinator	No
last	Conj-adverb	Yes
lastly	Conj-adverb	Yes
later	Conj-adverb	Yes
lest	Subordinator	No
let us assume	Phr.w/scomp	Yes
likewise	Conj-adverb	Yes
luckily	Conj-adverb	No
mainly because	Subordinator	No
meanwhile	Conj-adverb	Yes
merely	Conj-adverb	Yes
merely because	Subordinator	Yes
mind you	Conj-adverb	No

Phrase	Syntactic category	In corpus analysis
more Xly	Conj-adverb	Yes
moreover	Conj-adverb	Yes
most Xly	Conj-adverb	Yes
much as	Subordinator	No
much later	Conj-adverb	Yes
much sooner	Conj-adverb	Yes
naturally	Conj-adverb	Yes
neither is it the case	Phr.w/scomp	Yes
nevertheless	Conj-adverb	Yes
next	Conj-adverb	Yes
next time	Subordinator	No
no doubt	Conj-adverb	Yes
no sooner than	Subordinator	No
nonetheless	Conj-adverb	Yes
not	Conj-adverb	Yes
not because	Conj-adverb	Yes
not only	Conj-adverb	Yes
not that	Conj-adverb	Yes
notably	Conj-adverb	Yes
notwithstanding that	Subordinator	Yes
notwithstanding that,	Conj-adverb	Yes
now	Conj-adverb	No
now	Subordinator	Yes
now that	Subordinator	Yes
obviously	Conj-adverb	Yes
of course	Conj-adverb	Yes
on balance	Conj-adverb	No
on condition that	Subordinator	Yes
on one hand	Conj-adverb	Yes
on one side	Conj-adverb	Yes
on the assumption that	Subordinator	Yes
on the contrary	Conj-adverb	Yes
on the grounds that	Subordinator	Yes
on the one hand	Conj-adverb	Yes
on the one side	Conj-adverb	Yes
on the other hand	Conj-adverb	Yes
on the other side	Conj-adverb	Yes
on top of this	Conj-adverb	Yes
once	Subordinator	Yes
once again	Conj-adverb	Yes
once more	Conj-adverb	Yes
only	Conj-adverb	No
only after	Subordinator	No

Phrase	Syntactic category	In corpus analysis
only because	Subordinator	No
only before	Subordinator	No
only if	Subordinator	No
only when	Subordinator	No
or	Coordinator	Yes
or again	Coordinator	No
or else	Coordinator	Yes
originally	Conj-adverb	No
otherwise	Conj-adverb	Yes
overall	Conj-adverb	Yes
particularly because	Subordinator	No
particularly if	Subordinator	No
particularly when	Subordinator	No
plainly	Conj-adverb	Yes
presently	Conj-adverb	No
presumably because	Subordinator	Yes
previously	Conj-adverb	Yes
provided that	Subordinator	Yes
providing that	Subordinator	Yes
put another way	Conj-adverb	Yes
rather	Conj-adverb	Yes
reciprocally	Conj-adverb	Yes
regardless of that	Conj-adverb	Yes
regardless of whether	Subordinator	No
second	Conj-adverb	Yes
secondly	Conj-adverb	Yes
seeing as	Subordinator	No
similarly	Conj-adverb	Yes
simply because	Subordinator	Yes
simultaneously	Conj-adverb	Yes
since	Subordinator	Yes
so	Subordinator	Yes
so that	Subordinator	Yes
soon	Conj-adverb	No
specifically	Conj-adverb	Yes
still	Conj-adverb	Yes
subsequently	Conj-adverb	Yes
such that	Subordinator	Yes
suddenly	Conj-adverb	No
summarising	Conj-adverb	Yes
summing up	Conj-adverb	Yes
suppose	Phr.w/scomp	Yes
suppose that	Phr.w/scomp	Yes

Phrase	Syntactic category	In corpus analysis
supposing that	Subordinator	Yes
sure enough	Conj-adverb	Yes
surely	Conj-adverb	Yes
that is	Conj-adverb	Yes
that is to say	Conj-adverb	Yes
that's all	Sentence	No
that's how	Phr.w/scomp	No
that's when	Phr.w/scomp	No
that's why	Phr.w/scomp	No
the fact is that	Phr.w/scomp	Yes
the first time	Subordinator	No
the moment	Phr.w/scomp	No
the more often	Subordinator	Yes
the next time	Subordinator	No
the one time	Phr.w/scomp	No
the thing is	Phr.w/scomp	No
then	Coordinator	Yes
then again	Conj-adverb	Yes
thereafter	Conj-adverb	Yes
thereby	Conj-adverb	Yes
therefore	Conj-adverb	Yes
third	Conj-adverb	Yes
thirdly	Conj-adverb	Yes
this means	Phr.w/scomp	No
this time	Conj-adverb	Yes
though	Subordinator	Yes
though	Conj-adverb	No
thus	Conj-adverb	Yes
to be precise	Prep Phrase	No
to be sure	Conj-adverb	Yes
to begin with	Conj-adverb	Yes
to conclude	Conj-adverb	Yes
to make matters worse	Conj-adverb	No
to start with	Conj-adverb	Yes
to sum up	Conj-adverb	Yes
to summarise	Conj-adverb	Yes
to take an example	Conj-adverb	Yes
to the degree that	Subordinator	Yes
to the extent that	Subordinator	Yes
too	Conj-adverb	Yes
true	Conj-adverb	Yes
ultimately	Conj-adverb	Yes
undoubtedly	Conj-adverb	Yes

Phrase	Syntactic category	In corpus analysis
unfortunately	Conj-adverb	No
unless	Subordinator	Yes
until	Subordinator	Yes
until then	Prep Phrase	No
we might say	Phr.w/scomp	Yes
well	Conj-adverb	No
what is more	Conj-adverb	Yes
when	Subordinator	Yes
whenever	Subordinator	Yes
where	Subordinator	Yes
whereas	Conj-adverb	Yes
wherein	Subordinator	Yes
whereupon	Conj-adverb	No
wherever	Subordinator	Yes
whether or not	Subordinator	No
which is why	Phr.w/scomp	No
which means	Phr.w/scomp	No
which reminds me	Phr.w/scomp	No
while	Subordinator	Yes
whilst	Subordinator	No
with that	Prep. Phrase	No
yet	Coordinator	Yes
you know	Phr.w/scomp	No

Appendix B

The Taxonomy of Cue Phrases

This appendix contains the taxonomy of cue phrases which forms the basis for the feature definitions motivated in Chapter 6. It was created prior to the development of these definitions; the intention is that it should reflect theory-neutral intuitions about the substitution relationships between cue phrases, and that the feature definitions should then be based on these intuitions.[1]

The taxonomy presented here is complex: around 150 phrases have so far been incorporated. In order to achieve a complete taxonomy, the substitutability relationship between each pair of phrases must be represented; in other words, for any two phrases X and Y in the corpus, it must be specified whether X is synonymous with, hyponymous to, hypernymous to, exclusive with, or contingently intersubstitutable with Y. For a taxonomy of 150 phrases, this means that $(150 \times 149)/2$ (= 11175!) relationships must be documented. Using diagrams of the type described in Chapter 4, the great majority of the work can be done by inheritance; but there is still large amount of information to be depicted. Some further principles of organisation have thus been used; these were described briefly in Section 4.6.4, and are described in more detail below.

B.1 Exclusive Phrases and Multicategory Phrases

The taxonomy is roughly organised into ten **categories**, as can be seen in the 'top level' diagram in Figure B.1. The categories are as follows: SEQUENCES, CAUSES, RESULTS, RESTATEMENTS, TEMPORAL RELATIONS, NEGATIVE POLARITY RELATIONS, ADDITIONAL INFORMATION RELATIONS, HYPOTHETICAL RELATIONS, SIMILARITY RELATIONS, and DIGRESSION RELATIONS. These categories only permit a rough division of phrases: some phrases (termed **exclusive phrases**) are unique to a single category, but others (termed **multicategory phrases**) appear in more than one category. Because of multicategory phrases, the categories in Figure B.1 are all be depicted as

[1] In fact, a few of the phrases discussed in Chapter 6 have not yet been incorporated into the taxonomy. However, for these phrases, the relevant substitution relationships are given in Chapter 6 itself. The decision not to include them here too was made in an effort to keep 'pre-theoretical' and 'post-theoretical' intuitions separate.

contingently substitutable with each other. Figure B.2 provides more detail, showing the exclusivity between the exclusive phrases in each category.

Relationships between the phrases within the individual categories are given in Figures B.3 to B.12. Each of these figures isolates a single category, and shows

- how the exclusive phrases in the category relate to each other;
- how the exclusive phrases in the category relate to the multicategory phrases which belong to that category;
- how the exclusive phrases in the category relate to the multicategory phrases which do *not* belong in that category. (These are called **non-shared multicategory phrases**, and obviously none of them are substitutable for any of the exclusive phrases in the category.)

Finally, in Figure B.13, the relationships between all the multicategory phrases are given. The figures thus provide a complete description of the relationships between all the cue phrases, multicategory and exclusive.

B.2 A Note about 'Re-Entrancy' in the Taxonomy

Because of the way the taxonomy is organised, many of its phrases appear in more than one diagram. Thus, all the multicategory phrases appear in Figures B.3 through B.12, and in Figure B.13. The higher-level categories in the taxonomy (such as SEQUENCES and EXCLUSIVE SEQUENCE PHRASES) are also depicted more than once: they appear in the top-level diagrams in Figures B.1 and B.2, as well as in the figure corresponding to their own category. Identical categories appearing in separate diagrams are to be regarded as 'one and the same object' in the taxonomy, and the sum of all the relationships of this object with other objects in the taxonomy is given by the union of the relations depicted in all the different diagrams.

This has an important consequence for the depiction of contingent substitutability relationships in the taxonomy. These relationships, it will be recalled, are only inferred as a default, if none of the other relationships can be inferred. The distribution of the taxonomy diagram over several pages means that *each appearance* of a item in the taxonomy must be considered before a contingent substitutability relationship can be inferred. For instance, from Figure B.1 in isolation, we would conclude that the phrases from the SEQUENCES category are all contingently substitutable with those from the CAUSES category. But other parts of the diagram present exceptions to this default: for example, in Figure B.2, the 'exclusive' SEQUENCE phrases are depicted as exclusive with the 'exclusive' CAUSE phrases.

B.3 Additional Notation Used in the Diagrams

Finally, there are two points to note about the notation used in Figures B.3 through B.12. Firstly, in each diagram, the non-shared multicategory phrases are always grouped together in a single box. This is just to make the diagrams easier to understand: the true relations between all the multicategory phrases are given in Figure B.13. Secondly, in each diagram, the arcs linking the exclusive phrases to the box entitled 'exclusive phrases' have been left out, again to make the diagrams easier to read. Instead, exclusive phrases appear in bold type. Thus, for example, in Figure B.3, the phrases *furthermore*, *lastly*, etc are all assumed to be connected to the EXCLUSIVE SEQUENCE PHRASES box.

B.4 A Note about the Linguistic Examples

Figures B.3 through B.2 each include a set of example texts; the kind of texts which can be used to motivate the substitutability diagrams. In each case, the examples have been picked to illustrate a *selection* of the substitutability relationships in the diagram.

It is not possible to provide *all* the examples needed to motivate the diagrams. For one thing, claims about synonymity, exclusivity, hyponymity or hypernymity all express general statements about *all possible linguistic contexts*: they cannot be *verified* by linguistic data, only *falsified*. Only the relationship of contingent intersubstitutability can be motivated by a finite number of examples: here, only three examples are needed; one providing a context in which two phrases X and Y are substitutable, one providing a context in which X is not substitutable for Y, and one providing a context in which Y is not substitutable for X. Even in this latter case, due to lack of space, motivating examples are not always given; but it is normally quite easy to think of them.

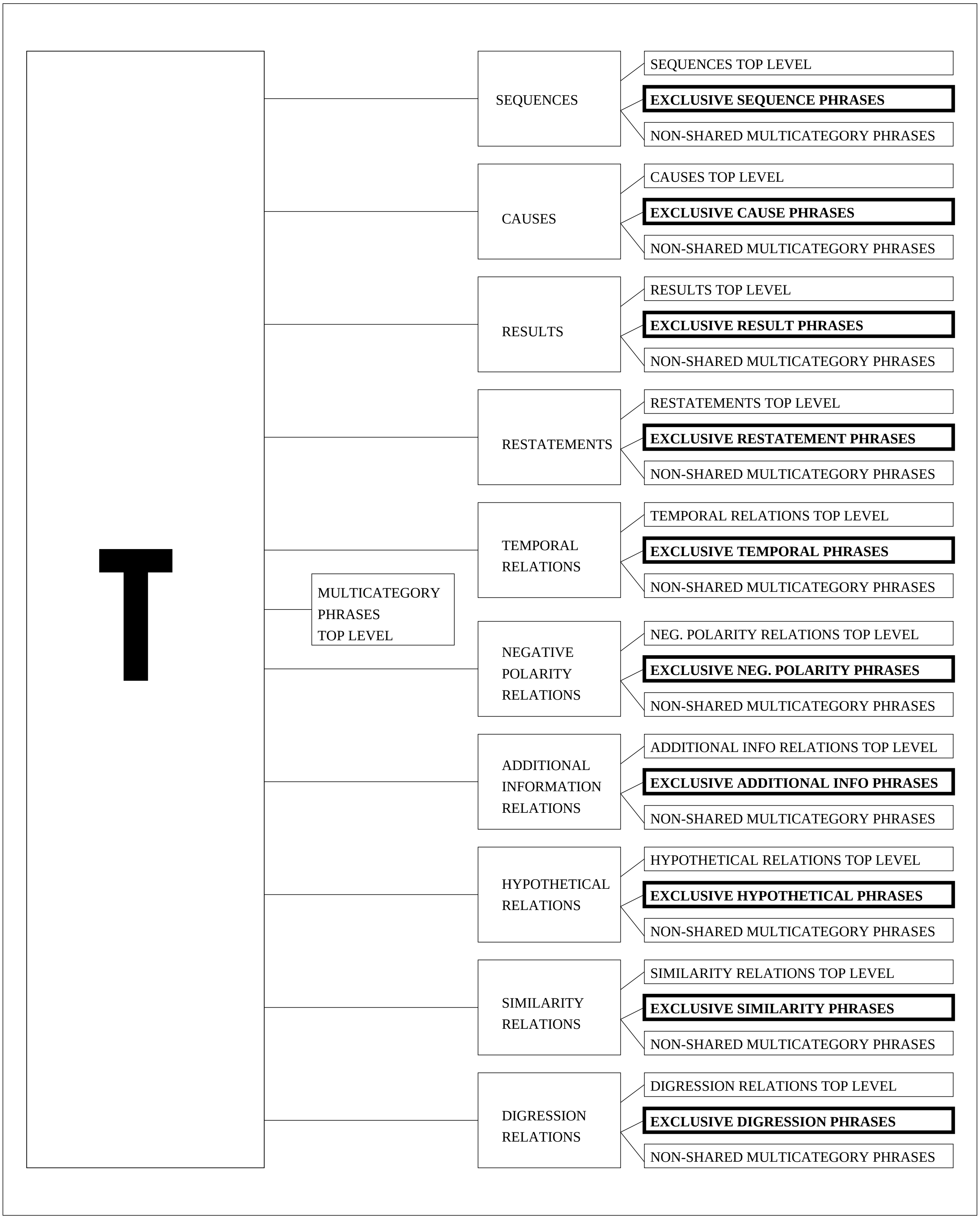

Figure B.1: The Top Level of the Taxonomy

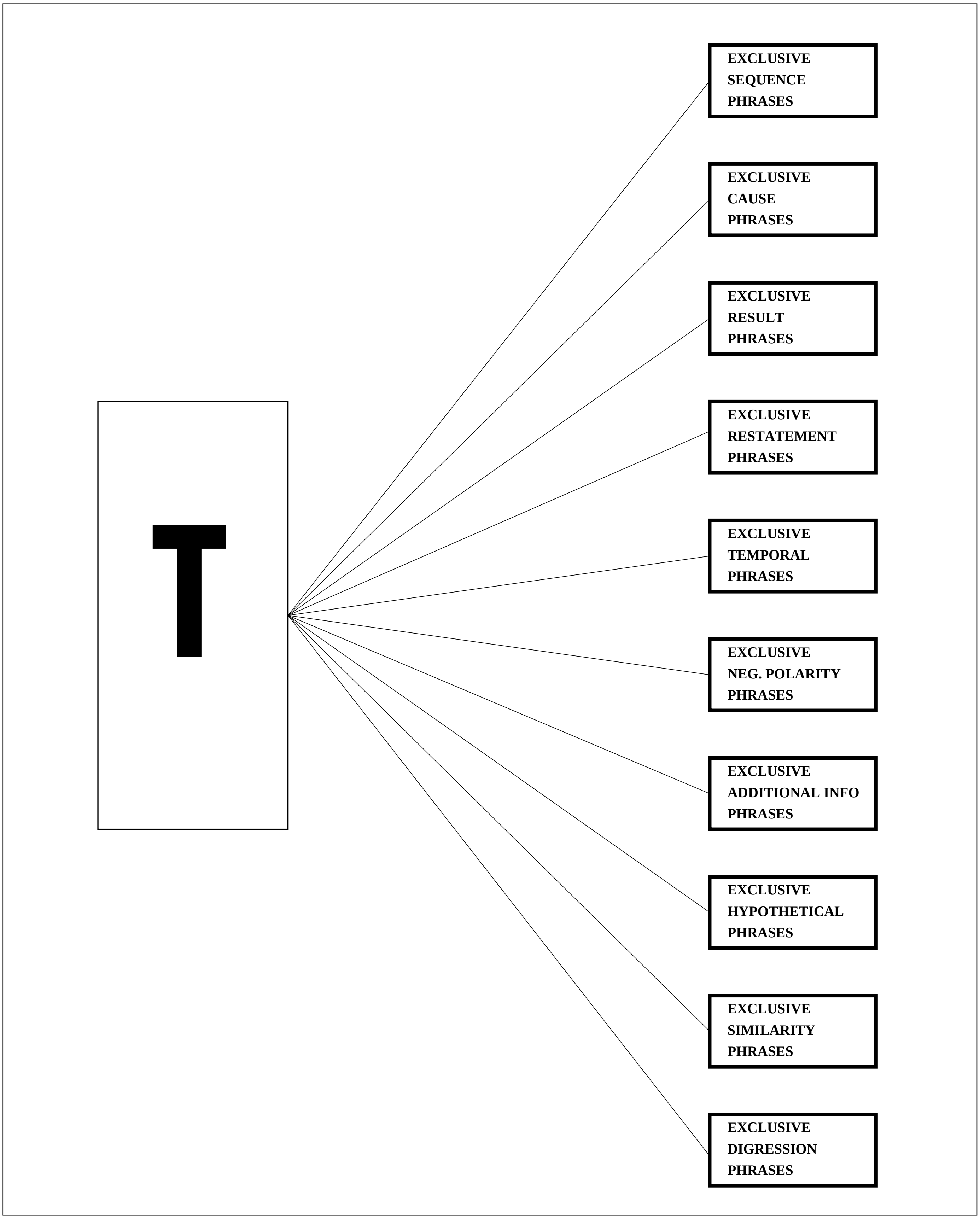

Figure B.2: 'Exclusive Phrases'

It looks as though Dan was preparing to sail.	*For a start,* ✓ *For one thing,* ✓ *Firstly,* ✓ *To start with,* # *First,* # *And*	he had taken off the sail covers and threaded the sheets; furthermore, I saw him checking the motor.
Dan set about making the boat ready to sail.	*First,* ✓ *To begin with,* ✓ *First of all,* # *For one thing,* # *Or*	he took off the sail covers and threaded the sheets; then he checked the motor.
Dan set about making the boat ready. He took off the sail cover and threaded the sheets;	*then* ✓ *and* ✓ *afterwards,* ✓ *later on,* ✓ *after this,* ✓ *in addition,* ✓ *too* → ✓ *as well* → # *furthermore,* # *besides,*	he checked the motor.
It looks as though Dan was preparing to sail. He had taken off the sail cover and threaded the sheets;	*what is more,* ✓ *and* ✓ *in addition,* ✓ *too* → ✓ *for another thing,* # *after this,* # *besides,*	I saw him checking the motor.
We should swap Liz and Kim. Liz is excellent in defence,	*whereas* ✓ *and* ✓ *while* ✓ *in addition,* ✓ *furthermore,* ✓ *on the other hand,* # *after this,* # *alternatively,*	Kim is much better in goal.
There are several possibilities. It could be in the office;	*or* ✓ *and* ✓ *alternatively,* ✓ *as well* → ✓ *then* ✓ *next* # *after this,* # *besides,*	it could be at home; finally, it could be at Phil's place.
I don't think that Bob can win. Firstly, he's out of training;	*secondly,* ✓ *next,* ✓ *(and)* ✓ *then* ✓ *on top of this* # *lastly,* # *besides,* # *after this,*	he's running at altitude; and finally, he's up against some pretty tough opponents.

Figure B.3: Sequence Phrases: Examples of Substitutability

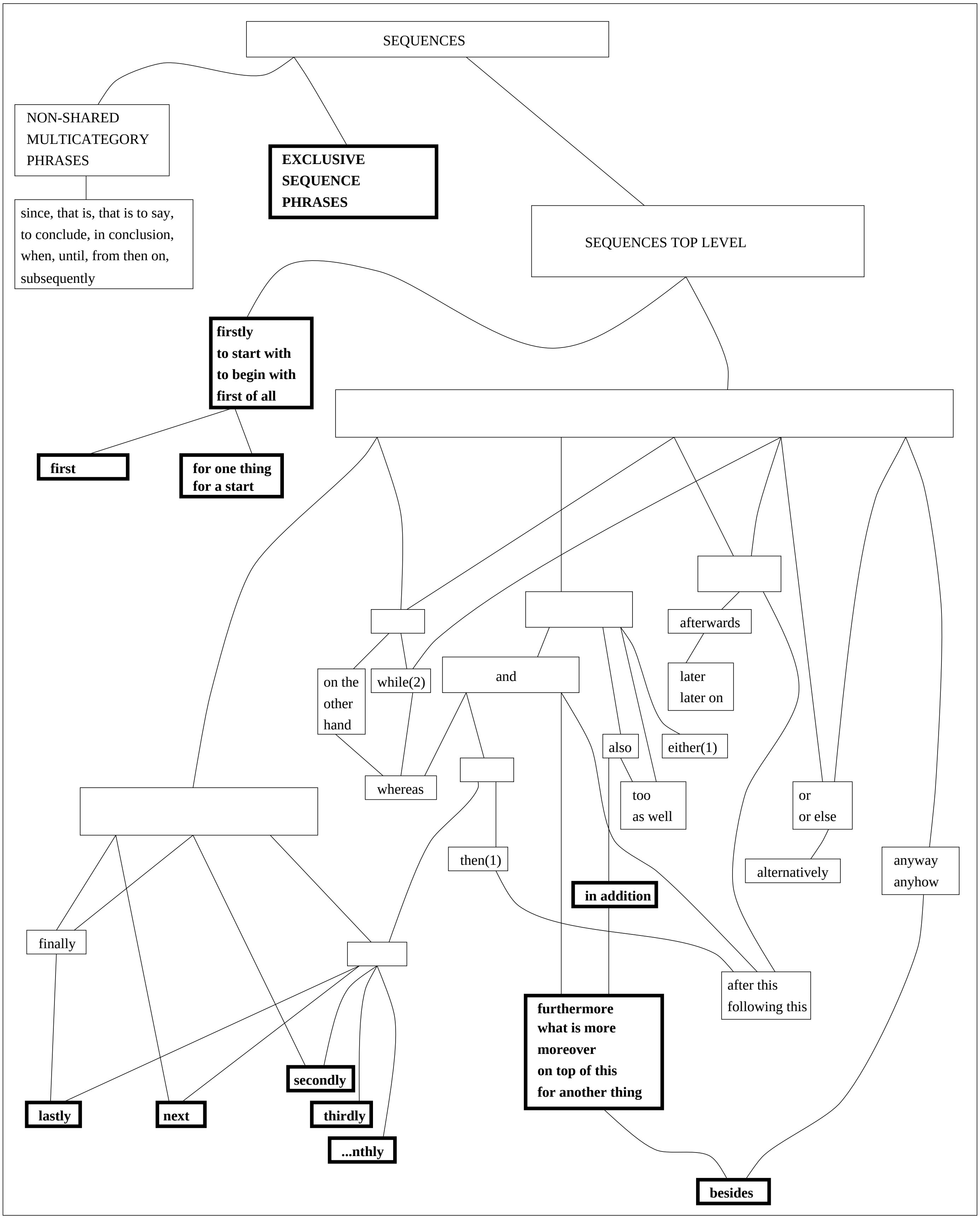

Figure B.3: Sequence Phrases: Substitutability Diagram

Sally left the engine running	*in case* # *in that* # *now that* # *because* # *when*	she was in a hurry when she came back.
It's a good proposal,	*except insofar as* √ *except in that* # *except because*	it makes no provision for bad debt.
It's a fairly good piece of work,	*considering that* √ *given that* √ *seeing as* # *because* # *in that* # *to the extent that* # *when*	you have been under a lot of pressure lately.
The proposal is useful,	*because* √ *since* √ *as* √ *in that* √ *insofar as* # *seeing as* # *to the extent that* # *considering that*	it gives us a fallback position if the forthcoming negotiations collapse.
	Seeing as √ *Given that* √ *Considering that* √ *Because* √ *Since* # *To the extent that*	we've got nothing but circumstantial evidence in this case, it's going to be difficult to get a conviction.
Compulsive gamblers fall into their addiction	*because* # *on the grounds that* # *since*	it provides an escape, however temporary, from the real world.
I have had this violin	*since* # *because* # *when*	I was a boy.
John stayed in bed that day,	*because* √ *as* √ *since* √ *on the grounds that* # *in that*	he was sick.
John must have been sick on Monday,	*because* √ *since* √ *as* √ *given that* # *on the grounds that* # *now that*	he stayed in bed all day.

Figure B.4: Cause Phrases: Examples of Substitutability

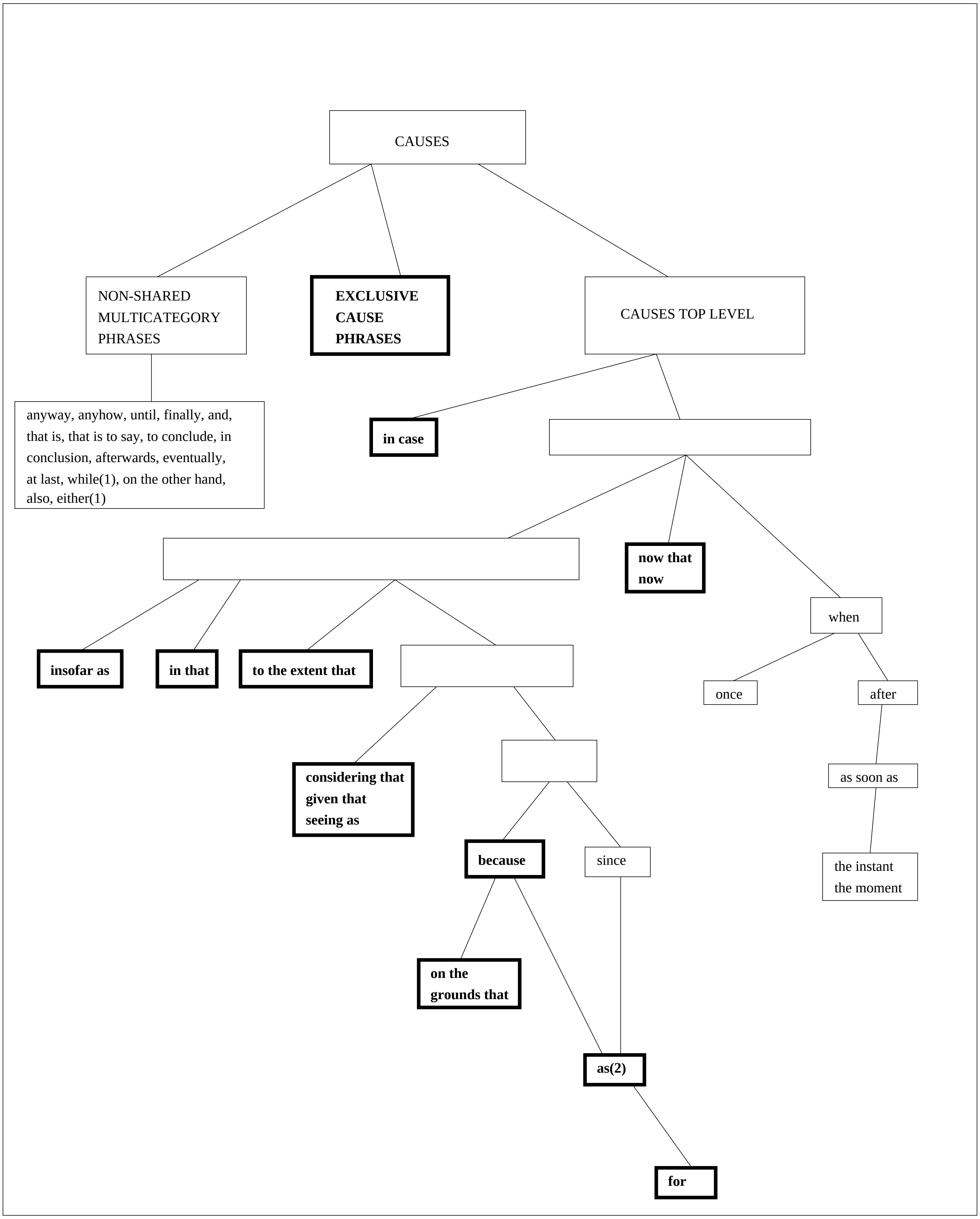

Figure B.4: Cause Phrases: Substitutability Diagram

Jim summoned his nerve and made a break for the door;	*at that,* ✓ *instantly,* ✓ *and* ✓ *after this,* ✓ *then* ✓ *as a result,* ✓ *consequently,* ✓ *so* # *in conclusion,* # *it follows that* # *to this end,* # *finally,* # *thereby,*	pandemonium broke out in the bar.
Sue left the country before the year was up;	*in so doing,* ✓ *thereby* → ✓ *immediately* → ✓ *and* ✓ *so* # *in short,* # *at that,* # *to this end,*	she lost her right to permanent residence.
We waited outside Mullen's door for three hours:	*at last,* ✓ *finally,* # *instantly,* # *at that,* # *in doing this,*	he agreed to see us.
The footprints are deep, and clearly defined.	*It follows that* ✓ *Plainly,* ✓ *That is to say,* ✓ *So* ✓ *Therefore,* # *As a result,* # *Finally,*	the thief was a heavy man.
The number is divisible by four;	*it follows that* ✓ *as a result,* ✓ *hence* ✓ *thus* # *instantly,* # *this way,*	it is divisible by two.
One night, Van Diemen nailed a pair of clogs to Mortimer's door.	*From then on,* ✓ *Thereafter,* ✓ *After this,* ✓ *(and)* # *All in all,*	Mortimer stopped coming to the Dog and Duck.
Mary held her breath,	*until* ✓ *so that* ✓ *and* ✓ *(finally)* # *from then on,* # *obviously*	she turned bright red.

Figure B.5: Result Phrases: Examples of Substitutability

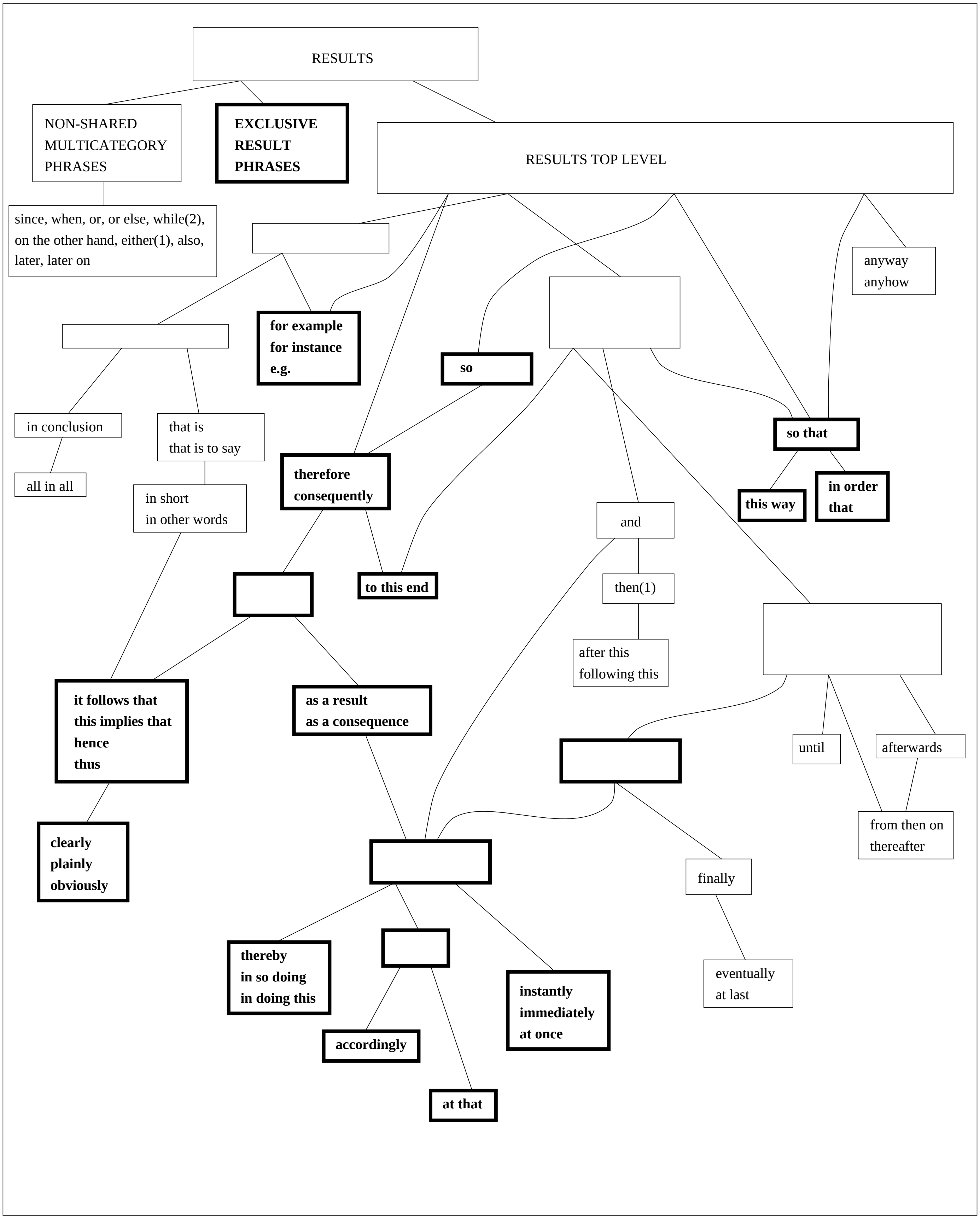

Figure B.5: Result Phrases: Substitutability Diagram

She'll be here at three o'clock this afternoon.	*At any rate,* √ *At least* √ *Or rather,* √ *That is,* # *In other words,* # *To recap,* # *In conclusion,* # *All in all,* # *Anyway,*	she hoped to be here at three: she might be a few minutes late.
It's high quality, good value for money, and easy to use:	*all in all,* √ *to conclude,* √ *in short,* √ *summing up,* # *or rather,* # *on the other hand,* # *at least,*	the RX-470 is an excellent package, and we highly recommend it.
As we have already seen, the story he told to Foley was completely different to the story he has given us.	*In conclusion,* √ *To conclude,* √ *In other words,* √ *In short,* # *All in all,* # *To summarise,* # *To recap,* # *Or rather,*	he must be lying to someone.
Foley doesn't like cowards; and if you don't agree to meet him, he'll think you're a coward. If you face up to him, he'll respect you, and he'll be more likely to make concessions.	*To sum up,* √ *To recap,* √ *Summing up:* √ *In conclusion,* √ *That is to say,* √ *In other words,* √ *All in all,* # *Or rather,* # *Anyway,*	we do think you should go to the meeting, but we don't think that you should agree to his demands.
The latest reports indicate that she doesn't want to visit our country.	*Or rather,* √ *That is,* √ *That is to say,* # *At any rate,* # *At least,* # *In conclusion,* # *All in all,* # *In other words,* # *To recap,*	she does want to visit, but not in the present political climate.

Figure B.6: Restatement Phrases: Examples of Substitutability

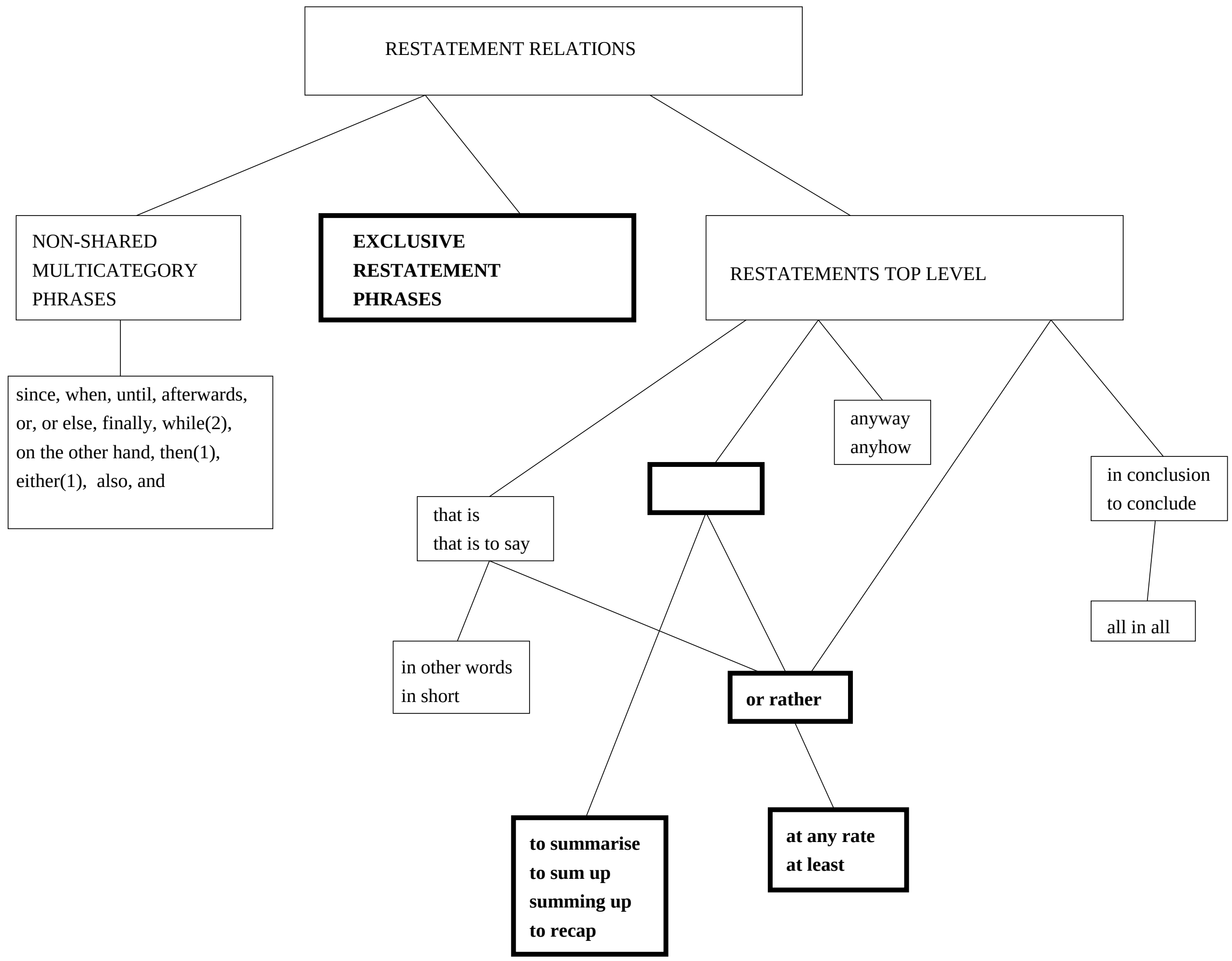

Figure B.6: Restatement Phrases: Substitutability Diagram

Jim felt the pain for the first time	*as* √ *when* √ *while* # *Meanwhile* # *and* # *once* # *after*	he was travelling to work.
It was a lazy Sunday afternoon. Bill tinkered with his old Honda,	*while* √ *and* √ *Meanwhile,* √ *as* # *when* # *then* # *finally* # *previously*	Bob mooched about on the deck reading the papers.
	As √ *When* # *While* # *After* # *Until* # *Ever since*	the children rushed by, Bob noticed that some of them were crying.
	The instant √ *As soon as* √ *After* # *As* # *Once* # *While*	we had stepped out of the house, there was a huge clap of thunder.
	Once √ *As soon as* √ *After* √ *The moment* # *As*	we had left the house, Jim began to talk more freely.
I have been frightened of bees	*ever since* √ *since* # *; previously,* # *; beforehand,* # *once* # *after*	I was a child.
The service begins at nine o'clock;	*afterwards,* √ *from then on,* √ *after this,* √ *and* √ *then* # *later on,* # *after*	it runs every half an hour.
Jim had a tattoo done while he was on service in the Philippines.	*Later on* √ *Later* √ *Afterwards* # *And* # *Following this*	he regretted it, because it made it harder for him to find a job.

Figure B.7: Temporal Phrases: Examples of Substitutability

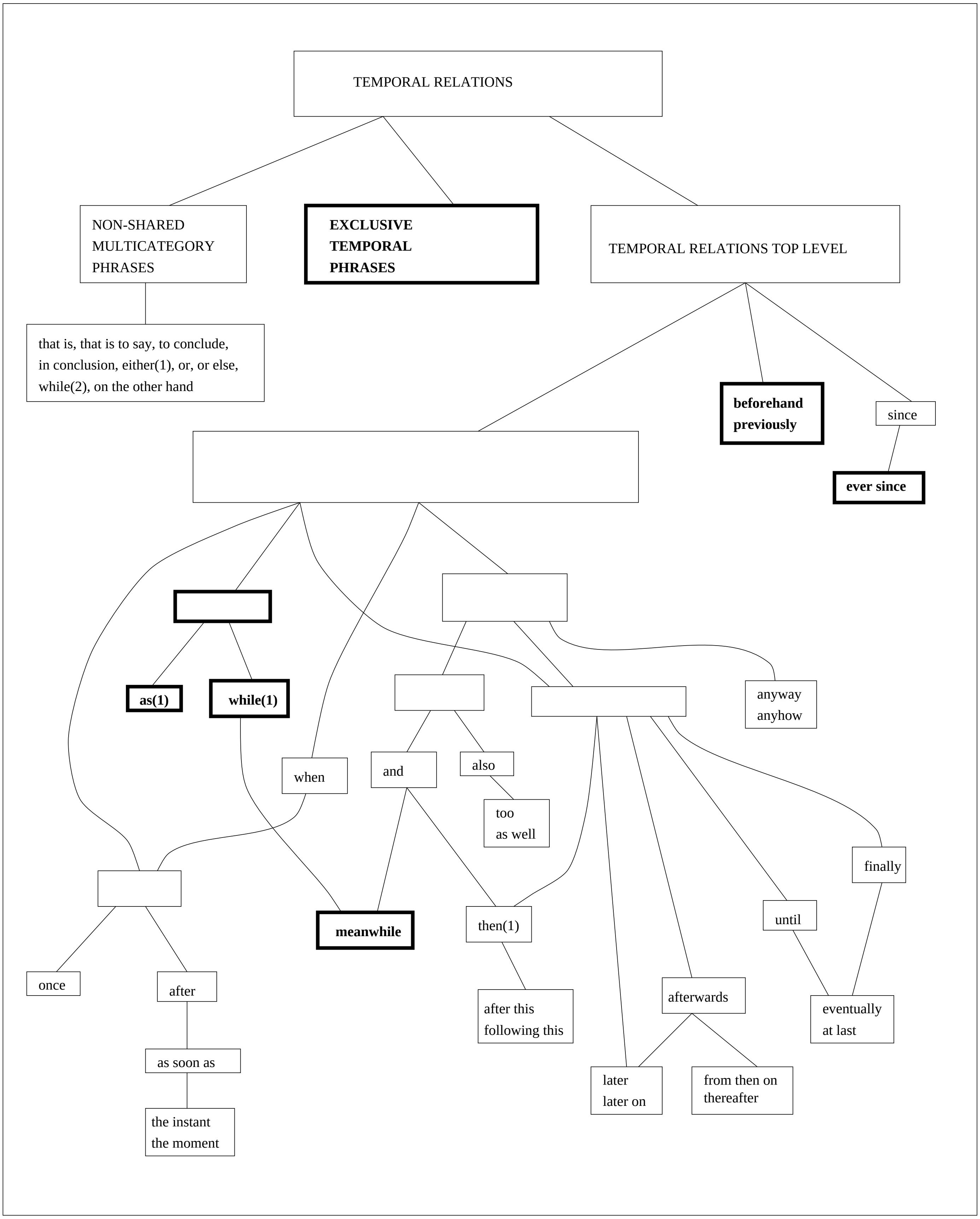

Figure B.7: Temporal Phrases: Substitutability Diagram

Jim didn't go to school that morning;	*instead,* *# rather,* *# but*	he stayed in bed.
It is not the government that is to blame;	*rather,* *# instead,* *# but*	it is the civil service.
She's part-time;	*but* ✓ *however,* ✓ *even so,* ✓ *in spite of this,* # *although* # *while*	she does more work than the rest of us put together.
She does more work than the rest of us put together,	*even though* # *but* # *while*	she's part-time.
That night, Bill and Bob were dressed very differently. Bill was wearing a red suit;	*whereas* ✓ *and* ✓ *on the other hand,* # *but* # *having said that,* # *though* # *all the same,*	Bob was wearing a blue one.
Jim was starving;	*but* ✓ *however,* # *nevertheless,*	there was no food in the house.
We could go out for a walk.	*Having said that,* ✓ *Then again,* ✓ *On the other hand,* # *Whereas* # *And* # *Though* # *Despite this,*	it's raining: we may be better off indoors.
Give me your money;	*otherwise,* ✓ *or* # *alternatively,* # *and*	I'll punch you.
	Even though ✓ *While* # *Whereas*	I am normally a timid man, on this occasion I was roused to anger.
Bill and Bob could not be more different.	*Whereas* ✓ *While* # *Though*	Bill excels in all kinds of sports, Bob lives for his schoolbooks.
	Either ✓ *(Unless)* ✓ *(Until)* # *On one hand*	you settle the matter amicably, or you will never be friends again.

Figure B.8: Negative Polarity Phrases: Examples of Substitutability

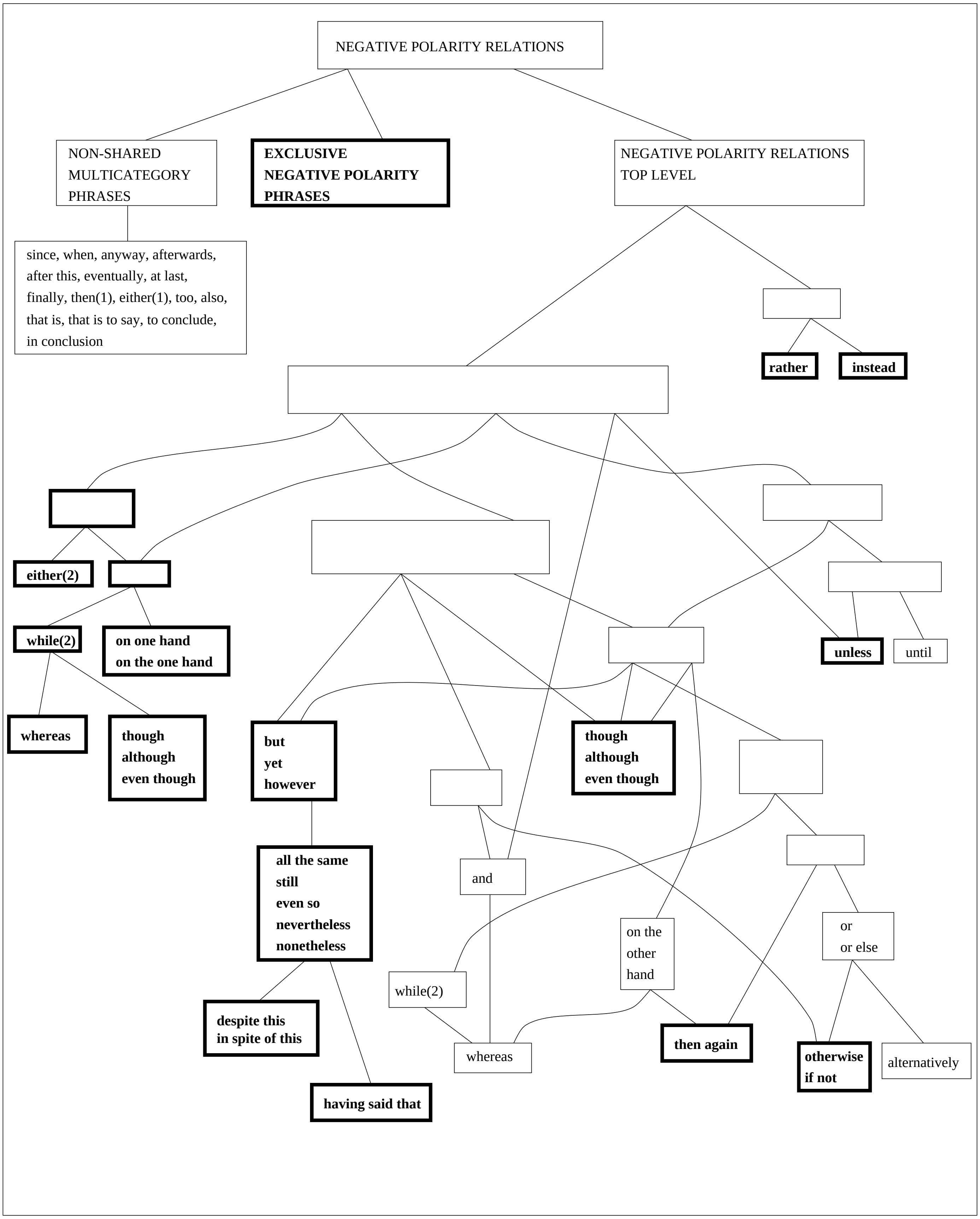

Figure B.8: Negative Polarity Phrases: Substitutability Diagram

I don't despise him at all:	*on the contrary,* √ *even* → √ *indeed* √ *as a matter of fact,* √ *in fact* # *in point of fact,*	I have a lot of respect for him.
I liked Perkins a great deal:	*indeed* √ *actually,* √ *in fact,* # *even* # *on the contrary,* # *in point of fact,* # *in truth,*	the whole squadron liked him.
We all liked Perkins.	*Even* # *Indeed,* # *On the contrary,* # *In fact,* # *Actually,*	the sergeant-major liked him.
Fred treated us very well while we were staying with him: he was very thoughtful. He	*even* # *indeed* # *on the contrary* # *in fact* # *actually*	drove us to the airport when we left.
Grandpa's always going on about his experiences during the war.	*In truth,* √ *In point of fact,* √ *In fact,* √ *Actually,* √ *As a matter of fact,* # *Indeed,* # *Even,* # *On the contrary,*	he was only in the catering corps; but he still seems to have had a pretty hairy time.

Figure B.9: Additional Information Phrases: Examples of Substitutability

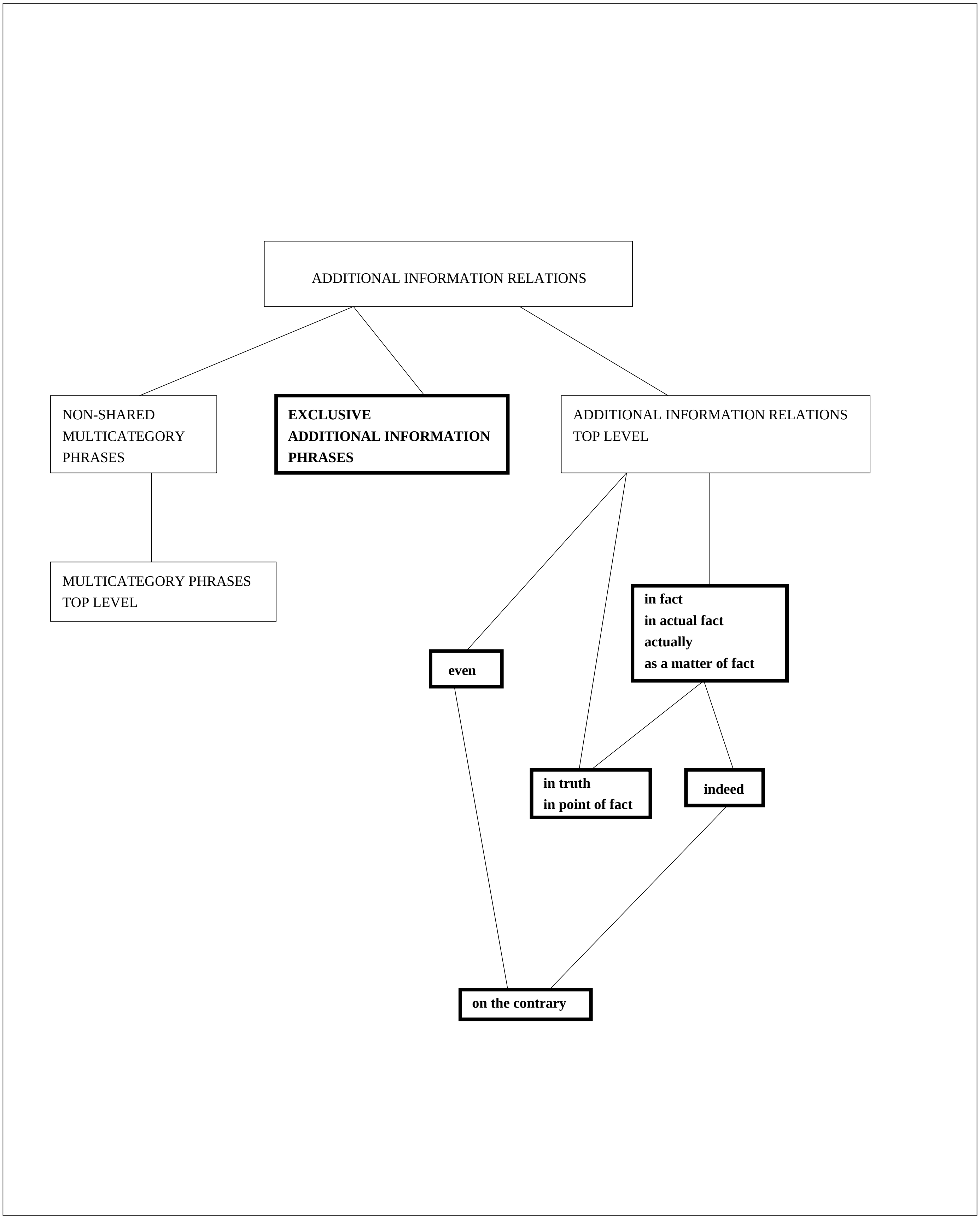

Figure B.9: Additional Information Phrases: Substitutability Diagram

	If only √ *If* √ *(Supposing that)* # *If ever* # *As long as* # *On condition that*	I had spoken up in yesterday's meeting, they would have been eating out of my hand by now!
	If ever √ *If* # *If only*	I strike it rich, I promise that I'll give you a fifty percent share in the business.
	If ever √ *If* √ *(Supposing)* # *If only* # *As long as* # *On condition that*	she decided to leave me, I would be heartbroken.
	As long as # *If* # *On condition that* # *If ever* # *If only*	the Tories are in power, we have no chance of electoral reform.
You can stay up with us	*on condition that* √ *as long as* √ *if* # *if ever* # *if only*	you promise to be quiet.
	Assuming that √ *Supposing that* √ *If* √ *As long as* # *If ever* # *On condition that*	I really am one of the lucky winners, then that holiday in Tonga may be on the cards after all!
	Assuming that √ *Supposing that* √ *(Suppose that)* √ *If* # *If ever*	they are travelling at about sixty miles an hour, they will arrive in about twenty minutes.
Suppose that Liz had genuinely forgotten our appointment:	*in that case,* √ *if so,* √ *(then)*	she wouldn't have sounded so apologetic on the phone this evening.
If you behave well tonight,	*then* # *if so,* # *in that case*	you'll have extra pocket money on Friday.

Figure B.10: Hypothetical Phrases: Examples of Substitutability

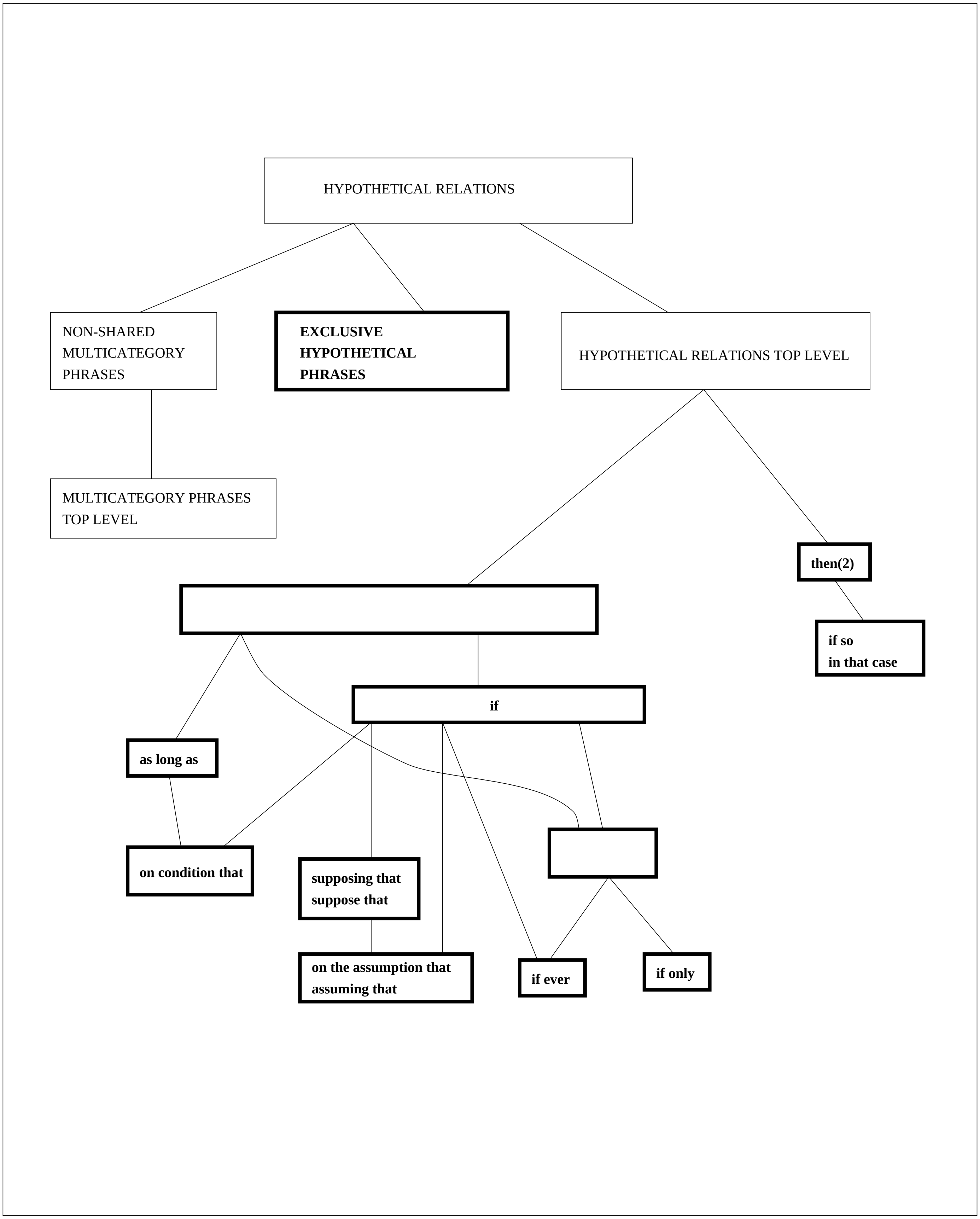

Figure B.10: Hypothetical Phrases: Substitutability Diagram

Jim jumped off the cliff, so Bill	*also* ✓ *as well* → ✓ *too* → # *in addition* # *either* # *just as*	jumped off.
I don't like Jim. He has no sense of humour;	*In addition,* ✓ *either* → ✓ *also* # *as well* # *too* # *as*	he has no brains.
Jip swam	*just as* ✓ *as* ✓ *the way* # *and* # *also* # *either*	all dogs swim: head up, with an expression of panic.

Figure B.11: Similarity Phrases: Examples of Substitutability

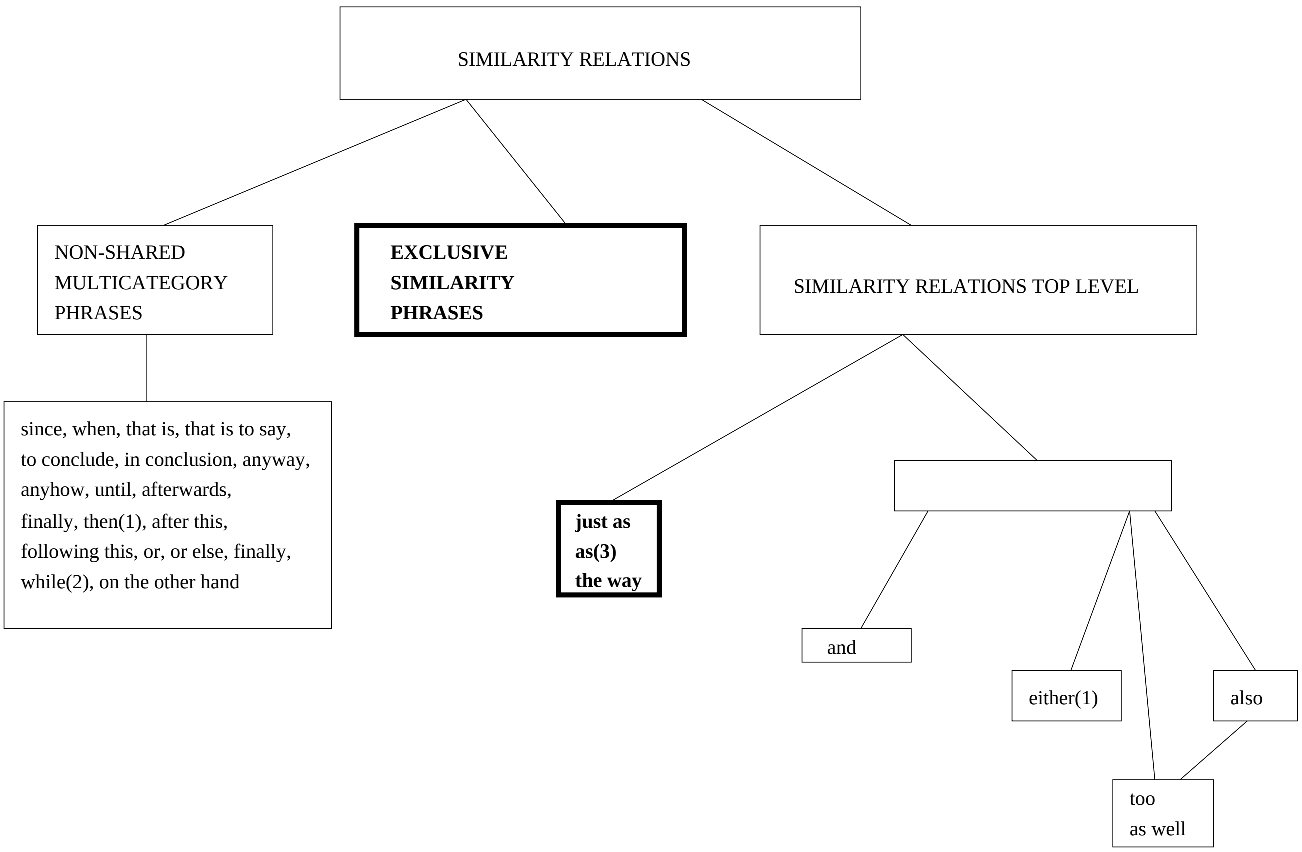

Figure B.11: Similarity Phrases: Substitutability Diagram

We were talking about Frank, and about how tidy his flat is these days. It's quite amazing, considering how it used to look.	*Incidentally,* ✓ *By the way,* # *And* # *Anyway,* # *Actually,* # *Besides,* # *Then* # *However,*	did you know that Frank has given up his job? He's now writing full time. Maybe that's why he's turned over a new leaf...

Figure B.12: Digression Phrases: Examples of Substitutability

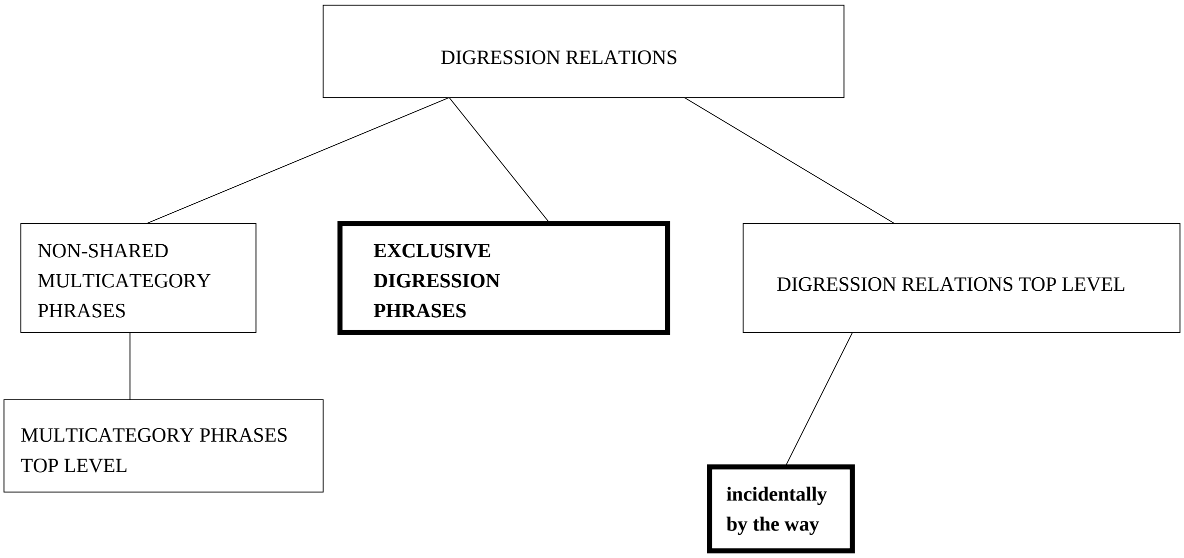

Figure B.12: Digression Phrases: Substitutability Diagram

We were all shocked	*when* # *since* # *in other words* # *and*	we heard that Kennedy had been assasinated.
Forensic has drawn a blank, door-to-door questioning is getting nowhere, and we can't even begin to think of a motive.	*In short,* # *When* # *Ever since* # *And* # *Anyway,* # *On the other hand,*	the case is a complete shambles.
John walked all afternoon	*, until* ✓ *, and* ✓ *, then* ✓ *; finally* ✓ *; eventually* # *, or* # *; on the other hand,* # *; either*	he reached a huge forest.
I'm not sure whether she should get the job. She is pretty smart;	*on the other hand,* # *whereas,* # *finally,* # *after this,* # *or esle* # *until*	she didn't show much enthusiasm in the interview.
It could be in the kitchen;	*or* ✓ *alternatively* ✓ *on the other hand,* ✓ *also* → ✓ *as well* → # *either* # *whereas* # *until* # *anyway,*	it could be in the living room.
I'm more or less a committed socialist,	*whereas* # *also* # *either* # *too*	she's far to the right of Genghis Khan.
It's good idea to keep them in their present jobs. Mullen has a flair for research;	*and* ✓ *also,* ✓ *whereas* # *as soon as* # *thereafter,* # *either*	Clarke is well suited to fieldwork.

Figure B.13: Multicategory Phrases: Examples of Substitutability

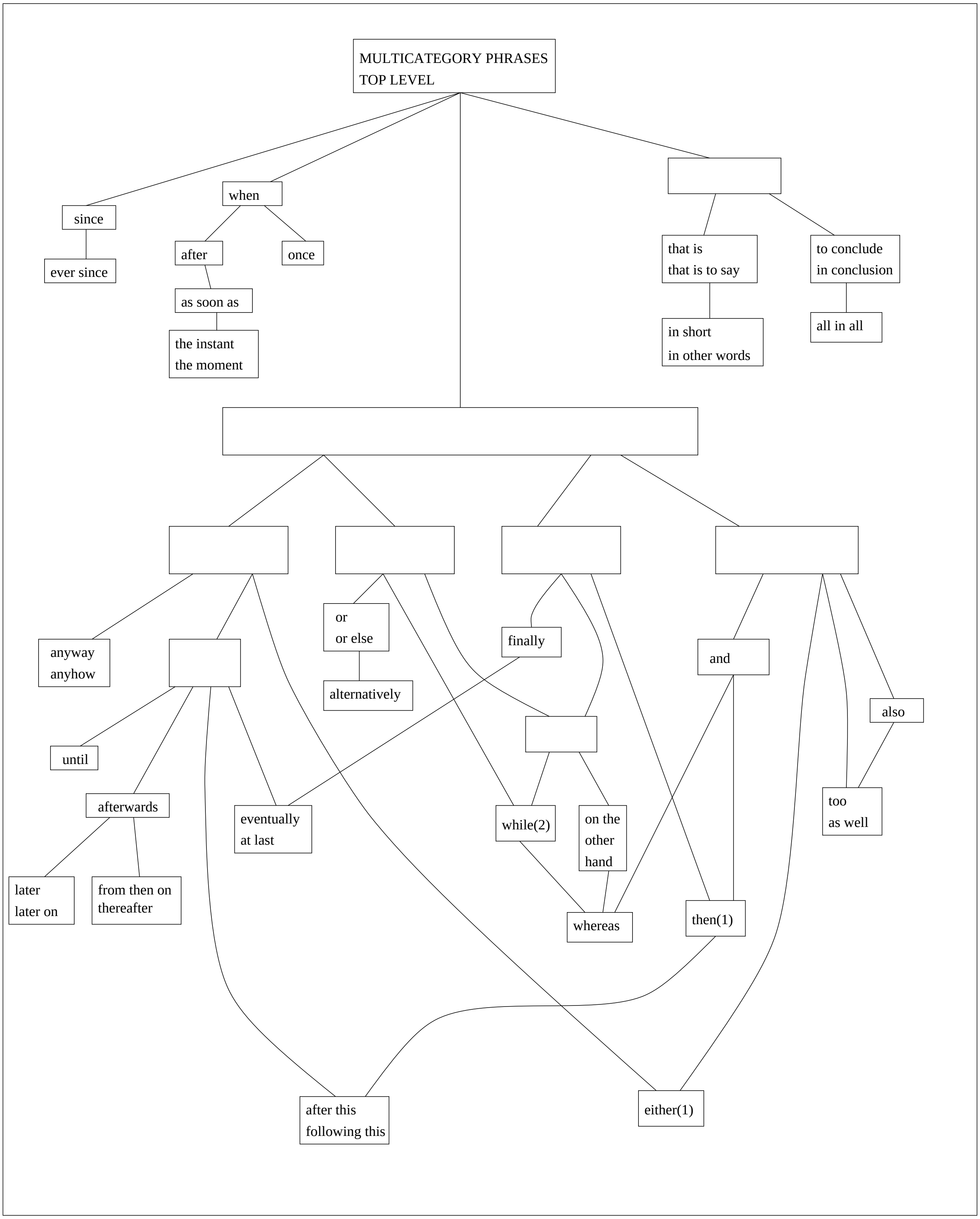

Figure B.13: Multicategory Phrases: Substitutability Diagram

Appendix C

The Core Set of Features Motivated from the Taxonomy

This appendix gives definitions for the eight features motivated in Chapter 6, along with some explanation of the variables in terms of which they are defined. The definitions given here are slightly different from those given in Chapter 6, as in this chapter they were introduced one by one, and reference could not be made to constructs yet to be introduced.

A relation is specified as holding between two adjacent text spans, S_A and S_C. The order of these spans is not initially specified.

Variables used in one definition have scope across all other definitions: thus, for instance, the value of A or C in one definition is expected to unify with the values of A and C in other definitions.

SOURCE OF COHERENCE

> SEMANTIC: the intended effect of the text containing the relation is that the reader believes some relation holds between two propositions A and C. A and C are the propositional contents of the two related text spans S_A and S_C.
>
> PRAGMATIC: the intended effect of the text containing the relation is that some relation actually holds between two propositions A and C. A and C are the intended effects of the two related text spans S_A and S_C.

ANCHOR

> CAUSE-DRIVEN: $A \in P_1 \ldots P_n$; $P_1 \wedge \ldots \wedge P_n$ is true.
>
> RESULT-DRIVEN: A corresponds to Q; and A is desired by the protagonist.

PATTERN OF INSTANTIATION

UNILATERAL: C' is on the same side of the rule as A' (but not the same as A).

BILATERAL: C' is on the opposite side of the rule to A'.

FOCUS OF POLARITY

ANCHOR-BASED: $F = A$; $F' = A'$; $I = C$; $I' = C'$.

COUNTERPART-BASED: $F = C$; $F' = C'$; $I = A$; $I' = A'$.

POLARITY

POSITIVE: $F = F'$; $I = I'$;

NEGATIVE: $F = \neg F'$; $I = I'$.

PRESUPPOSITIONALITY

PRESUPPOSED: $Precond$ is part of the left-hand side of the rule $X_1 \wedge \ldots \wedge X_n \rightarrow Y$, and A is Y.

NON-PRESUPPOSED: $Precond$ is A itself.

MODAL STATUS

ACTUAL: $Precond$ is known by the protagonist/writer.

HYPOTHETICAL: $Precond$ is not known by the protagonist/writer.

RULE TYPE

CAUSAL: the defeasible rule $P_1 \wedge \ldots \wedge P_n \rightarrow Q$ is a causal rule.

INDUCTIVE: the defeasible rule $P_1 \wedge \ldots \wedge P_n \rightarrow Q$ is an inductive rule.

There follows a brief description of each of the variables used in these definitions.

- S_A: the text span associated with the anchor of the relation.
- S_C: the text span associated with the counterpart of the relation.
- $P_1 \wedge \ldots \wedge P_n \rightarrow Q$: the rule linking A' and C'.
- A: the **anchor** of the relation.

- C: the **counterpart** of the relation.
- A': the anchor after the polarity transformation has taken place.
- C': the counterpart after the polarity transformation has taken place.
- F: the **focus of polarity**. Represents either the anchor or the counterpart before the polarity transformation.
- I: the **invariant**. Represents either the anchor or the counterpart before the polarity transformation.
- F': represents either the anchor or the counterpart after the polarity transformation.
- I': represents either the anchor or the counterpart after the polarity transformation.
- $X_1 \wedge \ldots \wedge X_n \rightarrow Y$: the rule for linking A to $Precond$.
- $Precond$: the proposition associated with the context preceding (or overlapping with) spans S_A and S_C.

Appendix D

A Preliminary Set of Relation Definitions

This appendix contains a table giving the set of relation definitions as it has so far been worked out. In fact, it is better thought of as containing definitions of cue phrases than definitions of relations. As argued in Section 6.4.2, the set of relation definitions should probably be regarded as an overlay on the set of cue phrase definitions. The issue of whether 'high-level' phrases should ever be thought of as markers of 'more abstract' relations was left unresolved; therefore it might be that the definitions of some of the higher-level cue phrases in the table should be left out of the final set of relation definitions. But in any case, it should not be necessary to devise any definitions in addition to those presented in the table.

A number of other caveats should be mentioned before the table is presented. Firstly, I have just selected a subset of cue phrases to look at at this stage. I have aimed to choose a wide range of different phrases, rather than to consider them in any particular order. In fact, the table is best seen as bringing together a collection of additional hypotheses to investigate in the style of Chapter 6. As has already been stressed, it was not possible to consider each phrase or each combination of feature values in detail in this chapter.

Secondly, the set of features used to frame the definitions is still not complete. Again, as emphasised in Chapter 6, there are many patterns of substitutability in the taxonomy which the current set of features does not yet seem able to explain; further features have still to be motivated.

Finally, there are many places where I am not sure of the value of a particular feature. These places are left blank, and should be distinguished from boxes containing a dash (—), which denote that a feature is undefined.

D.1 A Table of Cue Phrase Definitions

Cue Phrase	SOURCE OF COHERENCE	ANCHOR	PATTERN OF INSTANTIATION	FOCUS OF POLARITY	POLARITY	PRESUPPOSITION-ALITY	MODAL STATUS	RULE TYPE
if S_A, S_C	—	—	—	COUNT	—	PRES	HYP	CAUS
S_A; despite this, S_C	PRAG	CAUS	BILAT	COUNT	NEG	NON	ACT	CAUS
even if S_A, S_C		—	—	COUNT	NEG	PRES	HYP	CAUS
even when S_A, S_C			—	COUNT	NEG	PRES	ACT	CAUS
S_A; otherwise S_C		RES	BILAT	ANCH	NEG	NON	ACT	CAUS
unless S_A, S_C	PRAG	CAUS	BILAT	ANCH	NEG	PRES	HYP	CAUS
until S_A, S_C	SEM	CAUS	BILAT	ANCH	NEG	PRES	ACT	CAUS
because S_A, S_C	PRAG		BILAT	COUNT	POS	PRES	ACT	CAUS
S_A, in order that S_C	PRAG	RES	BILAT		POS	PRES		CAUS
to S_A, S_C	PRAG	RES			POS	PRES		CAUS
by S_A, S_C	PRAG	RES	BILAT		POS	PRES	ACT	CAUS
S_A; instead S_C			UNIL	COUNT	NEG	NON	ACT	CAUS
S_A; rather S_C			UNIL	COUNT	NEG	NON	ACT	CAUS
S_A or S_C	—	—	—	ANCH	NEG	NON	HYP	CAUS
S_A; besides, S_C	PRAG		UNIL	COUNT	NEG	NON		CAUS
S_C, before S_A	SEM					PRES	ACT	CAUS
S_A; instantly S_C	SEM					NON	ACT	CAUS
S_A; before long S_C	SEM					NON	ACT	CAUS
S_A; suddenly S_C	SEM					NON	ACT	CAUS
S_A; not that S_C	PRAG					NON		
S_A; even then S_C					NEG		—	CAUS
S_A; then S_C	—				POS		—	
S_A; next S_C	—				POS	NON	ACT	
S_A; finally S_C	—				POS	NON	ACT	
only if S_A, S_C	—				NEG	PRES	HYP	CAUS
only when S_A, S_C	SEM				NEG	PRES	ACT	CAUS
S_A; in that case, S_C					POS	NON	HYP	CAUS
S_A, so S_C	—	—	BILAT		POS			CAUS
S_A; it follows that S_C	PRAG		BILAT		POS	NON	ACT	CAUS
S_A; as a result S_C	SEM	CAUS	BILAT	COUNT	POS	NON	ACT	CAUS
S_A; but S_C	—	—	—	COUNT	NEG	NON	ACT	—
S_A; as it happened, S_C		RES	UNIL		—	NON	ACT	CAUS
S_A; fortunately, S_C	SEM	RES	UNIL		POS	NON	ACT	CAUS
S_A; unfortunately, S_C	SEM	RES	UNIL		NEG	NON	ACT	CAUS
S_A; on the other hand S_C	—		UNIL		NEG	NON	ACT	IND
S_A, whereas S_C	SEM		UNIL		NEG	PRES	ACT	IND
S_A; then again, S_C	PRAG		UNIL		NEG	NON	ACT	IND
admittedly, S_A; but S_C	PRAG		—		NEG	NON	ACT	
provided that S_A, S_C			BILAT	COUNT	—	PRES	HYP	CAUS
when S_A, S_C	SEM	—	—	COUNT		PRES	ACT	
S_A; meanwhile S_C	SEM		UNIL			NON	ACT	CAUS
while S_A, S_C	—		—		—	PRES	ACT	—?
S_A and S_C	—		—		—	NON		—?
S_A; furthermore, S_C	PRAG		UNIL		POS	NON	ACT	IND?