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.

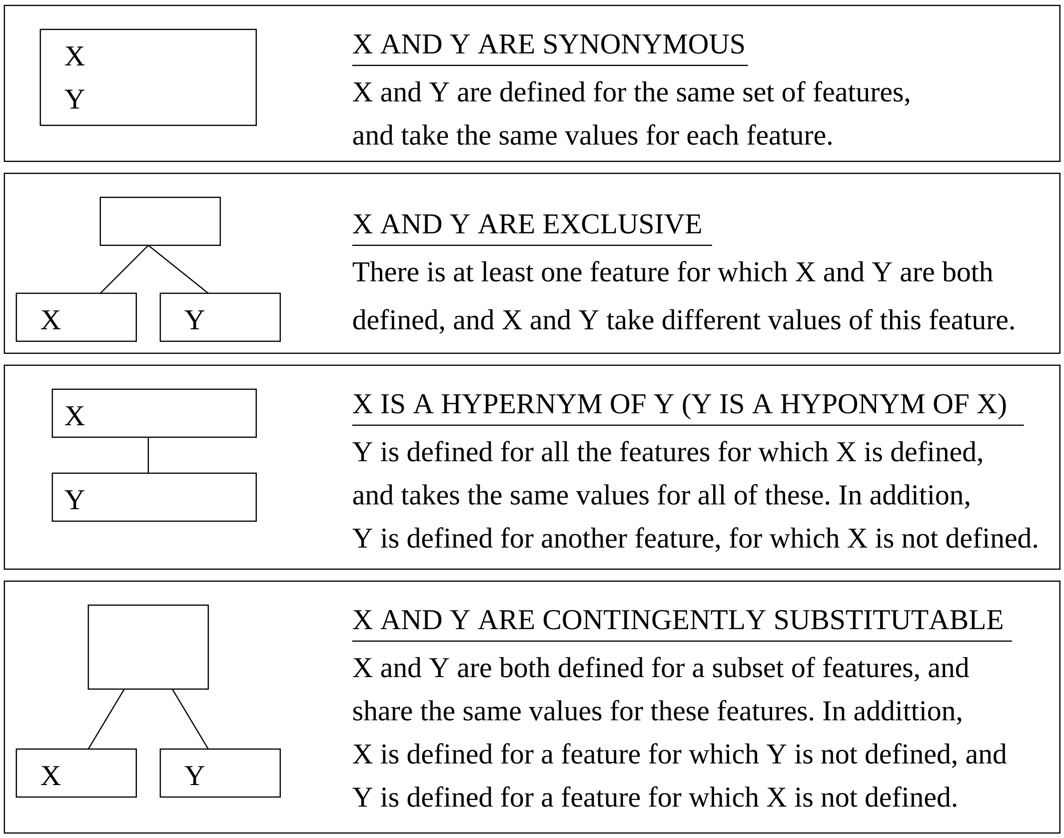

Figure 5.1: Feature-Theoretic Interpretations of Substitutability Relationships

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.

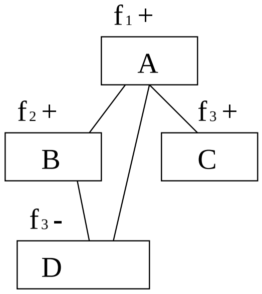

Figure 5.2: A Feature-Theoretic Account of Phrases A, B, C and D

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 $f_1 \ldots f_n$ 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:

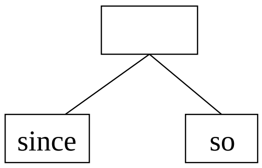

Figure 5.3: Two 'Swap-Substitutable' Phrases

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

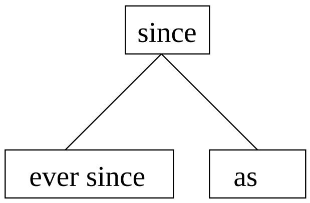

Figure 5.4: A Disjunctive Cue Phrase and its Hyponyms

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:

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

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

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

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*.

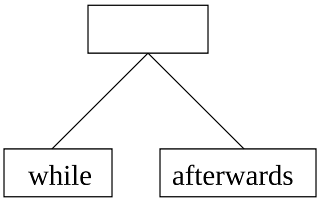

Figure 5.5: Exclusive Phrases

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

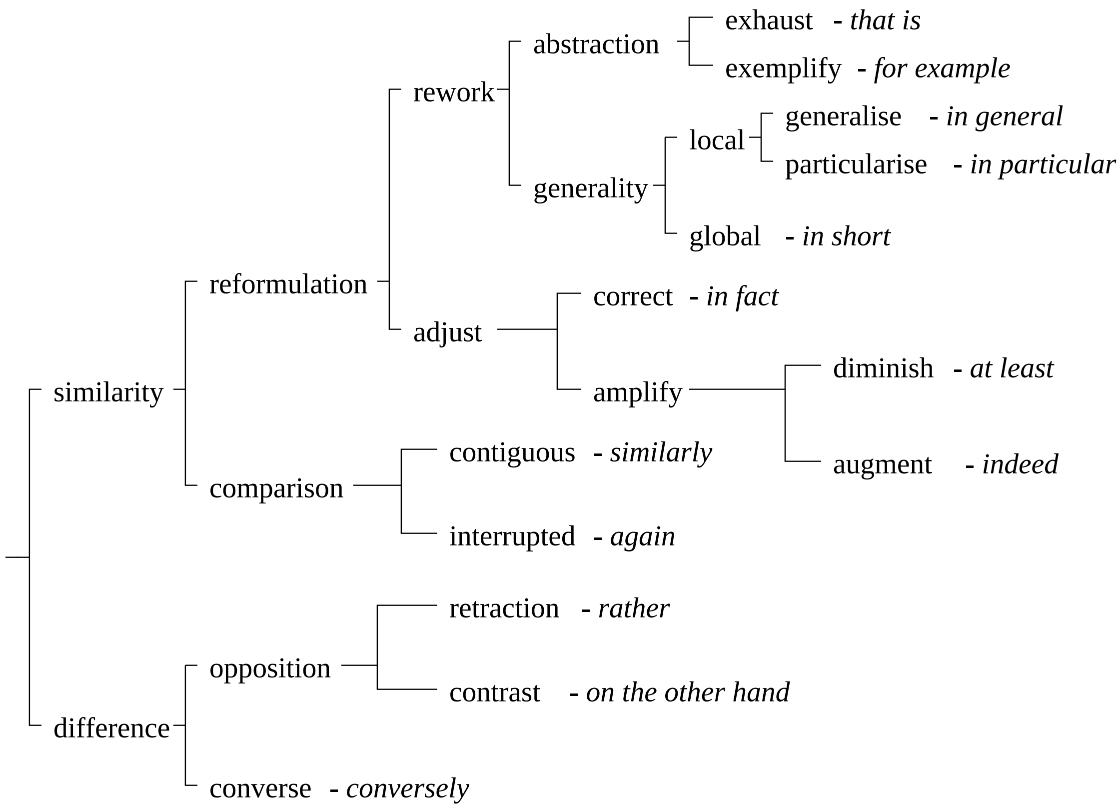

Figure 5.6: An Extract from Martin's Systemic Networks (Martin (1992) p.217)

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.

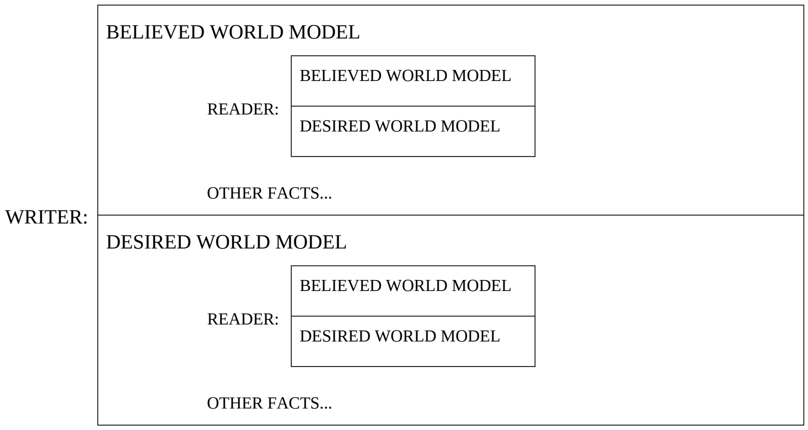

Figure 5.7: Primitives for Feature Definitions: The Top Level of Structure

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.