

Chapter 5

DISCOURSE-ORIENTED ANAPHORA SYSTEMS AND THEORIES

It is indeed harmful to come under the sway of utterly new and strange doctrines.

– Confucius¹

The relationship between theme on the one hand and pronominalization, anaphora and reference in general on the other has often been noted – for example by Kuno (1975), Givón (1975), Hirst (1976b) and Hinds (1977). In this section we will look at some work which attempts to explicate and/or exploit this relationship in resolving anaphora.

5.1. Concept activatedness

Robert Kantor (1977) has investigated the problem of why some pronouns in discourse are more comprehensible than others, even when there is no ambiguity or anomaly. In Kantor's terms, a hard-to-understand pronoun is an example of INCONSIDERATE discourse, and speakers (or, more usually, writers) who produce such pronouns lack SECONDARY [LINGUISTIC] COMPETENCE. In our terms, an inconsiderate pronoun is one that is not properly in focus.

I will first summarize Kantor's work, and then discuss what we can learn about focus from it.

5.1.1. Kantor's thesis

Kantor's main exhibit is the following text:

- (5-1) A good share of the amazing revival of commerce must be credited to the ease and security of communications within the empire. The Imperial fleet kept the Mediterranean Sea cleared of pirates. In each province, the Roman emperor repaired or constructed a number of skillfully designed roads. They were built for the army

¹From: Ware, James R (translator). *The sayings of Confucius*. New York: Mentor, 1955.

but served the merchant class as well. Over them, messengers of the Imperial service, equipped with relays of horses, could average fifty miles a day.

He claims that the *they* in the penultimate sentence is hard to comprehend, and that most informants need to reread the previous text to find its referent. Yet the sentence is neither semantically anomalous nor ambiguous – *the roads* is the only plural NP available as a referent, and it occurs immediately before the pronoun with only a full-stop intervening (cf (4-9)). To explain this paradox is the task Kantor set himself.

Kantor's explanation is based on discourse topic and the listener's expectations. In (5-1), the discourse topic of the first three sentences is *easing and securing communication*. In the fourth sentence, there is an improper shift to the roads as the topic: improper, because it is unexpected, and there is no discourse cue to signal it. Had the demonstrative *these roads* been used, the shift would have been okay. (Note that a definite such as *the roads* is not enough.) Alternatively, the writer could have clarified the text by combining last three sentences with semicolons, indicating that the last two main clauses were to be construed as relating only to the preceding one rather than to the discourse as a whole.

Kantor identifies a continuum of factors affecting the comprehension of pronouns. At one end is UNRESTRICTED EXPECTATION and at the other NEGATIVE EXPECTATION. What this says in effect is that a pronoun is easy to understand if expected, and difficult if unexpected. This is not as vacuous as it at first sounds; Kantor provides an analysis of some subtle factors which affect expectation.

The most expected pronouns are those whose referent is the discourse topic, or something associated with it (though note the qualifications to this below). Consider:

- (5-2) The final years of Henry's reign, as recorded by the admiring Hall, were given over to sport and gaiety, though there was little of the licentiousness that characterized the French court. The athletic contests were serious but very popular. Masques, jousts and spectacles followed one another in endless pageantry. He brought to Greenwich a tremendously vital court life, a central importance in the country's affairs, and above all, a great naval connection.²

In the last sentence, *he* is quite comprehensible, despite the distance back to its referent, because the discourse topic in all the sentences is *Henry's reign*. An example of the converse – an unexpected pronoun which is difficult despite recency – can be seen in (5-1) above. Between these two extremes are other cases involving references to aspects of the local topic, changes in topic, syntactic parallelism, and, in topicless instances, recency (though the effect of recency decays very fast). I will not describe these here; the interested reader

²From: Hamilton, Olive and Hamilton, Nigel. *Royal Greenwich*. Greenwich: The Greenwich

is referred to section 2.6.5 of Kantor's dissertation (1977).

Kantor then defines the notion of the ACTIVATEDNESS of a concept. This provides a continuum of concept givenness, which contrasts with the simple binary given-new distinction usually accepted in linguistics (for example Chafe (1970)). Kantor also distinguishes activatedness from the similar "communicative dynamism" of the Prague school (Firbas 1964). Activatedness is defined in terms of the comprehensibility phenomena described above: the more activated a concept is, the easier it is to understand an anaphoric reference to it. Thus activatedness depends upon discourse topic, context, and so forth.

5.1.2. The implications of Kantor's work

What are the ramifications of Kantor's thesis for focus? Clearly, the notions of activatedness and focus are very similar, though the latter has not previously been thought of as a continuum. It follows that the factors Kantor finds relevant for activatedness and comprehensibility of pronouns are also important for those of us who would maintain focus in computer-based NLU systems; we will have to discover discourse topic and topic shifts, generate pronominalization expectations, and so forth.

In other words, if we could dynamically compute (and maintain) the activatedness of each concept floating around, we would have a measure for the ordering of the focus set by preferability as referent – the referent for any given anaphor would be the most highly activated element which passes basic tests for number, gender and semantic reasonableness. And to find the activatedness of the concepts, we follow Kantor's pointers (which he himself concedes are very tenuous and difficult) to extract and identify the relevant factors from the text.

It may be objected that all we have done is produce a mere notational variant of the original problem. This is partly true. One should not gainsay the power of a good notation, however, and what we can buy here even with mere notational variance is the (perhaps limited, but non-zero) power of Kantor's investigations. And there is more to it than that. Previously, it has been thought that items either are in focus or they aren't, and that at each separate anaphor we need to compute a preference ranking of the focus elements for that anaphor. What Kantor tells us is that such a ranking exists independent of the actual use of anaphors in the text, and that we can find the ranking by looking at things like discourse topic.

Some miscellaneous comments on Kantor's work:

1 It can be seen as a generalization albeit a weakening of Grosz's (1977a, 1977b, 1978) findings on focus in task-oriented dialogues (where each sub-task becomes the new discourse topic, opening up a new set of possible referents), which are discussed below in section 5.2. (Kantor and Grosz were apparently

Bookshop, 1969. Quoted by Halliday and Hasan (1976:14), quoted by Kantor (1977).

5.1.2 *The implications of Kantor's work*

unaware of each other's work; neither cites the other.)

2 It provides an explanation for focus problems that have previously baffled us. For example, in section 4.2 I contemplated the problem of the ill-formedness of this text:

(5-3) *John left the window and drank the wine on the table. It was brown and round.

I had previously (Hirst 1977a) thought this to be due to a syntactic factor – that cross-sentence pronominal reference to an NP in a relative clause or adjectival phrase qualifying an NP was not possible. However, it can also be explained as a grossly inconsiderate pronoun which does not refer to the topic properly – *the table* occurs only as a descriptor for the wine, and not as a concept in its own right. This would be a major restriction on possible reference to sub-aspects of topics.³

3 Kantor makes many claims about comprehensibility and the degree of well-formedness of sentences which others (as he concedes) may not agree with. He uses only himself (and his friends, sometimes) as an informant, and then only at an intuitive level.⁴ Claims as strong and subtle as Kantor's cry out for empirical testing. Kieras (1978), to mention but one, has performed psycholinguistic experiments on the comprehensibility of paragraphs. Kantor's claims need verification by similar experiments. (Unfortunately, I myself am not in a position to do this.)⁵

5.2. Focus of attention in task-oriented dialogues

5.2.1. Motivation

Barbara Grosz (1977a, 1977b, 1978) studied the maintenance of the focus of attention in task-oriented dialogues and its effect on the resolution of definite reference, as part of SRI's speech understanding system project (Walker 1976, 1978). By a TASK-ORIENTED dialogue is meant one which has some single major well-defined task as its goal. For example, Grosz collected and studied dialogues in which an expert guides an apprentice in the assembly of an air compressor. She found that the structure of such dialogues parallels the

³Note however that this restriction may apply to all relative clauses and adjectival phrases. Then the syntactic explanation would still be correct and would be descriptively simpler.

⁴I do not deny that I am guilty too. But I at least try to do penance, in footnote 8 of Chapter 4 and in section 7.3. I also suggest that Kantor is more culpable than I, because of the peculiar subtlety of the phenomena he studied and because his results rely so heavily on his claims of well- and ill-formedness.

⁵Kantor tells me that he hopes to test some of his assertions by observing the eye movements of readers of considerate and inconsiderate texts, to find out if inconsiderate texts actually make readers physically search back for a referent.

structure of the task. That is, just as the major task is divided into several well-defined sub-tasks, and these perhaps into sub-sub-tasks and so on, the dialogue is likewise divided into sub-dialogues, sub-sub-dialogues, etc,⁶ each corresponding to a task component, much as a well-structured Algol program is composed of blocks within blocks within blocks. As the dialogue progresses, each sub-dialogue in turn is performed in a strict depth-first order corresponding to the order of sub-task performance in the task goal (though note that some sub-tasks may not be ordered with respect to others). As we will see, this dialogue structure can be exploited in reference resolution.

Grosz's aim was to find ways of determining and representing the FOCUS OF ATTENTION of a discourse – that is, roughly speaking, its global theme and the things associated therewith – as a means for constraining the knowledge an NLU system needs to bring to bear in understanding discourse. In other words, the focus of attention is that knowledge which is relevant at a given point in a text for comprehension of the text.⁷ Grosz claims that antecedents for definite reference can be found in the focus of attention. That is, the focus of attention is a superset of focus in our sense, the set of referable concepts (in this case definite reference, not just anaphoric reference). Moreover, no element in the focus of attention is excluded from being a candidate antecedent for a definite NP. Grosz thereby implies that all items in the focus of attention can be referred to, and that hence the two senses of the word *focus* are actually identical.

5.2.2. Representing and searching focus

In Grosz's representation, which uses a partitioned semantic net formalism (Hendrix 1975a, 1975b, 1978), an EXPLICIT FOCUS corresponds to a sub-dialogue, and includes, for each concept in it, type information about that concept and any situation in which that concept participates. For each item in the explicit focus, there is an associated IMPLICIT FOCUS, which includes subparts of objects in explicit focus, subevents of events in explicit focus, and participants in those subevents. The implicit focus attempts to account for reference to items that have a close semantic distance to items in focus (see sections 2.4.2 and 6.7), or which have a close enough relationship to items in focus to be able to be referred to (see section 2.4.2). The implicit focus is also used in detecting focus shifts (see below).

Then, at any given point in a text, antecedents of definite non-pronominal NPs can be found by searching through the explicit and implicit focus for a match for the reference. After checking the other non-pronominal NPs in the same sentence to see if the reference is intrasentential, the CURRENTLY ACTIVE

⁶Below I will use the prefix *sub-* generically to include *sub-sub-sub-* . . . to an indefinite level.

⁷In her later work (Grosz 1978), Grosz emphasizes *focusing* as an active process carried out by dialogue participants.

5.2.2 Representing and searching focus

explicit focus (the focus corresponding to the present sub-dialogue) is searched, and then if that search is not successful, the other currently open focus spaces (that is, those corresponding to sub-dialogues that the present sub-dialogue is contained in) are searched in order, back up to the top of the tree. As part of the search the implicit focus associated with each explicit focus is checked, as are subset relations, so that if a novel, say, is in focus, it could be referred to as *the book*. If there is still no success after this, one then checks whether the NP refers to a single unique concept (such as the sun), contains new information (such as *the red coat*, when a coat is in focus, but not yet known to be red), or refers to an item in implicit focus.

A similar search method could be used for pronouns. However, since pronouns carry much less information than other definite NPs, more inference is required by the reference matching process to disambiguate many syntactically ambiguous pronouns, and it would be necessary to search focus exhaustively, comparing reasonableness of candidate referents, rather than stopping at the first plausible one. In addition, other constraints on pronoun reference, such as local (rather than global) theme, and default referent, would also need to be taken into account; Grosz's mechanisms do not do this. However, Grosz does show how a partitioned network structure can be used to resolve certain types of ellipsis by means of syntactic and semantic pattern matching against the immediately preceding utterance, which may itself have been expanded from an elliptical expression. She leaves open for future research most of the problems in relating pronouns to focus.

5.2.3. Maintaining focus

Given this approach, one is then faced with the problem of deciding what the focus is at a given point in the discourse. For highly constrained task-oriented dialogues such as those Grosz considered, the question of an initial focus does not arise; it is, by definition, the overall task in question. The other component of the problem, handling changes and shifts in the focus, is attacked by Grosz in a top-down manner using the task structure as a guide.

A shift in focus can be indicated explicitly by an utterance, such as:

- (5-4) Well, the reciprocating afterburner nozzle speed control is assembled. Next, it must be fitted above the preburner swivel hose cover guard cooling fin mounting rack.

In this case, the reciprocating afterburner nozzle speed control assembly sub-task and its corresponding sub-dialogue and focus are closed, and new ones are opened for the reciprocating afterburner nozzle speed control fitting, dominated by the same open sub-tasks/sub-dialogues/focuses in their respective trees that dominated the old ones. If however the new sub-task were a sub-task of the old one, then the old one would not be closed, but the new one added to

the hierarchy below it as the new active focus space. The newly created focus space initially contains only those items referred to in the utterance, and those objects associated with the current sub-task. (Being ABLE to bring in the associated objects at this time is, of course, the crucial point on which the whole system relies.) As subsequent non-shift-causing utterances come in, their new information is added to the active focus space.

Usually, of course, speakers are not as helpful as in (5-4), and it is necessary to look for various clues to shifts in focus. For Grosz, the clues are definite NPs. If a definite NP from an utterance cannot be matched in focus, then this is a clue that the focus has shifted, and it is necessary to search for the new focus. If the antecedent of a definite NP is in the current implicit focus, this is a clue that a sub-task associated with this item is being opened. If the task structure is being followed, then the new focus will reflect the opening or closing of a sub-task.

Shifting cannot be done until a whole utterance is considered, as clues may conflict, or the meaning of the utterance may contraindicate the posited shift. In particular, recall that the task structure is only a guide, and does not define the dialogue structure absolutely. For example, the focus may shift to a problem associated with the current sub-task with a question like this:

(5-5) Should I use the box-end ratchet wrench to do that?

This does not imply a shift to the next sub-task requiring a box-end ratchet wrench (assuming that the current task doesn't require one) (cf Grosz 1977b:105).

We can see here that the problem of the circularity of language comprehension looms dangerously – to determine the focus one must resolve the references, and to resolve the references, one must know the focus. In Grosz's work, the strong constraints of the structure of task-oriented dialogues provide a toehold. Whether generalization to the case of discourse with other structures, or with no particular structure, is possible is unclear, as it may not be possible to determine so nicely what the knowledge associated with any new focus is. (See however my remarks in section 5.1.2 on the relationship between Grosz's work and that of Kantor, and section 5.5 on approaches which attempt to exploit local discourse structure.)

In addition, Grosz's mechanisms are limited in their ability to resolve inter-sentential and/or inference-requiring anaphora. The assumption that global focus of attention equals all and only possible referents (except where the focus shifts), while perhaps not unreasonable in task-oriented domains, is probably untrue in general. For example, could such mechanisms handle the 'table' examples of Chapter 4, excluding the table from focus when the second sentence of each of these texts is considered? Recall that local as well as global theme is involved (see section 5.1). Similarly, could the level of world knowledge and inference required by the 'sukiyaki' examples of Chapter 3 be integrated into the partitioned semantic net formalism? Could entities evoked by, but not explicit in, a text of only moderate structure be identified and instantiated in focus? Grosz did not address these issues (nor did she need to

5.2.3 *Maintaining focus*

for her immediate goals), but they would need to be resolved in any attempt to generalize her approach. (Some other related problems, including those of focus shifting, are discussed in Grosz (1978).)

Grosz's contribution was to demonstrate the role of discourse structure in the identification of theme, relevant world knowledge and the resolution of reference; we now turn to another system which aspires to similar goals, but in a more general context.

5.3. Focus in the PAL system and Sidner's theory

The PAL personal assistant program (Bullwinkle 1977a) is a system designed to accept natural language requests for scheduling activities. A typical request (from Bullwinkle 1977b:44) is:

(5-6) I want to schedule a meeting with Ira. It should be at 3 pm tomorrow. We can meet in Bruce's office.

The section of PAL that deals with discourse pragmatics and reference was developed by Candace Sidner [Bullwinkle] (Bullwinkle 1977b; Sidner 1978a). Like Grosz's system (see section 5.2), PAL attempts to find a focus of attention in its knowledge structures to use as a focus for reference resolution. Sidner sees the focus as equivalent to the discourse topic; in fact in Bullwinkle (1977b) the word *topic* is used instead of *focus*.

There are three major differences from Grosz's system:

- 1 PAL does not rely heavily on discourse structures.
- 2 Knowledge is represented in frames.
- 3 Focus selection and shifting are handled at a more superficial level.

I will discuss each difference in turn.

5.3.1. PAL's approach to discourse

Because a request to PAL need not have the rigid structure of one of Grosz's task-oriented dialogues, PAL does not use discourse structure to the same extent, instead relying on more general local cues. However, as we shall see below, in focus selection and shifting, Sidner was forced to use ad hoc rules based on observations of the typical requests to PAL.

5.3.2. The frame as focus

The representation of knowledge in PAL is based on the FRAME concept first introduced by Minsky (1975),⁸ and its implementation uses the FRL frame representation language (actually a dialect of LISP) developed by Roberts and Goldstein (1977a, 1977b; Goldstein and Roberts 1977).

In PAL, the frame corresponds to Grosz's focus space. Following Rosenberg's (1976, 1977) work on discourse structure and frames, the antecedent for a definite NP is first assumed to be either the frame itself, or one of its slots.⁹ So, for example, in (5-7):

(5-7) I want to have a meeting with Ross⁽¹⁾. It should be at three pm.
The location will be the department lounge. Please tell Ross⁽²⁾.

it refers to the MEETING frame (NOT to the text *a meeting*) which provides the context for the whole discourse; *the location* refers to the LOCATION slot that the MEETING frame presumably has (thus the CLOSELY ASSOCIATED WITH relation (see section 2.4.2) is easily handled), and Ross⁽²⁾ to the contents¹⁰ of the CO-MEETER slot, previously given as Ross.

If the antecedent cannot be found in the frame, it is assumed to be either outside the discourse or inferred. In (5-7), PAL would search its database to find referents for Ross⁽¹⁾ and *the department lounge*. Personal names are resolved with a special module that knows about the semantics of names (Bullwinkle 1977b:48). PAL carries out database searches for references like *the department lounge* apparently by searching a hierarchy of frames, looking at the frames in the slots of the current focus, and then in the slots of these frames, and so on (Sidner 1978a:211) though it is not apparent why this should usefully constrain the search in the above example.¹¹

⁸I will have to assume the reader is familiar with the basic concept of frames. Readers who require further background should read the section of Charniak (1976) on frames and/or Minsky's original paper (1975).

⁹In Sidner (1978b:91) it is claimed that a definite NP cannot refer to the focus if it contains more information than the focus. This is often true, but (2-100) is a counterexample to the complete generality of her assertion.

¹⁰Sidner only speaks of reference to slots (1978a:211), without saying whether she means the slot itself or its contents; it seems reasonable to assume, as I have done here, that she actually means both.

¹¹In fact there is no need in this particular example for a referent at all. The personal assistant need only treat *the department lounge* as a piece of text, presumably meaningful to both the speaker and Ross, denoting the meeting location. A human might do this when passing on a message they didn't understand:

(i) Ross asked me to tell you to meet him in the arboretum, whatever the heck that is.

On the other hand, an explicit antecedent WOULD be needed if PAL had been asked, say, to deliver some coffee to the meeting in the department lounge. Knowing when to be satisfied with ignorance is a difficult problem which Sidner does not consider, preferring the safe course of always requiring an antecedent.

5.3.2 The frame as focus

5.3.3. Focus selection

In PAL, the initial focus is the first NP following the VP of the first sentence of the discourse – usually, the object of the sentence – or, if there is no such NP, then the subject of that sentence. This is a short-cut method, which seems to be sufficient for requests to PAL, but which Sidner readily admits is inadequate for the general case (Sidner 1978a:209). I will briefly review some of the problems.

Charniak (1978) has shown that the frame-selection problem (which is here identical to the initial focus selection problem, since the focus is just the frame representing the theme of the discourse) is in fact extremely difficult, and is not in the most general case amenable to solution by either strictly top-down or bottom-up methods. Sidner's assumption that the relevant frame is given by an explicitly mentioned NP is also a source of trouble, even in the examples she quotes, such as these two (Sidner 1978b:92):

- (5-8) I was driving along the freeway the other day. Suddenly the engine began to make a funny noise.
- (5-9) I went to a new restaurant with Sam. The waitress was nasty. The food was great.

(Underlining indicates what Sidner claims is the focus.) In (5-8), Sidner posits a chain of inferences to get from *the engine* to the focus, the FREEWAY frame. This is more complex than is necessary; if the frame/focus were DRIVING (with its LOCATION slot containing the FREEWAY frame), then the path from the frame to *the engine* is shorter and the whole arrangement seems more natural. Thus we see that focus need not be based on an NP at all.

In (5-9), our problem is what to do with Sam, who could be referenced in a subsequent sentence. It is necessary to integrate Sam into the RESTAURANT frame/focus, since clearly he should not be considered external to the discourse and sought in the database. While the RESTAURANT frame may indeed contain a COMPANION slot for Sam to sit in, it is clear that the first sentence could have been *I went <anywhere at all> with Sam*, requiring that any frame referring to something occupying a location have a COMPANION slot. This is clearly undesirable. But the RESTAURANT frame IS involved in (5-9), otherwise *the waitress* and *the food* would be external to the discourse. A natural solution is that the frame/focus of (5-9) is actually the GOING-SOMEWHERE frame (with Sam in its COMPANION slot), containing the RESTAURANT frame in its PLACE slot, with both frames together taken as the focus. Sidner does not consider mechanisms for a multi-frame focus.

It is, of course, not always true that the frame/focus is explicit. Charniak (1978) points out that (5-10) is somehow sufficient to invoke the MAGICIAN frame:

- (5-10) The woman waved as the man on stage sawed her in half.

(See also Hirst (1982) for more on frame invocation problems.)

Focus shifting in PAL is restricted: the only shifts permitted are to and from sub-aspects of the present focus (Sidner 1978a:209). Old topics are stacked for possible later return. This is very similar to Grosz's open-focus hierarchy. It is unclear whether there is a predictive aspect to PAL's focus-shift mechanism,¹² but the basic idea seems to be that any new phrase in a sentence is picked as a potential new focus. If in a subsequent sentence an anaphoric reference is a semantically acceptable coreferent for that potential focus, then a shift to that focus is ipso facto indicated (Sidner 1978a:209). Presumably this check is done after a check of focus has failed, but before any database search. A potential focus has a limited life span, and is dropped if not shifted to by the end of the second sentence following the one in which it occurred.

An example (Sidner 1978a:209):

(5-11) I want to schedule a meeting with George, Jim, Steve and Mike. We can meet in my office. {It's kind of small, but the meeting won't last long anyway | It won't take more than 20 minutes}.

In the second sentence *my office* is identified as a potential focus, and *it*, in the first reading of the third sentence, as an acceptable coreferent to *my office* confirms the shift. In the second reading, *it* couldn't be *my office*, so no shift occurs. The acceptability decision is based on selectional and case-like restrictions.

While perhaps adequate for PAL, this mechanism is, of course, not sufficient for the general case, where a true shift, as opposed to an expansion upon a previously mentioned point, may occur. This is exemplified by many of the shifts in Grosz's task-oriented dialogues.

Another problem arising from this shift mechanism is that two different focus shifts may be indicated at the same time, but the mechanism has no way to choose between them. For example:

(5-12) Schedule a meeting of the Experimental Theology Research Group, and tell Ross Andrews about it too. I'd like him to hear about the decommunication work that they're doing.

Each of the underlined NPs in the first sentence would be picked as a potential focus. Since each is pronominally referenced in the second sentence, the mechanism would be confused as to where to shift the focus. (Presumably *Ross Andrews* would be the correct choice here.)

¹²On page 209 of Sidner (1978a) we are told: "Focus shifts cannot be predicted; they are detectable only after they occur". Yet on the following page, Sidner says: "Sentences appearing in mid-discourse are assumed to be about the focus until the coreference module predicts a focus shift. . . Once an implicit focus relation is established, the module can go onto [sic] predictions of focus shift". My interpretation of these remarks is that one cannot be certain that the next sentence will shift focus, but one CAN note when a shift MIGHT happen, requiring later checking to confirm or disconfirm the shift.

*I always get buggered by the
bottom-up approach.*

– “*Sydney J Hurtubise*”¹³

5.3.4. Sidner's general theory

In another paper (Sidner 1978b) Sidner describes a more general theory of focus whose relation to PAL is not explicitly stated. For example, for details of focus shifting one is simply referred to the section of Bullwinkle (1977b) on PAL's shift mechanism, which, as we saw, is inadequate for the general case. One can't tell if Sidner intends that PAL's mechanism be part of her general theory, or merely makes the reference as a stopgap.

Her theory is based on Grosz's system, but does not rely on a rigid discourse structure, nor does it suggest a knowledge representation for focus. However, Sidner does suggest (1978b:92) that a semantic association network should be involved as well. This would be used in determining CLOSELY ASSOCIATED WITH relations (Sidner 1978b:92), though she doesn't say how an acceptable closeness would be determined in the net. The net would be used instead of, or together with, the database search, the search starting from concepts closely related to the focus and working outwards. When a reference's relationship to the focus requires inference, this too would use the semantic net, though we are not told if this is attempted before, after, in parallel with or as part of the database search, nor exactly how it would be done.

Sidner is also concerned, in her general theory, with deciding whether or not a definite NP is generic. (Grosz did not attempt this, assuming all references to be specific.)¹⁴ Sidner gives some heuristics for determining whether a U-AMBIGUOUS NP – one that could be either generic or non-generic – has a preferred generic or non-generic reading. She then points out that those NPs whose head nouns match the focus usually have the same genericity as the focus, with which they are coreferential. She gives these examples (1978b:91):

¹³While presenting a paper at the first national conference of the Canadian Society for Computational Studies of Intelligence/Société canadienne pour études d'intelligence par ordinateur, on 26 August 1976.

¹⁴A SPECIFIC NP refers to a certain entity, a GENERIC NP to a class of entity, but via a single member of the class. For example, (i) shows specific NPs and (ii) a generic NP:

(i) When Ross returned to his car, the wheels were gone.

(ii) Today we will discuss rare marsupials. First let me tell you about the narbalek.

Note that the second sentence of (ii) has a generic reading in this context, but can be specific in a different context:

(iii) Ross gave Nadia a narbalek and a bandicoot. First let me tell you about the narbalek.

An NP may be ATTRIBUTIVE instead of either specific or generic – this usage need not concern us here.

(5-13) I'm going to tell you about the elephant⁽¹⁾. The elephant⁽²⁾ is the largest of the jungle mammals. He weighs over 3000 pounds.

(5-14) I sent George an elephant⁽³⁾ last year for a birthday present. The elephant⁽⁴⁾ likes potatoes for breakfast.

The underlined NPs are u-ambiguous without context. But since the focus of (5-13), *the elephant*⁽¹⁾, is generic, so are *the elephant*⁽²⁾ and *he*; the focus of (5-14), *an elephant*⁽³⁾, is specific, and therefore so is *the elephant*⁽⁴⁾. The focus can thus be used to u-disambiguate such NPs. Unfortunately there are counterexamples to this; Sidner's is (5-15):

(5-15) Mary got a ferret⁽¹⁾ for Christmas last year. The ferret⁽²⁾ is a very rare animal.

The ferret⁽²⁾ is so strongly generic that the specific focus, Mary's ferret, cannot override it, and *the ferret*⁽²⁾ therefore does not refer to the focus. Hence genericity must also be checked at the sentence level before testing NPs to see if they refer to the focus. In other words, there is a top-down/bottom-up conflict here. Sidner's solution is apparently to first check whether an NP is overwhelmingly generic at the sentence level; if not, only then is the genericity of the focus used. No threshold for overwhelming genericity is suggested.

Sidner's general theory has a more complex initial focus selection mechanism than PAL; she refers the reader to her forthcoming thesis (Sidner 1979) for details.

5.3.5. Conclusions

The shortcomings of Sidner's work are mainly attributable to two causes: her avoidance of relying on the highly constrained discourse structures that Grosz used, and the limited connectivity of frame systems, compared to Grosz's semantic nets. Recognizing the latter point, Sidner proposed the use of an association network in her general theory (1978b:87), though she does not say whether this should supplant or supplement other knowledge structures like PAL's frames. (Perhaps a synthesis, such as a network whose nodes are frames (cf McCalla 1977), is the answer.) With respect to the former point, perhaps Sidner's main contribution has been to show the difficulties and pitfalls that lie in wait for anyone attempting to generalize Grosz's work, even to the extent that PAL does.

5.3.5 Conclusions

5.4. Webber's formalism

In the preceding sections of this chapter, we saw approaches to focus that were mainly top-down in that they relied on a notion of theme and/or focus of attention to guide the selection of focus (although theme determination may have been bottom-up). An alternative approach has been suggested by Bonnie Lynn [Nash-]Webber (Nash-Webber and Reiter 1977; Webber 1978a, 1978b), wherein a set of rules is applied to a logical-form representation of the text to derive the set of entities that that text makes available for subsequent reference. Webber's formalism attacks problems caused by quantification, such as those we saw in (2-5)¹⁵ that have not otherwise been considered by workers in NLU.

I can only give the flavour of Webber's formalism here, and I shall have to assume some familiarity with logical forms. Readers who want more details should see her thesis (1978a); readers who find my exposition mystifying should not worry unduly – the fault is probably mine – but turn to the thesis for illumination.

In Webber's formalism, it is assumed that an input sentence is first converted to a parse tree, and then, by some semantic interpretation process, to an EXTENDED RESTRICTED-QUANTIFICATION PREDICATE CALCULUS REPRESENTATION. It is during this second conversion that anaphor resolution takes place. When the final representation, which we shall simply call a LOGICAL FORM, is complete, certain rules are applied to it to generate the set of referable entities and descriptions that the sentence evokes. Webber considers three types of antecedents – those for definite pronouns (IRAs), those for *one*-anaphora, and those for verb phrase ellipsis. Each has its own set of rules, at which we will briefly look.

5.4.1. Definite pronouns

The antecedents for definite pronouns are INVOKING DESCRIPTIONS (IDs), which are derived from the logical form representation of a sentence by a set of rules that attempt to take into account factors, such as NP definiteness or references to sets, that affect what antecedents are evoked by a text. There are six of these ID-rules;¹⁶ which one applies depends on the structural description of the logical form.

Here is one of Webber's examples (1978a:64):

(5-16) Wendy bought a crayon.

This has this representation:

¹⁵(2-5) Ross gave each girl a crayon. They used them to draw pictures of Daryl in the bath.

¹⁶Webber regards her rules only as a preliminary step towards a complete set which considers all relevant factors. She discusses some of the remaining problems, such as negation, in Webber (1978a:81-88).

(5-25) e_3 is $\text{maxset}(\lambda(u : \text{Peach})[(\exists x \in e_1)(\exists y \in e_2) . \text{Gave } x, y, u$
 & evoke (5-21), $y]$) z

"the set of peaches, each one of which is linked to (5-21) by virtue of some member of e_1 giving it to some member of e_2 ."

Although such rules could (in principle) be used to generate all IDs (focus elements) that a sentence evokes, Webber does not commit herself to such an approach, instead allowing for the possibility of generating IDs only when they are needed, depending on subsequent information such as speaker's perspective. She also suggests the possibility of "vague, temporary" IDs for interim use (1978a:67).

There is a problem here with intrasentential anaphora, since it is assumed that a sentence's anaphors are resolved before ID rules are applied to find what may be the antecedents necessary for that resolution. Webber proposes that known syntactic and selectional constraints may help in this conflict, but this is not always sufficient. For example:

(5-26) Mary bought each girl a cotton T-shirt, but none of them were the style de rigeur in high schools.

The IDs for both the set of girls and the set of T-shirts are needed to resolve *them*, but *them* needs to be resolved before the IDs are generated. In this particular example, the clear solution is to work a clause at a time rather than at a sentence level. However, this is not always an adequate solution, as (5-27) shows:

(5-27) The rebel students annoyed the teachers greatly, and by the end of the week none of the faculty were willing to go to their classes.

In this ambiguous sentence one possible antecedent for *their*, *the faculty*, occurs in the same clause as the anaphor. Thus neither strictly intraclausal nor strictly interclausal methods are appropriate. Webber is aware of this problem (1978a:48), and believes that it suffices that such information as is available be used to rule out impossible choices; the use of vague temporary IDs then allows the anaphor to be resolved.

5.4.2. One-anaphors

The second type of anaphor Webber discusses is the *ONE-ANAPHOR*.¹⁷ By this, she means an anaphor that refers to a description rather than a specific entity (see section 2.5). For example (Webber 1978a:97):

¹⁷I feel *one-anaphor* is a misleading (as well as clumsy) term, since a *one-anaphor* can be instantiated by *that*, *those*, *it*, or ϕ as well as *one*. Perhaps Webber's earlier term *descriptive anaphor* (Nash-Webber 1976) would have been better.

(5-28) Wendy didn't give either boy a green tie-dyed T-shirt, but she gave Sue a red one.

Here *one* is either *T-shirt* or *tie-dyed T-shirt*, but not *green tie-dyed T-shirt*.

Webber believes that the logical-form representation, as used above for deriving IDs, is an adequate representation from which such descriptions may be derived when needed by an appropriate reasoning procedure. She argues that this representation fulfils four desiderata:

- 1 It must retain the structure of noun phrases as a unit (so that, for example, in (5-28) *tie-dyed* remains connected to *T-shirt* to provide a single antecedent).
- 2 Yet it must allow decomposition of the description (so that, for example, in (5-28) *green* can be broken off *green tie-dyed T-shirt* when found inappropriate).
- 3 It should allow identification of word sense, to prevent inadvertent syllepsoid/zeugmoid interpretations (so that, for example, (5-29):

(5-29) *The ruler [i.e. head of state] picked one [i.e. a ruler, i.e. a measuring stick] up and measured the lamp.

can be flagged as anomalous).¹⁸

- 4 It must retain definite pronouns in both their resolved and unresolved forms (so that, for example, in (5-30) (after Webber 1978a:106):

(5-30) I compared Ross's behaviourist analysis of his mother with Daryel's gestalt one.

one is resolved as *analysis of Ross's mother*, not *analysis of Daryel's mother*, while in (5-31) (after Webber 1978a:106):

(5-31) Sue will pay up to seventy dollars for a dress she can wear without alteration, but Nadia refuses to pay more than fifty for one.

one is a dress that Nadia, not Sue, can wear without alteration).

Given this approach, the problem remains of determining when an anaphor is a *one*-anaphor and when it is a definite anaphor, as some pronouns, such as *it*, can be either. Webber offers some tentative suggestions:

- 1 *That* and *those* are *one*-anaphors if and only if they are followed by one or more NP postmodifiers (such as a prepositional phrase or relative clause).
- 2 An ellipsis can be used as a *one*-anaphor when preceded by an adjective but not followed by a postmodifier, or when preceded by a possessive, ordinal, comparative or superlative (with optional postmodifier). However, the problem of detecting the ellipsis in the first place remains, as structural

¹⁸See footnote 32 of Chapter 2.

ambiguities can arise (Webber 1978a:116).

- 3 It is problematic, but it seems to be a *one*-anaphor whenever followed by a postmodifier, and it requires as an antecedent a description of a unique entity in the discourse.

Webber asserts (1978a:111) that only recency, independent of discourse structure, controls the availability of descriptions as antecedents. I'm not sure that this is entirely correct. For example:

(5-32) ?Ross drank the wine on the table. Meanwhile Nadia and Sue played cards on another one next door.

(5-33) ?Ross moved the wine on the table to another one.

In each of these texts an attempt to reference a recent description with *one* is ill-formed, or at best marginal. That is, not all recent descriptions are in focus. Are, conversely, all referable descriptions textually recent? The answer is probably yes; I for one have not found any counterexamples.

Only descriptions explicitly present in the text are available as antecedents in the approach mentioned so far. What of implicit descriptions evoked by the text, which are also referable? Webber divides these into three categories, and gives suggestions on the handling of each (1978a:118-124):

1 Strained anaphora (see section 2.3.5). Webber suggests strained anaphora can occur with only a certain few words, and therefore can be handled by noting all such cases in the lexicon. I find this intellectually unsatisfying – I'm sure there is a general principle lurking about waiting to be discovered – but I have no better suggestions to offer.

2 References to IDs evoked by existential quantifiers. For example (after Webber 1978a:120):

(5-34) Nadia gave Ross some cotton T-shirts. The most expensive ϕ was too large, but the other ones fitted.

The referents in (5-34) are not just *cotton T-shirt(s)* but *cotton T-shirt(s) that Nadia gave Ross*. Two ways of deriving these are suggested: either (a) the *one*-anaphors could be treated as above, referring only to *cotton T-shirt(s)*, and these references are in turn treated as again anaphoric (cf section 2.4.2) and resolved as definite references to the ID for the T-shirts that Nadia gave Ross; or (b) the *one*-anaphors may be viewed as direct references to the ID. The latter has problems with negation¹⁹ and blurs the useful line between *one*- and definite anaphors; the former requires great care with determiners when checking whether a resolved *one*-anaphor has turned into a definite anaphor.

3 Abstraction of list elements. For example (Webber 1978a:122-123):

¹⁹*One*-anaphors can refer to descriptions of entities that don't exist in the discourse model and therefore don't have IDs. See Webber (1978a:121).

(5-35) I have in my cellar a '76 Beaujolais, a '71 Chateau Figeac, a '75 Durkheimer Feuerburg and a '75 Ockfener Bockstein. Shall we drink the German ones now and the others later?

(5-36) I know about Advent, Bose, AR and KLH, but about Japanese ones you'll have to ask Fred.

According to Webber, *ones* is *wines* in (5-35) and something like *speakers* or *speaker manufacturers* in (5-36). This sort of sentence varies in acceptability (I personally find (5-36) ill-formed) and Webber suggests that the poorer sentences are exactly those where the anaphor occurs in an indefinite NP, requiring an explicit abstraction on the list to be carried out for use as an antecedent, whereas in sentences such as (5-35) *one(s)* can be interpreted simply as *member(s) of the just-mentioned list*.²⁰

5.4.3. Verb phrase ellipsis

The third and last class of anaphor that Webber treats is verb phrase ellipsis (VPE) (in which she includes the pro-verb *to do*),²¹ extending Sag's (1976) theory of logical forms and VPE. A verb phrase may be elided if its logical form representation (written such that the predicate of the sentence applies to the subject) is identical to that of some preceding²² VP, called the ellipsis TRIGGER. (The ANTECEDENT is the deleted VP itself.) For example:

(5-37) Ross gave Nadia a book. Sue did ϕ too.
 $\lambda(s)$ [Gave, s, Nadia, book] Ross
 $\lambda(s)$ [Gave, s, Nadia, book] Sue

Webber proposes that a syntactic variant of her abovementioned representation is adequate for resolving VPE, discussing (1978a:129-149) the requirements that it must and does fulfil, including the problems caused by negation

²⁰In my idiolect such a sentence is ill-formed exactly when this simpler interpretation of *one(s)* is not possible. Webber believes that the additional requirement that the list be composed of names, not descriptions, is necessary, and thus does not like this example of hers (1978a:124):

(i) At the Paris zoo, Bruce saw a lion, a tiger, a giraffe, a hippopotamus and an elephant. It was feeding time, and the carnivorous ones were eating boeuf bourgignon, and the herbivorous ones, salad nipoise.

However, this is acceptable to me, and is amenable to the simpler interpretation. On the other hand, the list of animals in (i) is, in a very real sense, a list of names rather than descriptions. (Where is the dividing line between a name and a description?) It may therefore be that Webber's explanation is correct and that she has misconstrued her own example.

²¹Webber sees *to do* as a dummy verb sitting in the void left by a VPE, rather than as an anaphor in its own right.

²²Cataphoric VPE is also possible, but heavily restricted. Webber discusses it briefly (1978a:152).

5.4.3 Verb phrase ellipsis

and sloppy identity (see section 2.6).

The focus for VPE is then the set of all possible triggers in the logical form representation. Recency, with the additional constraints of sentence structure, voice, negation and tense, determines what is available as a trigger. When an ellipsis is detected, the appropriate trigger is sought; Webber discusses this and associated problems in (1978a:157-162). In particular, it is necessary to resolve VPE before definite pronouns, to avoid problems of missing antecedents (see footnote 59 of Chapter 2).

As Webber herself points out, this approach only works where the trigger is textually similar to the elided VP. But this is not always the case. Recall texts (2-16) and (2-17),²³ for example. This type of VPE requires inference and/or alternative ways of looking at the text; Webber makes some very tentative suggestions on how this might be handled (1978a:162-167).

5.4.4. Conclusions

It remains to discuss the strengths and weaknesses of Webber's approach, and she herself (in contradistinction to some other AI workers) is as quick to point out the latter as the former. The reader is therefore referred to her thesis (1978a) for this. However, I will make some global comments on the important aspects relevant here.

Webber's main contributions, as I see them, are as follows:

- 1 The focus problem is approached from the point of view of determining what an adequate representation would be, rather than trying to fit (to straitjacket?) focus into some pre-existing and perhaps arbitrarily chosen representation; and the criteria of adequacy for the representation are rigorously enumerated.
- 2 A formalism in which it is possible to compute focus elements as they are needed, rather than having them sitting round in advance (as in Grosz's (1977) system), perhaps never to be used, is provided (but compare my further remarks below).
- 3 Webber brings to NLU anaphora research the formality and rigour of logic, something that has been previously almost unseen.
- 4 Previously ignored problems of quantification are dealt with.
- 5 The formalism itself is an important contribution.

The shortcomings, as I see them, are as follows:

- 1 The formalism relies very much on antecedents being in the text. Entities evoked by, but not explicit in, the text cannot in general be adequately handled (contrary to Grosz's system).

²³(2-16) Nadia wants to climb Mt Everest, and Ross wants to tour Africa, but neither of them will ϕ because they are both too poor.

(2-17) Ross and Nadia wanted to dance together, but Nadia's mother said she couldn't ϕ .

- 2 The formalism is not related to discourse structure. So, for example, it contains nothing to discourage the use of *the table* as the antecedent in the 'table' examples of Chapter 4. It remains to be seen if discourse pragmatics can be adequately integrated with the formalism or otherwise accounted for in a system using the formalism.
- 3 Intrasentential and intraclausal anaphora are not adequately dealt with.
- 4 Webber does not relate her discussions of representational adequacy to currently popular knowledge representations. If frames, for example, are truly inadequate we would like to have some watertight proof of this before abandoning current NLU projects attempting to use frames.

You will have noticed that contribution 2 and shortcoming 1 are actually two sides of the same coin – it is static pre-available knowledge that allows non-textual entities to be easily found – and clearly a synthesis will be necessary here.

5.5. Discourse-cohesion approaches to anaphora resolution

Another approach to coreference resolution attempts to exploit local discourse cohesion, building a representation of the discourse with which references can be resolved. This approach has been taken by (inter alia) Klappholz and Lockman (again hereafter *K&L*) (1977; Lockman 1978). By using only cues to the discourse structure at the sentence level or lower, one avoids the need to search for referents in pre-determined dialogue models such as those of Grosz's task-oriented dialogues (see section 5.2), or rigidly predefined knowledge structures such as scripts (Schank and Abelson 1975, 1977) and frames (Minsky 1975), which K&L, for example, see as overweight structures that inflexibly dominate processing of text. K&L emphasize that the structure through which reference is resolved must be dynamically built up as the text is processed; frames or scripts could assist in this building, but cannot, however, be reliably used for reference resolution as deviations by the text from the pre-defined structure will cause errors.

The basis of this approach is that there is a strong interrelationship between coreference and the cohesive ties in a discourse that make it coherent. By determining what the cohesive ties in a discourse are, one can put each new sentence or clause, as it comes in, into the appropriate place in a growing structure that represents the discourse. This structure can then be used as a focus to search for coreference antecedents, since not only do coherently connected sentences tend to refer to the same things, but knowledge of the cohesion relation can provide additional reference resolution restraints. Hobbs (1978) in particular sees the problem of coreference resolution as being automatically solved in the process of discovering the coherence relations in a text. (An example of this will be given in section 5.5.2.) Conversely, it is frequently helpful or necessary to resolve coreference relations in order to discover the coherence relations. This is not a vicious circle, claims

5.5 *Discourse-cohesion approaches to anaphora resolution*

Hobbs, but a spiral staircase. (This helical approach to understanding also occurs elsewhere in artificial intelligence; compare for example Mackworth's (1978) Cycle of Perception.)

In our discussion below, we will cover four issues:

- 1 deciding on a set of possible coherence relations;
- 2 detecting them when they occur in a text;
- 3 using the coherence relations to build a focus structure; and
- 4 searching for referents in the structure.

5.5.1. Coherence relations

The first thing required by this approach is a complete and computable set of the coherence relations that may obtain between sentences and/or clauses. Various sets have been suggested by many people, including Eisenstadt (1976), Phillips (1977), Pitkin (1977a, 1977b), Hirst (1977b, 1978b), Lockman (1978), Hobbs (1978) and Reichman (1978a, 1978b).²⁴ None of these sets fulfil all desiderata; and while Halliday and Hasan (1976) provide an extensive analysis of cohesion, it does not fit within our computational framework of coherence relations, and those, such as Hobbs, Lockman, Eisenstadt and Hirst, who emphasize computability, provide small sets which cannot, I believe, capture all the semantic subtleties of discourse cohesion. Nevertheless, the works cited above undoubtedly serve as a useful starting point for development of this area.

To illustrate what a very preliminary set of cohesion relations could look like, I will briefly present a set abstracted from the various sets of Eisenstadt, Hirst, Hobbs, Lockman and Phillips (but not faithful to any one of these).

The set contains two basic classes of coherence relations: (a) expansion or elaboration on an entity, concept or event in the discourse, and (b) temporal continuation or time flow. Expansion includes relations like CONTRAST, CAUSE, EFFECT, SYLLOGISM, ELABORATION, PARALLEL and EXEMPLIFICATION. In the following examples, '▪' is used to indicate the point where the cohesive tie illustrated is acting:

- (5-38) [CONTRAST] The hoary marmot likes to be scratched behind the ears by its mate, ▪ while in the lesser dormouse, nuzzling is the primary behaviour promoting pair-bonding.
- (5-39) [CAUSE] Ross scratched his head furiously. ▪ The new Hoary Marmot™ shampoo that he used had made it itch unbearably.
- (5-40) [EFFECT] Ross pulled out the bottom module. ▪ The entire structure collapsed.

²⁴Reichman's coherence relations operate at paragraph level rather than sentence or clause level.

- (5-41) [SYLLOGISM] Nadia goes to the movies with Ross on Fridays. Today's Friday, ▪ so I guess she'll be going to the movies.
- (5-42) [ELABORATION] To gain access to the latch-housing, remove the control panel cover. ▪ Undo both screws and rock it gently until it snaps out from the mounting bracket.
- (5-43) [PARALLEL] Nearly all our best men are dead! Carlyle, Tennyson, Browning, George Eliot! – ▪ I'm not feeling very well myself!²⁵
- (5-44) [EXEMPLIFICATION] Many of our staff are keen amateur ornithologists. ▪ Nadia has written a book on the Canadian triller, and Daryel once missed a board meeting because he was high up a tree near Gundaroo, watching the hatching of some rare red-crested snipes.

(You may disagree with my classification of some of the relations above; the boundaries between categories are yet ill-defined, and it is to be expected that some people will find that their intuitive boundaries differ from mine.)

Temporal flow relations involve some continuation forwards or backwards over time:

- (5-45) VICTORIA – A suntanned Prince Charles arrived here Sunday afternoon, ▪ and was greeted with a big kiss by a pretty English au pair girl.²⁶
- (5-46) SAN JUAN, Puerto Rico – Travel officials tackled a major job here Sunday to find new accommodations for 650 passengers from the burned Italian cruise liner *Angelina Lauuro*.
▪ The vessel caught fire Friday while docked at Charlotte Amalie in the Virgin Islands, but most passengers were ashore at the time.²⁷

Temporal flow may be treated as a single relation, as Phillips, for example, does, or it may be subdivided, as Eisenstadt and Hirst do, into categories like TIME STEP, FLASHBACK, FLASHFORWARD, TIME EDIT, and so on. Certainly, time flow in a text may be quite contorted, as in (5-47) (from Hirst 1978b); "•" indicates a point where the direction of the time flow changes:

- (5-47) Slowly, hesitantly, Ross approached Nadia. ▪ He had waited for this moment for many days. ▪ Now he was going to say the words ▪ which he had agonized over ▪ and in the very room ▪ he had often dreamed about. ▪ He gazed lovingly at her soft green eyes.

It is not clear, however, to what extent an analysis of time flow is necessary for anaphor resolution. I suspect that relatively little is necessary – less than is required for other aspects of discourse understanding. Temporal anaphora

²⁵From: A lament [cartoon caption], *Punch, or the London charivari*, CIC, 1893, page 210.

²⁶From: *The Vancouver express*, 2 April 1979, page A1.

²⁷From: *The Vancouver express*, 2 April 1979, page A5.

5.5.1 Coherence relations

(see section 5.6.1) probably makes the strongest demands here, though the definitive set of temporal cohesion relations will probably be a superset of those actually required to resolve anaphors.

I see relations like those exemplified above as PRIMITIVES from which more complex relations could be built. For example, the relation between the two sentences of (5-40) above clearly involves FORWARD TIME STEP as well as EFFECT. I have hypothesized elsewhere (Hirst 1978b) the possibility of constructing a small set of discourse relations (with cardinality about twenty or less) from which more complex relations may be built up by simple combination, and, one hopes, in such a way that the effects of relation $R1+R2$ would be the sum of the individual effects of relations $R1$ and $R2$. Rules for permitted combinations would be needed; for example, FORWARD TIME STEP could combine with EFFECT, but not with BACKWARD TIME STEP.

What would the formal definition of a coherence relation be like? Here is Hobbs's (1978:11) definition of ELABORATION: Sentence $S1$ is an ELABORATION of sentence $S0$ if a proposition P follows from the assertions of both $S0$ and $S1$, but $S1$ contains a property of one of the elements of P that is not in $S0$.

5.5.2. An example of anaphor resolution using a coherence relation

It is appropriate at this stage to give an example of the use of coherence relations in the resolution of anaphors. I will present an outline of one of Hobbs's; for the fine details I have omitted, see Hobbs (1978:18-23). The text is this:

(5-48) John can open Bill's safe. He knows the combination.

We want an NLU system to recognize the cohesion relation operating here, namely ELABORATION, and identify *he* as John and *the combination* as that of Bill's safe. We assume that in the world knowledge the system has are various axioms and rules of inference dealing with such matters as what combinations of safes are and knowledge about doing things. Then, from the first sentence of (5-48), which we represent as (5-49):

(5-49) can (John, open (Bill's-safe))

(we omit the details of the representation of *Bill's safe*), we can infer:

(5-50) know (John, cause (do (John, *ACT*), open (Bill's-safe)))

"John knows an action *ACT* that he can do that will bring about the state in which Bill's-safe is open"

From the second sentence of (5-48), namely:

(5-51) know (*HE*, combination (*COMB*, *Y*))

"someone, *HE*, knows the combination *COMB* to something, *Y*"

we can infer, using knowledge about combinations:

(5-52) know (*HE*, cause (dial (*COMB*, *Y*), open (*Y*)))

"*HE* knows that by causing the dialling of *COMB* on *Y*, the state in which *Y* is open will be brought about"

Recognizing that (5-50) and (5-52) are nearly identical, and assuming that some coherence relation does hold, we can identify *HE* with John, *Y* with Bill's-safe, and the definition of the ELABORATION relation is satisfied. In the process, the required referents were found.

5.5.3. Lockman's contextual reference resolution algorithm

Given a set of discourse cohesion relations, how may they be computationally determined in the processing of a text and used to build a structure representing the discourse that can be used for reference resolution? Only Hobbs (1978) and Lockman (1978; Klappholz and Lockman 1977) seem to have considered these aspects of the problem, though Eisenstadt (1976) discusses some of the requirements in world knowledge and inference that would be required. In this section we look at Lockman's work; a full description of Hobbs's program was not available at the time of writing.

Lockman does not separate the three processes of recognizing cohesion, resolving references and building the representation of the discourse. Rather, as befits such interrelated processes, all three are carried out at the same time. His contextual reference resolution algorithm (CRRA) works as follows:

The structure to be built is a tree, initially null, each node of which is a sentence. As each new sentence comes in, the CRRA tries to find the right node of the tree to attach it to, starting at the leaf that is the previous sentence and working back up the tree in a specified search order (see below) until a connection is indicated. Lockman assumes the existence of a judgement mechanism which generates and tests hypotheses as to how the new sentence may be FEASIBLY CONNECTED to the node being tested. The first hypothesis whose likelihood exceeds a certain threshold is chosen.

The hypotheses consider both the coherence and the coreference relations that may obtain. Each member of the set of coherence relations is hypothesized, and for each one coreference relations between the conceptual tokens of the new sentence and tokens either in the node under consideration or nearby it in the tree. (The search for tokens goes back as far as necessary in the tree until suitable ones are found for all unfulfilled definite noun phrases.) The hypotheses are considered in parallel; if none are judged sufficiently likely, the next node or set of nodes will be considered for feasible

5.5.3 Lockman's contextual reference resolution algorithm

connection to the current sentence.

The search order is as follows: First the IMMEDIATE CONTEXT, the previous sentence, is tried. If no feasible connection is found, then the immediate ancestor of this node, and all its other descendents, are tried in parallel. If the algorithm is still unsuccessful, the immediate ancestor of the immediate ancestor, and the descendents thereof, are tried, and so on up the tree. If a test of several nodes in parallel yields more than one acceptable node, the one nearest the immediate context is chosen.

If the current sentence is not a simple sentence, it is not broken into clauses dealt with individually, but rather converted to a small sub-tree, reflecting the semantic relationship between the clauses. The conversion is based simply upon the structure of the parse tree of the sentence and uses a table look-up. One of the nodes is designated by the table look-up as the head node, and the sub-tree is attached to the pre-existing context, using the procedure described above, with the connection occurring at this node. Similarly one (or more) of the nodes is designated as the immediate context, the starting point for the next search. (The search will be conducted in parallel if there is more than one immediate context node.)

There are some possible problems with Lockman's approach. The first lies in the fact that the structure built grows without limit, and therefore searches in it could, in theory, run right through an enormous tree. Normally, of course, a feasible connection or desired referent will be found fairly quickly, close to the immediate context. However, should the judgement mechanism fail to spot the correct one, the algorithm may run wild, searching large areas of the structure needlessly and expensively, possibly lighting on a wrong referent or wrong node for attachment, with no indication that an error has occurred. In other words, Lockman's CRRA places much greater trust in the judgement mechanism than a system like Grosz's (1977) (see section 5.2) which constrains the referent search area – more trust than perhaps should be put in what will of needs be the most tentative and unreliable part of the system.

Secondly, I am worried about the syntax-based table look-up for sub-trees for complex sentences. On the one hand, it would be nice if it were correct, simplifying processing. On the other hand, I cannot but feel that it is an oversimplification, and that effects of discourse theme cannot reliably be handled like this. However, I have no counterexamples to give, and suggest that this question needs more investigation.

The third possible problem, and perhaps the most serious, concerns the order in which the search for a feasible connection takes place. Because the first hypothesis exceeding the likelihood threshold is selected, it is possible to miss an even better hypothesis further up the tree. In theory, this could be avoided by doing all tests in parallel, the winning hypothesis being judged on both likelihood and closeness to the immediate context. In practice, given the ever-growing context tree as discussed above, this would not be feasible, and some way to limit the search area would be needed.

The fourth problem lies in the judgement mechanism itself. Lockman frankly admits that the mechanism, incorporated as a black box in his algorithm, must have abilities far beyond those of present state-of-the-art

5.5.3 Lockman's contextual reference resolution algorithm

inference and judgement systems. The problem is that it is unwise to predicate too much on the nature of this unbuilt black box, as we do not know yet if its input-output behaviour could be as Lockman posits. It may well be that to perform as required, the mechanism will need access to information such as the sentence following the current one (in effect, the ability to delay a decision), or more information about the previous context than the CRRA retains or ever determines; in fact, it may need an entirely different discourse structure representation from the tree being built. In other words, while it is fine in theory to design a reference resolver round a black box, in practice it may be computationally more economical to design the reference resolver round a knowledge of how the black box actually works, exploiting that mechanism, rather than straitjacketing the judgement module into its pre-defined cabinet; thus Lockman's work may be premature.

None of these problems are insurmountable. However it is perhaps a little unfortunate that Lockman's work offers little of immediate use for NLU systems of the present day.

5.5.4. Conclusion

Clearly, much work remains to be done if the coherence/cohesion paradigm of NLU is to be viable. Almost all aspects need refinement. However, it is an intuitively appealing paradigm, and it will be interesting to see if it can be developed into functioning NLU systems.

5.6. Non-noun-phrase focusing

The theories and approaches discussed heretofore in this chapter have been almost exclusively concerned with anaphors whose antecedents are NPs or other noun-like entities in consciousness, and indeed this is where most of the interesting problems lie. However, as we saw in Chapter 2, there are many other kinds of anaphor, and in this section I would like to describe the focus that temporal and locative anaphors require. These are simpler than the nominal case, and I present what I believe to be a complete theory (i.e. one which accounts for all cases).²⁸

²⁸A note on methodology:

In what I say below, I will make assertions like the following:

- (i) Linguistic phenomenon X occurs in English in exactly n ways: X_1, X_2, \dots, X_n .
- (ii) Linguistic phenomenon Y cannot occur in English.

These assertions will not be proved, in the sense that a mathematical or scientific assertion might be proved, for they cannot be. So, when I say (i) or (ii), what I really mean is this:

5.6 Non-noun-phrase focusing

*Rush on into the Aramis counter . . . now!
Discover Aramis 900,
the revolutionary grooming system for men.
Our trained Aramis consultant
will take you through the 900
systems programmer first,
after you receive a complementary
bottle of herbal after shave.²⁹*

5.6.1. The focus of temporal anaphors

Linguists have spent considerable time analyzing time and tense, and in recent years a few AI workers have examined the problems of computer understanding and representation of temporal concepts and temporal reference in natural language (Bruce 1972; Cohen 1976; Kahn and Gorry 1977; Sondheimer 1977a, 1977b). Strangely, AI workers have not considered temporal anaphora. My discussion below will assume the availability of an understander for non-anaphoric temporal references. I will show that temporal anaphors – the temporally

(iii)Although I've thought about it quite a bit, neither I, in my capacity as a native speaker of Australian English, nor anyone else I've asked (if any), can come up with an example of well-formed English text in which X_p ($p > n$) or Y occurs.

It is possible, therefore, that X_p ($p > n$) or Y may in fact occur in English, perhaps even rampantly – the language after all is infinite – but has managed to avoid my investigations. Maybe you, faithful reader, can easily come up with an example of X_p or Y . If so, I would be interested in seeing it.

The problem here is that of the "boundary of language". Wilks (1975c) expresses the situation well:

"Suppose that tomorrow someone produces what appears to be the complete AI understanding system, including of course all the right inference rules to resolve all the pronoun references in English. We know in advance that many ingenious and industrious people would immediately sit down and think up examples of perfectly acceptable texts that were not covered by those rules. We know they would be able to do this, just as surely as we know that if someone were to show us a boundary line to the universe and say 'you cannot step over this', we would promptly do so.

Do not misunderstand my point here: it is not that I would consider the one who offered the rule system as refuted by such an example, particularly if the latter took time and ingenuity to construct. On the contrary, it is the counterexample methodology that is refuted."

Because language is inherently infinite, one cannot prove the non-occurrence of X_p ($p > n$) or Y by enumeration of all possible sentences. And, a fortiori, it is claimed by some (such as Wilks 1971, 1973a, 1975c) that a natural language cannot even be understood or generated by a finite set of rules; that almost ANYTHING can be understood by a human's language system, provided it is accompanied by enough context or explanation. Thus a language understanding system cannot be refuted on the basis of a counterexample, provided its level of performance is by some criterion adequate, for a counterexample could be generated for ANY system we could ever possibly construct; and therefore we need special rules and recovery mechanisms to handle these counterexamples. While I am not convinced that this view is entirely correct (I discuss it further in Hirst (1976a)), it is not unappealing. What it means to us for the present is that the method of argument expressed in (iii) is the best we can do here.

²⁹Advertisement for David Jones' department store in: *The Canberra times*, 21 June 1977, page 1. Spelling, punctuation and temporal location are as supplied.

5.6.1 The focus of temporal anaphors

relative phrases and certain uses of the word *then* that we saw in section 2.3.11 – refer to the "temporal location" of the preceding text, and that discourse structure and topic have little to do with such anaphors.

By the TEMPORAL LOCATION of a text, I simply mean the time at which the actions being described take place. This time may be specified explicitly, as in (5-53), or not, as in (5-54):

(5-53) After dinner, Ross retired to the bathroom with a copy of *Time*, while Nadia and Sue played cribbage. [after dinner]

(5-54) Nadia dropped the orange down the chute, fervently hoping for a miracle. [the time when Nadia, while hoping fervently for a miracle, dropped the orange down the chute]

The text in brackets after each example represents its temporal location.

Not all text has a temporal location. Some present-tense sentences are effectively tenseless in that they refer to "all eternity"; this case occurs, for example, when discussing abstract ideas, as in (5-55):

(5-55) Some present-tense sentences are effectively tenseless in that they refer to "all eternity"; this case occurs, for example, when discussing abstract ideas, as in (5-55).

Clearly, detecting tenselessness requires inference on the meaning of the text.³⁰ Tenseless texts do not, in general, involve temporal anaphors, except when describing repeated actions over time:

(5-56) On Saturdays at the Enver Hoxha Christian Gospel Commune, we always follow the same inspiring schedule. Reveille is sounded at six am, and the residents eat a hearty breakfast of hash-brown potato peels. The next two hours are spent in quiet meditation and prayer, and it is then that glossolalia sometimes occurs.

The referent of any temporal anaphor is always the most recent temporal location of the text. For example, in (5-56) the antecedent of *the next two hours* is the time the residents have breakfast, and of *then* is the two hours of meditation. I have been unable to construct any well-formed text which

³⁰Some languages allow a lexical disambiguation. For example, in Spanish the verb *to be* is *ser* if tenseless and *estar* if not; compare (i) and (ii):

- (i) Soy australiano. [I am an Australian.]
- (ii) Estoy enfermo. [I am sick.]

³¹One possible exception occurs when two times are contrasted as in (i):

- (i) Surely their plane is more likely to arrive on Tuesday than on Wednesday. If we want to meet them, we should go to the airport THEN.

This sentence, in which *then* is stressed and intended to be temporally anaphoric, was acceptable only to a small proportion of informants, who understood *then* as meaning Tuesday. (There was no general consensus among informants as to whether or not (i) was either gram-

5.6.1 The focus of temporal anaphors

violates this general rule.³¹ Temporal cataphors are not possible.³²

The problem then becomes one of establishing a temporal location for the text. This is one aspect of the problem that Bruce, Cohen, and Kahn and Gorry, in the work cited above, approached, and it is not appropriate to discuss it here – the interested reader should see the work mentioned – except for two points:

First, time tends to move forward in the discourse, as in this example:

(5-57) Nadia filled the kettle, put it on the stove, and busied herself with the task of icing the cake. Suddenly, the telephone rang.

Although there are no explicit indications in the text, when reading it we have no trouble in deciding that the four events described took place one after the other in this order:

- 1 Nadia fills the kettle.
- 2 Nadia puts the kettle on the stove.
- 3 Nadia commences icing the cake.
- 4 The telephone rings.

The assumption of discourse cohesion implies further that these events took place contiguously (when viewed at a certain level of detail). This is the default case, and variations from it must be explicitly flagged.³³ This means that the temporal location is constantly changing in text. Thus in (5-56), the referent of *the next two hours* is not six am precisely, but six am plus the time taken in breakfast plus some certain amount of time taken in relevant overheads (like getting out of bed). (Kahn and Gorry attempt to handle the natural inexactitude of temporal reference with an explicit "FUZZ" element in their representation.)

matical or meaningful. When I first tried it without the phrase *if we want to meet them*, some informants understood the referent to be Wednesday and the intent of the speaker to be AVOIDING meeting the plane.) This could be another example of a case in which stress on an anaphor is to be interpreted as meaning *the intended referent is not the one this word would normally have* (see section 7.1 on the effects of stress and intonation).

³²In Hirst (1978b) I described (i) as temporally cataphoric (and, a fortiori, as a prototype of the only possible temporal cataphor):

(i) #It was then, when Sue had given up all hope, that it began to rain fish.

I no longer believe this to be cataphoric. Rather, *then* here refers to the temporal location of the previous text, and the embedded clause is an expansion on that same temporal location rather than a cataphoric referent for *then*. When presented without preceding text, as it is here, (i) is not coherent, as it presumes a previous temporal context. This could be acceptable as a literary device at the start of a story (cf footnote 5 of Chapter 4).

³³If variations from the default are not flagged, the result is ill-formed; hence (i) sounds strange:

(i) #I wanna hold you till I die,
Till we both break down and cry.

[From: Hill, Dan. Sometimes when we touch. On: Hill, Dan. *Longer fuse*. LP recording, GRT 9230-1073.]

(One informant told me that they interpreted *die* metaphorically, and thereby restored forward sequential ordering to (i).)

5.6.1 The focus of temporal anaphors

Second, topic is relevant to temporal anaphora only insofar as it affects temporal location; a new topic will usually have a new temporal location. But sometimes a temporal anaphor will explicitly refer across a topic shift to establish the new location by relating it to that of the previous topic.

5.6.2. The focus of locative anaphors

The anaphor *there* and locative relations exactly parallel *then* and temporal relations in that they refer to what we shall (ambiguously) call a text's PHYSICAL LOCATION.³⁴ An example:

(5-58) The Church of Scientology met in a secret room behind the local Colonel Sanders' chicken stand. Sue had her first dianetic experience there⁽¹⁾. Across the street was a McDonald's where The Church Of God The Utterly Indifferent had their meetings, and Ross went there⁽²⁾ instead, because of the free Big Macs they gave to recent converts.

The referent of *there*⁽¹⁾ is the secret room behind the local Colonel Sanders stand, and the referent of *across the street* is either the secret room or the chicken stand – there is no semantic difference.³⁵ The McDonald's is the referent of *there*⁽²⁾.

Determining a text's physical location is quite a different task from finding its temporal location, as there is no locative equivalent to tense in English (nor in any other language, as far as I am aware), nor does text automatically move through space as it does time. Determining physical location therefore relies solely on understanding locative references in the text. A complicating factor in doing this is that a text may have a separate *here*-location – the place where the speaker/writer is producing the text. This requires understanding the text to the extent of being able to determine whether a locative reference applies to the first person or not. For example, in (5-59):

(5-59) Ross is in Canberra, while I am in Vancouver. In July it is warmer here than there.

³⁴Also paralleling temporal reference are the problematic contrastive usage and the impossibility of locative cataphora. Texts (i) and (ii) correspond exactly to the examples in footnotes 31 and 32:

(i) Surely they are more likely to go to Spuzzum than Vancouver. We should wait for them THERE.

(ii) It was there, where Sue had given up all hope, that the pile of dead fish lay rotting.

³⁵This suggests the possibility of a similar text in which there IS a semantic difference, and hence whose physical location is not uniquely determined. I have not, however, found a well-formed example of this.

5.6.2 The focus of locative anaphors

one must be able to work out that *here* is Vancouver and *there* is Canberra.³⁶

³⁶Text also has a *now*-location in time which parallels its *here*-location, and which an NLU system may have to distinguish from other temporal locations in the text.