COGNITIVE SCIENCE 3, 67–90 (1979)

Coherence and Coreference*

JERRY R. HOBBS

SRI International
Menlo Park, CA

Coherence in conversations and in texts can be partially characterized by a set of coherence relations, motivated ultimately by the speaker's or writer's need to be understood. In this paper, formal definitions are given for several coherence relations, based on the operations of an inference system; that is, the relations between successive portions of a discourse are characterized in terms of the inferences that can be drawn from each. In analyzing a discourse, it is frequently the case that we would recognize it as coherent, in that it would satisfy the formal definition of some coherence relation, if only we could assume certain noun phrases to be coreferential. In such cases, we will simply assume the identity of the entities referred to, in what might be called a "petty conversational implicature," thereby solving the coherence and coreference problems simultaneously. Three examples of different kinds of reference problems are presented. In each, it is shown how the coherence of the discourse can be recognized, and how the reference problems are solved, almost as a by-product, by means of these petty conversational implicatures.

1. INTRODUCTION

Successive utterances in coherent discourse refer to the same entities. The common explanation for this is that the discourse is coherent *because* successive utterances are "about" the same entities. But this does not seem to stand up. The text

> John took a train from Paris to Istanbul. He likes spinach.

is not coherent, even though "he" can refer only to John. At this point the reader may object, "Well, maybe the French spinach crop failed and Turkey is the only country. . . ." But the very fact that one is driven to such explanations indicates that some desire for coherence is operating, which is deeper than the notion of a discourse just being "about" some set of entities.

*The research reported in this paper was supported by the National Science Foundation under Grant No. MCS76-22004 and by the Advanced Research Projects Agency under Contract N00039-78-C-0060 administered through the Naval Electronic Systems Command.

In this paper I would like to turn the picture upside down. I will present an independent characterization of coherence, motivated ultimately by the need of speakers to be understood. I suggest that the sense we have that a discourse is "about" some entity or set of entities is frequently just the conscious trace of the deeper processes of coherence. In Section 2, certain coherence relations that hold between portions of a discourse are defined with computable precision in the framework of the inference component of a language processor. Viewed from above, from the Olympian vantage point of an investigator studying a paragraph or transcript, the relations give structure to a discourse. From the point of view of a speaker just uttering a sentence, the relations correspond to coherent continuation moves he can make, i.e., the means he has of continuing the discourse in a relevant way. The solutions to many problems of reference and coreference simply "fall out" in the course of recognizing the coherence relations. I discuss why this should be so, and in Section 3 present three examples in which it happens. These examples illustrate the close connection between coherence and the resolution of anaphora, in which coherence plays the dominant role.

2. CHARACTERIZING COHERENCE

2.1. Requirements for a Theory of Coherence

A number of linguists have investigated the relations that link clauses, sentences, or larger portions of discourse to each other. These have variously been called "rhetorical predicates" (Grimes, 1975), "conjuctive relations" (Halliday & Hasan, 1976), "paragraph types" (Longacre, 1977), and "sequiturity relations" (Fillmore, 1974). In this paper, I shall call them "coherence relations," or, where context allows, simply "relations." Typically, one studying these relations simply lists them, usually in the form of a taxonomy, and gives some examples. They are frequently correlated with various conjunctions, but otherwise there is no attempt to go beyond the intuitive characterization toward formal definitions.

The difficulty for traditional linguists in formalizing the study of coherence in an illuminating way has been that to deal seriously with discourse, one must deal with the information it conveys and the knowledge that the listener or reader brings to bear in understanding it. These can be of an arbitrarily detailed nature. Work in artificial intelligence, especially on inference systems (e.g., Rieger, 1974), now allows us to begin to construct a theory of coherence, for the representation and use of knowledge is precisely what AI is all about.

In Section 2.2, I describe briefly the basic design of an inference system for natural language processing. In Section 2.3 certain coherence relations are listed and given very abstract but computable formal definitions in terms of primitive operations of the inference system. The inferencing operations can establish the relations with more or less "difficulty," as described below. It is the claim of this theory that a relatively small number of coherence relations occur in coherent English discourse and together they define coherence in the following sense: If a text strikes one intuitively as coherent, then coherence relations can be found linking its various parts. More precisely, a text will strike one as coherent to a degree that varies inversely with the degree of "difficulty" the inferencing operations have in recognizing *some* coherence relation. Coherence thus plays a role beyond sentence boundaries analogous to the role played by grammaticality within sentences. It is the mortar with which extended discourse is constructed.

If such a theory is to be convincing, it should satisfy three requirements. We should see why discourse is coherent in the first place, what other problems are solved by recognizing coherence, and how coherence can be recognized.

First, we should be able to explain the function of each of the coherence relations. Out of the various possible orders in which a collection of ideas can be communicated, why is one particular organization chosen over another? I will attempt to answer this question in part by appealing to the speaker's goal of communicating his ideas via the imperfect medium of language, to a listener operating under certain processing constraints. The speaker seeks to have the listener understand him—that is, draw the right inferences and arrive at the correct interpretation of what he says. He seeks to ease the processing load on the listener by structuring his message in a way that will enable finding the right inferences quickly. He seeks to exercise control over the significance that the listener attributes to his utterances, for people tend to generalize from what they learn, and one role the coherence relations play is to allow the speaker to promote or inhibit these generalizations. As each of the coherence relations is introduced, I will attempt to show how it aids some or all of these goals.

All this seems to assume one speaker has control over the organization of the discourse, but this is not necessarily the case. In a conversation, all the participants interact in ways that serve these goals, probing when they don't understand, helping each other express their thoughts, implicitly or explicitly proposing generalizations, as they work together in the creation of a single meaning. This suggests correspondences between the coherence relations used by a single speaker or writer and the coherent moves in conversation. Some of these correspondences are pointed out below.

The second requirement is that the cohesive relations studied by Halliday and Hasan (1976)—identity, similarity and subpart relations between entities referred to in different sentences—can be seen as deriving from the coherence relations. That is, a theory of coherence should answer what is a rather surprising question to ask in the first place—why should successive sentences talk about the same things? The answer is built into the coherence relations, for they all depend on the ways in which information and entities are shared by the sentences they link. The computational corollary of this is that many cases of coreference beyond sentence boundaries are resolved as a by-product of recognizing coherence. Examples of this have accumulated; three are given in Section 3.

The final requirement, and what distinguishes this effort from previous, descriptive characterizations of coherence, is that the relations must be computable. The next two sections attempt to point a way toward this goal.

2.2. The Inference Component

The typical inference system[1] has four aspects—data, representation, operations, and control. "Data" refers to the knowledge available to the system; in a natural language processing system the enormous amounts of world knowledge that must be accessed in understanding the most ordinary texts. "Representation" refers to the formats in which the knowledge is stored. "Operations" refers to the procedures that work on the represented data. "Control" refers to the choice of which operations apply and the order in which they apply.

These aspects are probably inseparable in an AI theory of language use. Nevertheless, in this paper my aim is to concentrate on the *operations* that recognize coherence. I will try to deal with the essentials of the other three aspects in a quick and graceful manner, but where this is impossible, grace is sacrificed first. More details are discussed in Hobbs (1976b). It is convenient to discuss "data" last.

Representation. The representation scheme is a kind of production system. Thinking of it as predicate calculus may help if not pushed too far.[2] I assume a number of *predicates*—e.g., can, open, safe, own, find, . . .—corresponding roughly to English words, and an arbitrary number of *entities*—e.g., J, B, S, . . .—which have no semantic content but are used to keep track of reference. A *proposition* is formed by applying a predicate to one or more entities or other propositions as *arguments*—e.g., can(J,open(J,S)) ("J can open S"), safe(S) ("S is a safe"), own(B,S) ("B owns S"). The predicate and arguments of a proposition will be referred to as its *elements*. The *properties* of an entity are all the propositions in which the entity occurs as an argument.

It is assumed that each successive clause in a *text* is made available to and is operated on in turn by the inference component. Each clause is in the form of a collection of propositions. At least one proposition in each clause is marked as *asserted*, or is the *assertion* of the clause. For example, in

John can open Bill's safe.

the proposition "can(J,open(J,S))" is asserted, while "safe(S)" and "own(B,S)" are not. This form is produced by a syntactic "front end" (cf.

[1]Most of what is described here is embodied in a working computer program.

[2]In particular, whereas in predicate calculus, one may apply modus ponens freely to construct chains of inference of arbitrary length, in this inference system, what chains of inference are constructed is placed under the strict "higher" control of the operations. It is for this reason (and for other reasons beyond the scope of this paper to discuss) that I have avoided adopting wholesale the form and terminology of predicate calculus.

Hobbs & Grishman, 1976; Hobbs, 1976b); I will not discuss the important issues of how such a "front end" must interact with the inference component. This representation is intended to be fairly close to the surface, and should be viewed primarily as a way of handing some of the hard problems of language processing over to the inference component, where they belong.

The inference component also has available to it a large number of rules, or *axioms*, which encode the system's normally true, commonly known lexical and world knowledge. These are of the form

antecedent → consequent

where both the antecedent and consequent are sets of propositions with variables in place of entities as arguments. If instances of all the propositions in the antecedent occur in the text, and *if some operation determines the axiom to be appropriate*, then an instantiation of the consequent is added to the text. If a variable in the antecedent is matched with some object in the text, all occurrences of that variable in the consequent are instantiated as the same object. If a variable occurs in the consequent but not in the antecedent, a new entity is posited in the text. Thinking of "→" as implication is helpful in understanding the intended semantic content of the axioms, but is dangerous if carried too far in formal manipulations.

Axioms likely to be used in a natural language processing system encode superset relations such as

safe(x) → container(x)
("A safe is a container");

common world knowledge facts such as

safe(x) → combination(y,x)
("A safe has a combination");

and lexical decompositions such as

find(x,y) → come-about(know(x,at(y,z)))
("If x finds y, then it comes about that x knows that y is at some point z.")

The collection of axioms is intended to represent those things a speaker of English generally knows and can expect his listener to know. The axioms may not always be true, but we leave to the operations the decision as to whether to apply them; hence the caveat italicized above. A relation, called "follows-from," between propositions, or more properly sets of propositions, is defined as the inverse of the reflexive transitive closure of "→."

Operations. The text is processed by applying a number of operations to it in parallel, for such things as interpreting general words in context (or determining word sense), resolving anaphora, determining illocutionary force, and recognizing coherence. The operations work by attempting to construct *chains of inference* out of the axioms, satisfying certain demands. Only the operation for recognizing coherence will be described here. It attempts to construct chains of inference satisfying definitions like those in Section 2.3. We will see in Section 3 how the chains of inference used in recognizing coherence are also used by other operations.

Control. It is assumed that the axioms have associated with them some measure of *salience* to the text and task at hand. The basic control regime for the inferencing process is that the order of search for chains of inference depends on this salience and on the length of the chains of inference. This order gives a measure of the "difficulty" the system has in constructing the chains. That means that the relation "follows-from" is really a matter of degree, as are those things defined in terms of "follows-from," including coherence.[3]

Data. For the definitions of the coherence relations it will not be necessary to assume anything about the axioms the inference component has available. In recognizing a particular instance of any coherence relation, we will of course have to assume a number of very specific axioms. To control this, we will for the time being simply insist that the axioms be plausible and have the appearance of general applicability. They should not look as if they were cooked up to handle the example in question. Ultimately, such investigations will have to be integrated with an overall theory of the knowledge base. But it is likely that one of the chief criteria we will want to use in deciding what to include in our collections of lexical and world knowledge, will be that the knowledge base mesh well with the theory of coherence.

2.3. Some Coherence Relations Defined

In Hobbs (1979), I describe the typical discourse situation and show how various classes of coherence relations arise out of problems that face a speaker in this situation in trying to be understood by a listener, or in trying to "make sense" to him.[4] One such class is comprised of relations that can be said to expand the text in place and are generally used to manipulate in some way the inference processes of the listener. Three relations in this class—Elaboration, Parallel, and Contrast—are discussed in this paper. They point to some of the complex ways in which the information implicit in sentences overlaps and interacts. They each

[3]One way of implementing a measure of salience is described in Hobbs, 1976b. I will not elaborate on it here. In this paper, to be frank, salience functions principally as an oracle, allowing me to assume the correct chain of inference has been chosen where several are possible.

[4]Hobbs (1979) contains my current best guess of a complete list of the coherence relations. Earlier versions can be found in Hobbs (1974, 1976b). Taxonomies of coherence relations have also been presented and illustrated by Isenberg (1971), Grimes (1975), and Longacre (1976, 1977), but they do not say how the relations would be recognized computationally. Phillips (1975, 1977a, 1977b) also discusses coherence relations and the highly detailed world knowledge required to process them.

link two segments of discourse that say almost the same thing, and they can be differentiated by the ways in which the second segment fails to say the same thing as the first.

In what follows, a formal definition will be given for each coherence relation, together with a fairly straightforward example and a brief indication of the chains of inference involved in recognizing the relation. I then suggest how the relation might help overcome some of the processing obstacles to communication.

Certain portions of a discourse will be designated *sentential units*, which are defined recursively as follows: A clause is a sentential unit. (Recall that clauses, and thus sentential units, are sets of propositions.) If some coherence relation links two sentential units, the union of the sentential units is itself a sentential unit. If a proposition is *asserted* in either of the original two sentential units, it is *asserted* in the union. Each of the relations is defined in terms of what is implied by the asserted proposition or propositions in each sentential unit. This reflects the notion that coherence rests on the "new" information conveyed by sentences.

In each of the definitions, S1 refers to the sentential unit currently being processed, S0 to a previous one. "Sentence" will frequently be used for "sentential unit."

For expository reasons, I have defined the relations as though they were an all or none matter. But it should be kept in mind that, just as "follows-from" is a matter of degree, a particular coherence relation holds between two sentential units to a greater or lesser degree, depending ultimately on the salience of the axioms used to establish the relation.

These definitions should be viewed as first attempts. Where they err, it is most likely to be toward too great a generality, and the appropriate ways to constrain them further is an important problem for future research.

Elaboration. S1 is an Elaboration of 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).

At a sufficiently deep level the two sentences say the same thing. In the typical case, new information is conveyed by the second sentence, since there must be some reason for saying it again. This is why I have called the relation *Elaboration* rather than *Paraphrase*. However, I also mean to include under this heading such trivial moves as pure repetitions, repairs, tag questions, and the like. Hence, the parentheses around the second clause in the definition. In addition, there are examples in which we can infer a proposition P from the first sentence, and in the second sentence, often for clarification, P is stated explicitly. The example of Section 3.2 below is such a case.

An example of an Elaboration from a set of directions is

> Go down Washington Street. Just follow Washington Street three blocks to Adams Street.

It is important that anyone trying to follow these directions recognize the second sentence as an Elaboration and not as the next instruction. The pattern is recognized by inferring a "going" from "follow" and matching the paths—Washington Street—from the two sentences. Then "to Adams Street" elaborates on the unstated end point of the "going" in the first sentence, and "three blocks" adds measure to its path.

One function of Elaboration is obviously to overcome misunderstanding or lack of understanding. In procedural texts, when a sentence is insufficiently informative to determine the corresponding action, the reader or listener looks for an Elaboration next, and frequently finds it.[5] This is seen in the above example, and also in the following example from an algorithm description:

Initialize. Set the stack pointer to zero, and set link variable P to ROOT. (1)

From "Initialize" alone we cannot generate adequate code.

But this raises an interesting point. Example (1) comes from a published text (Knuth, 1973), so the first sentence can't be a mistake that is corrected by the second. Why should the first sentence appear at all, if it can't lead to code? This suggests another function of Elaboration—it enriches the understanding of the listener by expressing the same thought from a different perspective. In algorithm descriptions, the first sentence typically describes the action in terms of the overall flow of control and the purposes of the algorithm. The second sentence describes it in terms of code. A single clause in English cannot easily support more than one point of view.

This pattern also occurs in conversational exchanges in modified form. First of all, a question-answer sequence can be viewed as a kind of Elaboration. To see this, we must extend our formalism slightly by introducing a question-mark operator, "?," which can be applied to propositions—$p(A,B)?$—to indicate yes–no questions. It can also be applied to entity symbols—$?X$—if, in a wh-question, it is the identity of an entity which is being questioned. Finally we will let the predicate symbol "$?p$" indicate a request for a property of its argument. We also introduce several rules for manipulating this operator. One such rule says that the question "Is x identical with y, where p is true of y?" 'implies' the question "Is p true of x?"

$$[x=y?,\ p(y)] \rightarrow p(x)? \tag{2}$$

Another rule says that the question "Is p true of x?" 'implies' a request for a property of x.

$$p(x)? \rightarrow ?p(x)$$

Then a question-answer sequence, such as

A: Who bought the dog?

B: The boy bought the dog.

[5]I am indebted to William Mann for pointing this out to me.

would be represented (ignoring tense and articles)

A: buy(?X,D), dog(D)

B: buy(b,D), boy(b),dog (D).

The recognition that B's response is an answer to A's question is just the recognition that b=?X and that because of the proposition "boy(b)," B's response Elaborates A's question in the required way.

Another variety of Elaboration is a Request for Elaboration:

A: He bought the dog.

B: Who bought the dog?

or simply,

B: Who?

These would be represented

A: buy(X,D), dog(D)

B: buy(?X,D), dog(D).

Here the "elaboration" consists in the addition of the question-mark operator.

The same computational processes that recognize Elaborations will, with slight changes, also recognize Answers and Requests for Elaboration. Moreover, the functions of Answers and Requests for Elaboration are similar to the function of Elaboration. Answers resolve lack of understanding. Requests for Elaboration indicate it.

For the next two relations we need a definition of the complex notion of *similar entities*. Two entities A, B in a text are *similar* if $A = B$ or if a property P1 of A follows-from some property of A in the text and a property P2 of B follows-from some property of B in the text, where the predicates of P1 and P2 are identical and all pairs of corresponding arguments other than A and B are similar. For example, in the phrases "the foot F of ladder L" and "the top T of ladder L," the entities F and T are similar: from the property of F "foot(F,L)" we can infer "end(F,L)," "end(T,L)" follows-from "top(T,L)," these propositions have identical predicates, and the pair of corresponding arguments, L and L, are similar since identical.

The reader may object that almost any pair of entities would satisfy this definition. For example, Jimmy Carter and the planet Jupiter are both physical objects. Recall, however, that the relation "follows-from" is a matter of degree and thus imposes a matter of degree on the notion of "similarity." I would expect the knowledge base to be constructed in such a way that that pair would have very low similarity in most contexts.[6] In the example of Section 3.3, the similarity of a man and a ladder in the context of a physics problem turns out to be crucial. It may be, however, that further constraints are needed—e.g., that

[6]This is an example of the point made in the final paragraph of Section 2.2. It is not a circle; it's a spiral staircase.

they share some other property or that they exhaust some independently definable set.

Parallel. S0 and S1 are in Parallel if propositions P0 and P1 follow-from the assertions of S0 and S1 respectively, where P0 and P1 have identical predicates and the corresponding arguments of P0 and P1 are similar.

The second sentence of (1) is an example:

Set the stack pointer to zero, and set link variable *P* to ROOT.

Here propositions P0 and P1 are the assertions themselves; no inferences are required:

P0: set(*Pr*,*SP*,*0*),
P1: set(*Pr*,*P*,ROOT),

where *Pr* is the processor and *SP* the stack pointer. The predicates are identical, as are the first arguments. The second arguments—*SP* and *P*—are similar since both are variables. The third arguments are similar in that both are possible values.

This example also exhibits syntactic parallelism, but it should be emphasized that this is not an essential ingredient. The example in Section 3.3 illustrates the Parallel coherence relation without syntactic parallelism.

Why should a discourse tend to become organized along these lines? In spite of the fact that the second sentence in a Parallel construction may be largely new information, the Parallel pattern allows it to be handled with the minimum of reinterpretation, for processing the second sentence requires only abstracting away from the specific statement of the first sentence to a general framework with a number of slots—in the above example

set(*PR*, <data-structure>, <value>)

—and reinstantiating the framework with new specifics. The speaker or writer thus minimizes the cognitive load on his audience by streamlining the search needed for interpretation.

When we look at conversation, we find something very similar operating. A "That Reminds Me" move will be judged relevant to the extent that it exhibits the Parallel relation. Suppose, for example, you tell me about your backpack trip in the Sierras when it rained the whole weekend. If I respond with a story about how I hiked for two days in the rain in the Berkshires, it will be judged relevant, whereas if I tell you about how I got mugged in Philadelphia last year, it's likely to raise eyebrows. If I am able to generalize from your story, and reinstantiate it with details of my own, it signals an understanding of what you intended to convey.

It should be noted that since identical entities are trivially similar, the Elaboration relation is a limiting case of the Parallel relation. Where the Elaboration relation does in fact hold, we will say that the text exhibits Elaboration *rather than* Parallel. Where it is possible, as described below, to identify entities without contradiction and thereby match the Elaboration pattern, we will consider it a better match than if we had only recognized the Parallel relation.

Contrast. S0 and S1 are in Contrast if propositions P0 and P1 follow-from S0 and S1 respectively, where P0 and P1 have one pair of elements that are contraries, and the other pairs of corresponding elements are similar.

An example is

> You are not likely to hit the bull's eye, but you're more likely to hit the bull's eye than any other equal area.

Here the proposition P0 that follows-from the first clause is "$p < q$," where p is the probability of hitting the bull's eye and q is whatever probability counts as "likely." The proposition P1 that follows-from the second clause is "$p > r$," where r is the typical probability of hitting the other equal areas. "$<$" and "$>$" are contraries. The first arguments—p and p—are similar since identical. The second arguments—q and r—are similar in that both are probabilities.[7]

The reason given for the importance of the Parallel pattern operates here as well. The speaker has a mass of facts to impart in some order. He tries to choose an order that minimizes the processing needed for comprehension, by saying next a sentence that uses the same underlying framework. In the Contrast relation, a slightly greater cognitive load is probably placed on the listener since one of the slots in the framework has to be negated.

In conversation, a disagreement can be viewed as a Contrast in which the similar elements are in fact identical. This should give us a further insight into the function of the Contrast relation. One effect of the Parallel relation is to invite the generalization upon which the Parallelism is based. The Contrast pattern has the opposite function—to fend off illegitimate generalizations. This can be seen very clearly in the exchange

> A: I was hitchhiking in Norway, and nobody would pick me up.
>
> B: I found the Norwegians I met very friendly.

B's response resists what seems to be an invited generalization about the character of the Norwegian people. In fact, one could imagine A saying the second sentence himself as an afterthought, to fend off the generalization he is afraid a listener might make.

[7]Note that this example also exhibits the Parallel relation, for from "$p < q$" we can infer "$q > p$" which matches "$p > r$." Three things cause us to favor the Contrast pattern, however. The chain of inference establishing Parallel is one step longer, the match is not as strong since it lacks the "$p = p$" identity, and the conjunction "but" predisposes us to Contrast.

2.4 Coreference from Coherence

I have argued that people participating in the creation of a discourse tend to make it coherent, partly because it lightens the burden in comprehension and thus enhances the likelihood of being understood.[8] The devices for achieving this described in the last section all involve a high degree of overlap in the information conveyed by successive sentential units. A natural consequence of this is that successive sentenial units refer to the same entities. That is, coreference is due in part to coherence.

The speaker's strategy works rather better than might be expected. Because the speaker knows the discourse is coherent and knows the listener knows it is intended to be coherent, he can leave many entities unmentioned or minimally described. He knows the listener can use the coherence assumption to recover the entities. The listener's strategy is to do the best he can to recognize coherence, then to make those coreference assumptions that will allow coherence to go through. Following this strategy solves a remarkable number of coreference problems.

The examples in Section 3 illustrate different sorts of reference problems and how their solutions "just happen" once we direct our attention not to reference itself but to the deeper problem of coherence.[9]

3. THREE EXAMPLES OF REFERENCE RESOLVED

3.1. Resolving Reference Against Prior Discourse

Consider the text

John can open Bill's safe. He knows the combination. (3)

There is a common heuristic for resolving pronouns, defined in Hobbs (1976a), which says among other things that we should favor the subject over a noun phrase in the object position. That would work here. But I can change the example out from under the heuristic:

John can open Bill's safe. He's going to have to get the combination changed soon. (4)

or

Bill is worried because his safe can be opened by John. He knows the combination. (5)

[8]This does not mean that the speaker is necessarily conscious of the coherence relations. He is usually only vaguely aware that he is moving from idea to idea in a more or less orderly fashion. In a sense, the theory of coherence is a theory of the structure of how we are reminded of things, as we proceed toward our discourse goals.

[9]The notion that pronoun references can often be resolved as a by-product of other processes of comprehension was pointed out in Hobbs (1974) and developed at length in Hobbs (1976a). A similar approach is used by Lockman (1978). Schank and Abelson (1977) apparently resolve some pronoun references in the course of applying a script to a text, which can be viewed as implementing the same notion.

In these, "he" no longer refers to the subject. The heuristic not only gives the wrong answer. It gives no indication that it might be wrong.

Another commonly used technique is to try to find an entity in the prior text whose properties would imply the properties we know about the pronoun. In (3), all we know about the referent of "he" is that he knows the combination. We can infer this not only about John from the fact that he can open the safe, but also about Bill from the fact that he owns the safe. So this technique fails us here.

The second sentence of (3) poses three discourse problems—What is the antecedent of "he"? What is the combination a combination of? And what is the relevance of this sentence to the first? I will ignore the first two problems for the moment and concentrate on the third.

The two sentences exhibit the Elaboration relation. In fact, they are similar to (1) in that the first sentence describes the situation from a global perspective, while the second gives procedural detail. How is this recognized?

Suppose we have in our store of commonly possessed world knowledge the following axioms:

$$\text{can}(x,\text{state}) \rightarrow \text{know}(x,\text{cause}(\text{do}(x,a),\text{state})) \tag{6}$$

If x can bring about *state*, then there is an action a such that x knows that x doing a will cause *state* to hold;

$$\text{combination}(x,y),\ \text{person}(z) \rightarrow \text{cause}(\text{dial}(z,x,y),\text{open}(y)) \tag{7}$$

If x is the combination of y and z is a person, then z dialing x on y will cause y to be open;

and the following rule of plausible inference:

$$[\text{know}(x,p),\ p \rightarrow q] \vdash \text{know}(x,q) \tag{8}$$

One is normally able to draw the commonly known implications of what one knows (but of course not always).

Then the Elaboration relation in (3) is recognized as follows: The assertion of (3) is

$$\text{can}(\text{John},\text{open}(\text{Safe})).$$

From this we can infer

$$\text{know}(\text{John},\text{cause}(\text{do}(\text{John},a),\text{open}(\text{Safe}))). \tag{9}$$

That is, from "John can open the safe" we can infer by axiom (6) that John knows some action that he can do to cause the safe to be open. From

$$\text{know}(\text{he},\text{combination}(\text{Comb},y))$$

we can infer

$$\text{know}(\text{he},\text{cause}(\text{dial}(z,\text{Comb},y),\text{open}(y))) \tag{10}$$

by applying axiom (7) inside the predicate "know," as provided for by rule (8). That is, since it is common knowledge that dialing the combination of some