COHERENCE AND COREFERENCE

Technical Note 168

August 4, 1978

By: Jerry R. Hobbs
Computer Scientist
Artificial Intelligence Center

SRI Projects 5844 and 6891

The work reported herein was supported by the National Science Foundation under Grant No. MCS 76-22004 and by the Advance Research Projects Agency under Contract N00039-78-C0060 with the Naval Electronic Systems Command.

ABSTRACT

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.

I 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.

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., to 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 III 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.

II CHARACTERIZING COHERENCE

A. 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), "conjunctive 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 an 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 II.B I describe briefly the basic design of an inference system for natural language processing. In Section II.C 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 corresponding to 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 these questions 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 III.

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.

B. The Inference Component

The typical inference system* 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.

----

* Most of what is described here is embodied in a working computer program.

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.* 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"

----
* 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.

(cf. Hobbs & Grishman 1977, 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 underlined 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 II.C. We will see in Section III 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.

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

----

* One way of implementing such a measure is described in Hobbs 1976b.

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.

C. Some Coherence Relations Defined

There are at least two "directions" in which coherence relations can "carry" a discourse. Relations in one class cause the discourse to move along, either forward or backward. Among these are Temporal Sequence and Cause. Relations in the second class cause the discourse to be expanded in place. Three such relations -- 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 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.

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. But since there must have been some reason for saying it again, we require that new information be conveyed by the second sentence.

An example 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.* This is seen in the above example, and also in the following example from an algorithm description:

----

* I am indebted to William Mann for pointing this out to me.

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, question-answer sequences can be viewed as a kind of Elaboration. To see this, we must extend our formalism slightly by adding a question-mark operator, "?", which can be applied to entities (?X) or propositions (p(A,B)?) to indicate what is being questioned. A pragmatic component will use this operator to determine, at the deepest level, what an appropriate response would be. The inference component will need a few special rules for "?", such as

[x=y?, p(y)] --> p(x)? (2)

that is, the question "Is x identical to y, where p is true of y?" 'implies' the question "Is p true of x?" Otherwise, questions can be treated just like other propositions.

Then a question-answer sequence, such as

A: Who bought the dog?

B: The boy bought the dog.

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 <u>similar entities</u>. Two entities A, B in a text are <u>similar</u> 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.* In the example of Section III.C, 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 they share some other property or that they exhaust some independently definable set.

<u>Parallel</u>: 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 inferencing is 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 III.C 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

----

* This is an example of the point made in the final paragraph of Section II.B. It is not a circle; it's a spiral staircase.

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.

<u>Contrast</u>: 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.*

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.

D. <u>Coreference from Coherence</u>

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.* The devices for achieving this described in the last

----

* 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.

section all involve a high degree of overlap in the information conveyed by successive sentential units. A natural consequence of this is that successive sentential 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 III 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.

---

* 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.

III THREE EXAMPLES OF REFERENCE RESOLVED

A. 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)

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:

```
can(x,state) --> know(x,cause(do(x,a),state))          (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;

```
combination(x,y), person(z) -->                        (7)
        cause(dial(z,x,y),open(y))
```

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:

```
[know(x,p), p --> q]  |- know(x,q)                      (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: From

```
can(John,open(Safe))
```

we can infer

```
know(John,cause(do(John,a),open(Safe)))                 (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

```
know(he,combination(Comb,y))
```

we can infer

```
know(he,cause(dial(z,Comb,y),open(y)))                  (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 object causes it to be open, John's knowing the combination implies he knows that dialing it will cause the object to be open.

Propositions (9) and (10) are nearly identical. The formal definition of the Elaboration relation would be satisfied if we were to make certain further identifications. There is a strong assumption that the text is coherent -- i.e. that some relation holds between the two sentences. This assumption entitles us to make the required identifications, providing no obvious contradictions would result. Hence, we identify "he" with John, z with John, and y with Safe, and the definition is satisfied. The elaboration lies in the greater specificity regarding the action John would perform to open the safe.* The other two discourse problems posed by the sentence -- the antecedent of "he" and the missing argument of "the combination" -- are thus solved in the course of recognizing coherence.

This analysis involves a kind of conversational implicature, as discussed by Grice (1975). A conversational implicature is an assumption one makes without otherwise adequate evidence in order to see a discourse as coherent. (A slightly more coherent version of) Grice's principal example is

A: How is John doing on his new job at the bank?

B: Quite well. He likes his colleagues and he hasn't embezzled any money yet.

Grice argues that in order to see this as coherent, we must assume both A and B know that John is dishonest.

In the analysis of (3) we see what might be called a "petty conversational implicature". To see the coherence, we must assume the three identifications -- he = John, z = John, and y = Safe -- even though the principal evidence for this is the requirement of coherence itself. The kind of conversational implicature Grice gives examples of is a rather rare occurrence in conversation. The petty implicatures described here happen all the time, with almost every sentence we process.

----

* A paper by Moore (1977), which suggested this example, gives a proof of the connection between the two sentences in a more rigorous fashion, in terms of possible worlds. He does not address the pronoun or coherence issues. A similar example is also treated in McCarthy (1977).

Although analysis of (3) in terms of coherence yields a solution to the pronoun resolution problem, the heuristic mentioned above of favoring the subject over the object should not be dismissed so lightly. It is very powerful, working about 90% of the time in written texts (Hobbs 1976a) and about 75% of the time in dialogs I have examined. It seems to be at work in this example as well. For when we hear

John can open Bill's safe. He. . .

we are likely to assume "he" refers to John. If the sentence continues as in (3) then all is fine, but if it continues as in (4), we back up and change our commitment. This strongly suggests that some psychological reality underlies the heuristic.

How might this heuristic have arisen? The statistics show that it is a very good one. Why is it so good? We can get part of an answer by looking at the coherence relations of Section II. They all involve close correspondences between the assertions of the two sentences, and they are strongest when the corresponding arguments of the assertions are identical. So if an entity is the Agent of some description of an action, it is likely to be the Agent of most other descriptions of the action. Since the Agent usually appears as subject, matching a subject pronoun with the subject of the previous clause or sentence is a very good guess.*

This heuristic is especially effective when the pronoun is in subject position, for it allows us to begin processing the sentence right away, without juggling an ambiguity, and only rarely making us back up and start again. But I suspect that coherence underlies the heuristic, that it is because of coherence that the heuristic is so good. And the results of the heuristic must always be checked against what considerations of coherence tell us.

The coherence solution does not depend on whether the first clause is active or passive, so it would work in exactly the same way on (5) where the syntactic heuristic would fail. Example (4) is an instance of Causality. Recognizing this depends on knowledge of the purpose of a

----

* This does not explain the heuristic fully of course.

safe and the purpose of a combination, but a detailed analysis is beyond the scope of this paper.

It is interesting to compare the analysis given by the theory of coherence with another description that has been proposed for one of the processes in the comprehension of coherent discourse -- the three-step process of Clark and Haviland's (1977) given-new contract:

1. Divide the current sentence into the given and the new information.

2. Search memory for unique antecedents that match the given information.

3. Add the new information to what is already known about those antecedents.

Computationally speaking, this description raises certain difficulties.

In the second sentence of (3), I presume Clark and Haviland would label as given the entities referred to by "he" and "the combination" C and the fact that C is a combination of something:

Given: he, C, combination(C,x).

His knowing the combination would be labelled new:

New: know(he,combination(C,x)).

But as we have seen, the knowing is not exactly new. Some kind of knowledge is involved in John's ability to open the safe, and while it could be knowledge of how to use dynamite or knowledge of safecracking, knowledge of the combination is the most likely candidate. This knowledge seems to be almost as much <u>given</u> by "can open" as the existence of a combination is given by "safe".

Moreover, although the entity referred to by "he" is certainly given, step 2 provides no way of deciding which of the two men it is. It is precisely the supposedly new information -- his knowing the combination -- together with our assumption that the text is coherent, that allows us to choose the antecedent of "he" correctly, and gives us one path to the referent of "the combination" as well.

In short, it is not always clear in step 1 how to divide the sentence into given and new in a way that corresponds to our common understanding of these terms. Even if we could, the so-called given information is frequently insufficient in step 2 for us to identify the antecedents uniquely. The ways in which information in different parts of a discourse overlaps and interacts, and the ways in which the parts influence the meaning of one another are somewhat richer and more complex than is captured by the given-new contract.

B. Resolving Reference Against a World Model

The next example of a reference resolution problem comes from a set of dialogs between an Expert and an Apprentice involved in repairing an air compressor, collected by Grosz (1977):

E: Replace the pump and belt please.

A: I found a belt in the back [of the air compressor]. (11)
Is that where it should be?

There are two problems here that I will discuss. They will turn out to have the same solution.

First of all, this example illustrates a different kind of reference than in the previous example. There we were resolving against a world created by the first sentence itself. Here the world is already present in the form of a real air compressor. This has consequences. If the apprentice had said only the first sentence,

A: I found a belt in the back.

it still would have counted as a Request for Elaboration on the belt mentioned by the expert because of the word "a". Although the common way of handling indefinite noun phrases in computational systems is simply to posit a new entity, that won't work in this domain, for all entities are given beforehand. Even the indefinite noun phrases must be resolved against the model of the air compressor. Here the indefinite noun phrase indicates an uncertainty, and therefore the utterance functions as an implicit question about the location of the belt the expert is referring to.*

---

* In another dialog, the apprentice, after completely assembling the air compressor, says, "I found a screw on the floor." This functions as a warning, or a misgiving.

In the apprentice's second sentence, we must find the antecedents of "it" and "that". "That" presents no problems, but "it" does. The candidates for its antecedent, in the order given by the common heuristic, are the belt the apprentice found, the back, the pump, and the belt the expert mentioned. The intended antecedent is the last of these. In a sense, we have to skip over the apprentice's first sentence, while picking up enough of its information to know that it is the belt rather than the pump we are looking for in the expert's utterance.

Assume that the expert's utterance has been reduced to the propositions

replace(A,{P,B}), pump(P), belt(B). (12)

Consider the processing of the apprentice's first sentence, which in propositional form is

find(A,b), belt(b), in(b,Bck), back(Bck,x).

The symbol "b" is used for the belt since we have no guarantee that it is the same as the belt B. By lexical decomposition of "find" we can infer

come-about(know(A,in(b,Bck))), belt(b).

Since something which comes about is now true, we can infer

know(A,in(b,Bck)), belt(b).

Since something that is known is true, we infer

in(b,Bck), belt(b). (13)

What is needed to achieve a match with <u>some</u> coherence relation between (12) and (13)? The only match is on the predicate "belt". We are not free to strengthen the match by assuming that b and B are the same. However, we are entitled to assume there is an implicit question

b = B? (14)

This is a petty implicature similar to the ones drawn in the previous example. Then by the "substitution" rule (2), the first proposition of (13) becomes

in(B,Bck)? (15)

that is, "Is the belt you're referring to in the back?" This is a more specific way of asking "Where is the belt you're referring to?" which itself is a more specific way of asking "What belt?" Thus, we have matched the Request for Elaboration pattern.

Since the apprentice's second sentence

Is that where it should be?

is clearly, on an intuitive level, a Request for Elaboration on the belt, our problem in processing it is to see it as a paraphrase or Elaboration of the apprentice's first sentence. I will ignore the modal "should" since that complicates the detail without changing the substance. The sentence may then be represented

at(it,that)? (16)

where "at" may be viewed as a general locative operator and thus a generalization of "in". From (15) we can derive

at(B,Bck)? (17)

If we assume "it = B" and "that = Bck", (16) and (17) match. Elaboration is recognized and the reference problems are solved. The requirements of coherence have thus forced us to compute the very indirect speech act of the first sentence and given us the antecedents of the anaphors in the second.

Note that the two sentences ask the question from different perspectives: the first in the context of the apprentice's best efforts to answer it for herself; the second, because of the modal "should", in the context of the expert's knowledge.

At this point I might as well admit there is a certain dubious quality about my analysis of the apprentice's first sentence. It's a rather big jump made on rather slender evidence to assume the implicit question "b = B?" But isn't this precisely the same as the dubious quality of the illocutionary force of the sentence itself, stripped down to its computational kernel? That is, the issue of whether or not the sentence functions as a Request for Elaboration hinges, computationally speaking, on whether or not we are free to draw the implicature "b = B?"

But something like the following cognitive processing seems plausible to me in the comprehension of (11). The listener performs his equivalence of the inferencing down to (13). He does not draw (14) and (15), but in a sense, they are there, available. They may just be masked out by various other, equally dubious coherence possibilities. Then the next sentence comes in. Having failed to interpret the first sentence in an entirely satisfactory manner, the listener is ready for an Elaboration or clarification. The first word tells him it is a question, tending to raise the question "b = B?" to prominence. When the third word "where" comes in, it is known to be a question about location, causing (15) to be inferred. If there is any truth to this rank speculation, it is an interesting case of the surface form of one sentence influencing the deep interpretation of another.

Moreover, it is likely that the expert is poised to interpret any of the apprentice's utterances as questions, for he has just told her in one sentence to do the whole job, which he knows she cannot do without further aid. The expert's utterance functions primarily as a way of handing over to the apprentice the responsibility of determining the level at which the instruction will be conducted.

C. Resolving Reference Against an Alternate Representation

The final example of reference resolution illustrates a different kind of reference problem. Novak (1977) developed a system which first translates physics problems into the corresponding diagram,* associates the appropriate equations with the diagram, and then solves them. For each text, the system must discover the diagram the text refers to. Resolving reference then is not simply a matter of mapping a noun phrase in English into an entity in a semantic representation of the standard kind. Rather, it involves mapping a complex linguistic description into the corresponding complex "visual" representation. This differs from linguistic or propositional representations in what must be specified and what can be left vague. The present example illustrates a case in

---

* Or an internal model of the diagram.

which a decision that must be made in order to draw the diagram can be made on the basis of the text's coherence.

The text is

> The ladder weighs 100 lb with its center of gravity
> 20 ft from the foot, and a 150 lb man is 10 ft from the top. (18)

There is a minor problem in that the argument of "top" is not specified, but there are several paths to a solution, including the one shown here. But the real problem is the precise location of the man. We are not told where he is on the sphere of radius 10 feet with its center at the top of the ladder -- whether on the ladder, on a roof 10 feet from the top of the ladder, on the limb of a tree, or just where. These interpretations are not necessarily bizarre in all contexts. For example, in

> The firemen almost succeeded in saving John. He was 10 ft from the top of the ladder

our assumption is that John was not on the ladder. Novak's system assumes the man is on the ladder by convention. But it is possible to arrive at this fact from deeper considerations of coherence.

At first glance, (18) would seem to be a case of simple <u>logical conjunction</u>. But in spite of the fact that logical conjunction is usually mentioned in investigations of this sort (Grimes 1975, Halliday and Hasan 1976, Longacre 1977), I have not included it in my list of coherence relations in this paper or any other paper I have written about coherence. I believe logical conjunction is simply not enough by itself to confer coherence on a text. The best collection of examples illustrating this is in Robin Lakoff's "If's, And's, and But's about Conjunction" (1971), where she lists numerous examples like the following:

> John eats apples and many New Yorkers drive Fords.

But (18) poses a problem. If it is not simple logical conjunction, what is it?

The demands of the task dictate that we decompose both clauses of (18) into expressions involving a task primitive we might call "force".

"Force" has four arguments -- there is a force of a particular <u>magnitude</u> acting on a particular <u>object</u> (since forces can be exerted only on objects) in a particular <u>direction</u> at a particular <u>point</u>. The first clause then decomposes into the propositions

force(100 lb,L,Down,X1), distance(F,X1,20 ft), foot(F,L), ladder(L), (19)

i.e. there is a force of 100 lb acting on ladder L in a downward direction at point X1, where X1 is at a distance of 20 ft from F which is the foot of L. The force acts on L because the force at the center of gravity of an object always acts on the object. We know "the foot" refers to the foot of ladder L because the predicate "foot" requires for its second argument an object with a canonical, real or metaphorical vertical orientation (cf. Section 4.1, Hobbs 1976a, for the detailed analysis of a similar example.)

The second clause decomposes into

force(150 lb,x,Down,X2), distance(T,X2,10 ft), top(T,y), (20)

i.e. there is a force of 150 lb acting on some object x in a downward direction at point X2, where X2 is at a distance of 10 ft from T, which is the top of something y. We must identify x and y more exactly. In isolation, x could be the ladder, the man himself, or some other object, such as the floor. In seeking to establish coherence, however, we notice the close similarity of (19) and (20). This strongly suggests the Parallel relation. The match can be strengthened, first of all, by drawing the petty implicature x = L. That is, the text is more coherent if we assume the man's weight is a force acting on the ladder. Since mechanical forces between objects are transmitted only by contact, we can infer that the man is on the ladder.*

----

* Another possible solution uses the fact that ladders are normally for people to stand on. However, it happens that a floor is mentioned in the sentence preceding (18), and floors also are for people to stand on. Something more is required for the disambiguation. The impossibility of the man standing on the floor can be deduced from the geometry of the situation, but that involves rather complex reasoning, and we could change the geometry without changing the normal interpretation of (18).

Then since both a top and a foot are ends, F and T will be similar if we assume y = L. We thereby satisfy the definition of the Parallel relation, simultaneously resolving "the top", locating the man precisely, while recognizing the coherence of the text.

Is there any psychological validity to this analysis? I think there is. If we are asked what, intuitively, the two clauses are <u>about</u>, we would very likely say they are about forces acting on the ladder. This is precisely what the computational treatment discloses, for the ladder L and the predicate "force" are what are constant between the two clauses.

All three examples, (3), (11), and (18), illustrate an interesting fact about language. A sentence typically poses a number of problems -- reference or coreference, interpretation, ambiguity, and coherence problems. A significant amount of inferencing is required to solve them. But one of the things that saves us from a combinatorial explosion is that many of the problems have the same or almost the same solution. In Section III.B for example, the chains of inference required to compute the illocutionary force, resolve "that" and "it", and recognize the coherence of the whole exchange, are virtually identical. The very high degree of redundancy in natural language has been noted by many linguists (e.g. Joos 1972). The fact that one solution often solves several problems is part of the computational significance of this redundancy.*

----

* This redundancy has a further significance. One of the problems with the approach I have presented here is that the inference mechanism is operating in a very rich knowledge base. I have shown only the correct chain of inference in each case, and have assumed it is the most salient. But in reality there may be many chains of inference of roughly equal salience. How do we choose? Part of the answer may lie in the redundant nature of language itself. Where several competing solutions to a discourse problem present themselves, we choose the solution that solves the most other discourse problems.

IV THE PLACE OF COHERENCE IN A THEORY OF COMMUNICATION

Conversations are planned. For convenience, we may distinguish three levels of planning, although no clean separation is possible in fact. The levels can be characterized as the message level, the coherence level, and the description level.

1. Participants in a conversation begin with certain goals -- to obtain or impart information, to elicit some action from the other participants, to express some compelling idea or feeling, to present a particular image, or simply to maintain contact. They develop plans for these goals by breaking them into subgoals and breaking the subgoals into further subgoals until the subgoals can be implemented as a message to be uttered. We may view this process as the deepest level of planning.

2. Once the content of the principal message has been decided upon, the speaker may feel compelled to provide some necessary background information. He may decide to split the message into two or more utterances to give the information from several perspectives -- say, a sentence from a global perspective to orient the listener followed by utterances providing more detail. After saying something, he may decide a clarification or elaboration is necessary, whether through a question or questioning expression from another participant or through hearing himself speak. He may bolster his message by drawing generalizations, parallels, and contrasts and giving examples. He may feel called upon to give explanations, describe causes, suggest results. This second level of planning is the level of coherence, which has been investigated in this paper. The first two levels may be characterized roughly as follows: first-level planning involves the desire to communicate some message; second-level planning is motivated by the desire to have that message understood.*

3. The third level of planning is within the sentence itself. Once the global plan and the requirements of coherence have determined <u>what</u> is to be said, the speaker must decide* <u>how it</u> is to be said. This includes the choice of lexical items, grammatical constructions, and the appropriate descriptions of entities and events.

In modelling the production of utterances in ongoing conversation, we must take all three levels into account, as well as the interactions among them. The influence of the coherence level on the description level is illustrated by the examples of Section III, which show how the particular coherence move that is chosen, and the fact that it is a standard coherence move, enable certain entities to be minimally described by pronouns, or to be omitted altogether.

In modelling comprehension, however, it is not clear what we need to assume about how deeply the listener penetrates the goals of the speaker. Cohen (1978) has suggested that the process of comprehension is largely the process of discovering the speaker's goals, and there are certainly examples of conversations in which one of the participants is trying to "psych out" the other, trying to get behind the utterances to the deep goals that give rise to them. But it was shown in the example in Section III.B that rather indirect speech acts can sometimes be interpreted at the level of coherence. It seems quite possible to me that most conversations are one of two sorts: (1) conversations in situations so constructed that the goals of the participants mesh nicely with each other, such as teacher-student dialogs or exchanges at an information booth, so that responses can be made at the relatively superficial level of coherence; and (2) conversations in which the participants, for the most part, talk past each other, each person's utterances arising out of his own deep goals or merely as coherent

---

* Of course, not all discourse is coherent. In a conversation, we typically find a sequence of islands of coherence of varying sizes, as issues are taken up, explained, elaborated, developed and dropped. In written discourse, one structure frequently covers the entire text.

* "Decide" and "choice" are strong words for what is a subconscious or barely conscious process.

responses to another's utterances, so that deep communication fails precisely because the participants penetrate no deeper than the level of coherence. In either case, analysis only to the level of coherence suffices to explain ordinary behavior.

On the other hand, analysis at least to the level of coherence is necessary. A model, such as Clark and Haviland seem to propose, operating strictly at the level of description, is rarely adequate, as examples like those given in Section III demonstrate. A deeper mechanism for recognizing coherence must be present in order for listeners to solve the coreference problems they routinely solve. One of the major problems that now faces natural language processing is to elucidate this mechanism of coherence further, for it is in the problems of coreference that the discrepancies between human language use and our models of human language use stand out most starkly.

ACKNOWLEDGEMENTS

I am grateful to Barbara Grosz, Gary Hendrix, Bob Moore, Gordon Novak, Ann Robinson, Jane Robinson, Earl Sacerdoti, and Richard Waldinger for their comments on an earlier version of this paper.

REFERENCES

Clark, H., & S. Haviland 1977. Comprehension and the given-new contract. In R. Freedle (Ed.), _Discourse production and comprehension_, Vol. 1, 1-40. Norwood, N.J.: Ablex.

Cohen, P. 1978. On knowing what to say: planning speech acts. Technical Report No. 118, Department of Computer Science, University of Toronto. January 1978.

Fillmore, C. 1974. Pragmatics and the description of discourse. In C. Fillmore et al (Eds.), _Berkeley studies in syntax and semantics_, Vol. 1, pp. V-1-V-21. U. of California, Berkeley, California.

Grice, H. 1975. Logic and conversation. In P. Cole and J. Morgan (Eds), _Syntax and semantics_, vol. 3, 41-58. New York: Academic Press.

Grimes, J. 1975. _The thread of discourse_. The Hague: Mouton.

Grosz, B. 1977. The representation and use of focus in dialogue understanding. Stanford Research Institute Technical Note 151, July 1977.

Halliday, M., & R. Hasan 1976. _Cohesion in English_. London: Longman.

Hobbs, J. 1976a. Pronoun resolution. Research Report 76-1, Department of Computer Sciences, City College, City University of New York. August 1976.

Hobbs, J. 1976b. A computational approach to discourse analysis. Research Report 76-2, Department of Computer Sciences, City College, City University of New York. December 1976.

Hobbs, J. & R. Grishman 1977. The automatic transformational analysis of English sentences: an implementation. _International Journal of Computer Mathematics_.

Joos, M. 1972. Semantic axiom number one. _Language_, Vol. 48, 257-265.

Knuth, D. 1973. _The art of computer programming_, Vol.1. Reading, Mass.: Addison-Wesley.

Lakoff, R. 1971. If's, and's, and but's about conjunction. In C. Fillmore and D.T. Langendoen (Eds.), _Studies in linguistic semantics_, 115-150. New York: Holt, Rinehart, and Winston.

Longacre, R. 1977. The paragraph as a grammatical unit. Symposium on Discourse, UCLA. November 1977.

McCarthy, J. 1977. Epistemological problems of artificial intelligence. Proceedings, IJCAI-77, Cambridge, Mass.

Moore, R. 1977. Reasoning about knowledge and action. Proceedings, IJCAI-77, Cambridge, Mass.

Novak, G. 1977. Representations of knowledge in a program for solving physics problems. Proceedings, IJCAI-77, Cambridge, Mass.

Rieger, C. 1974. Conceptual memory: A theory and computer program for processing the meaning content of natural language utterances. Stanford AIM-233.