COHERENCE AND INTERPRETATION IN ENGLISH TEXTS

Jerry R. Hobbs
Department of Computer Sciences
City College, CUNY
New York, N.Y. 10031

Abstract

A theory and system are described for the semantic analysis of complex, coherent English texts. The principal question addressed in this paper is how the meanings of the smaller elements of lan* guage compose into the meanings of larger stretches of text. Within sentence boundaries, this is achieved by an operation, called predicate interpretation, which provides a mechanism for general words, especially those having a spatial flavor, to acquire specific interpretations in context. Ee-yond sentence boundaries it is achieved by an operation which matches successive sentences against a small number of common patterns and builds up a tree-like structure representing the text's patterns of coherence.

In isolation, words and sentences do not have specific meanings so much as they have the potential for acquiring a variety of specific meanings in particular contexts. In this paper we ask how the meanings of the smaller elements of language compose into the meanings of larger stretches of text. The answers proposed come out of work on an inferencing system being developed for the Semantic Analysis of complex, coherent Texts in English (called SATE). This system is intended to be general and is being applied to sets of directions (Hobbs 1975), algorithm descriptions (Hobbs 1977), and complex expository texts (Hobbs 1976b). Section 1 of this paper briefly describes the system and the collection of world knowledge axioms it rests on. Section 2 considers the question within sentence boundaries, addressing the problems of how words should be defined, how context influences the meanings of words, and how metaphors are to be interpreted. Sections 3 and 4 discuss how the relations between sentences can be discovered and hence how the structure of paragraph-length texts can be built up.

1. The Inferencing System

The data of the semantic analyzer is of two sorts — the Text and the Lexicon. The Text may be thought of as episodic memory and the Lexicon as permanent memory. The Lexicon contains type nodes, the Text token nodes. Initially, the Text contains the information explicit in the paragraph being analyzed, expressed in a simple logical notation produced by a syntactic preprocessor (Grishman et al 1973, Hobbs and Grishman 1977). The logical notation makes explicit the functional relationships between elements in the sentence by expressing the information in the form of logical propositions, and distinguishes between material which is asserted and material which is grammatically subordinated, or presupposed. In the course of semantic processing, certain <u>semantic operations</u> augment the Text by drawing inferences, interrelate it by identifying phrases in different parts of the paragraph which refer to the same entity, and structure it by discovering relations between sentences.

The semantic operations work by searching for chains of inference in the Lexicon, which is the store of lexical and world knowledge. The Lexicon contains the "definitions" of English words, where a definition is viewed simply as the set of inferences which may be drawn from the use of that word. These inferences include "superset" relations, such as the fact that a bank is a building,

BANK: (Vy(bank(y) •*• building(y))

and that to walk is to go from one place (z.) to another (z).

WALK: (Vy) (3 $^{z}{}_{1}$'$^{z}{}_{2}$) <^{walk}W "* g o t y ^ ^))

These axioms are part of the definitions of "bank" and "walk", respectively. The definitions also include lexical decompositions, e.g. that for y. to go from y to y is for y's being at y_2 to become or change into y^'s being at y_:

GO: (Vy_1,y_2,y_3)(go(y_1,y_2,y_3) •*
become (at ($y_{1\#}y_2$) ^atCy^y^))

(In the use of higher predicates like "become", we go beyond standard predicate calculus notation.) Also included are axioms having more of a "world knowledge" flavor, such as the fact that a building has a roof:

BUILDING: (Vy) (3z) (building (y) ■* roof(z,y))

No distinction is made among these various sorts of axioms, nor between "linguistic" and "non-linguistic" knowledge.

Inferences are not drawn freely, but only in response to the specific demands of semantic operations. These demands take two forms, i.e. they invoke one of two inferencing procedures:

Forward inferencing: From p(A) try to infer something of a given pattern.

Backward inferencing: Find something in the Text from which p(A) could be inferred.

A dynamic ordering determined by context is imposed on the axioms in the Lexicon. The axioms are divided into clusters, perhaps overlapping, roughly according to topic. The clusters are given an initial measure of salience according to their anticipated relevance to the text at hand. The measures of salience are modified in the course of semantic processing in response to changes in topic. When a fact in a cluster is used, the entire cluster is given maximum salience; while the facts in a cluster are not being used, its salience decays. The Text itself is treated as having maximum salience.

The search for a chain of inference is then an ordered heuristic search whose evaluation function depends upon this dynamic ordering and upon the length of the chain of inference. The search is discontinued after the evaluation function reaches a certain thresh old • For example, a chain of inference of five steps might be found if the axioms involved are of high salience, whereas with axioms of very low salience the search would be cut off after just one step.

This is more than just a device for efficiency. There may be more than one chain of inference satisfying the requirements, each leading to a different interpretation of the text. The one chosen is the first one encountered in the ordered heuristic search. Thus, not only the inferencing process but also the interpretations it produces are made highly dependent on global context, and context is not merely the current state of the Text, but also the current state of the dynamically ordered Lexicon.

The bulk of the axioms are classed as only "Normally" true, which means an inference is not drawn if it would result in a contradiction. For example, we can normally infer from "John walks" that "John balances on his feet", but if we have "John walks on his hands", we simply refrain from drawing that inference [cf. Cercone and Schubert 19753.

There are four principal semantic operations:

1. <u>Predicate interpretation</u>, which is described in Part 2.
2* <u>Detection of intersentence relations</u>, described in Part 3.
3* <u>Knitting</u>, which identifies and merges redundancies and secondarily resolves some pronoun references.
4» <u>Finding the antecedents</u> of definite noun phrases and the remaining pronouns.

The last two operations are described elsewhere [Hobbs 1975, 1976a, 1976b].

2. Interpretation and Coherence within Sentence Boundaries

The basic question addressed in this section is how to interpret words in context. To understand the scope of this problem, consider the variety of meanings the word "go" may take on, even in the reasonably well-behaved sublanguage of algorithm descriptions. In

Go to step T4.

"go" means that the processor next <u>executes</u> the instruction at step T4. In

N <u>goes</u> from 1 to 100 (1)

it means that the value of N successively <u>equals</u> the integers from 1 to 100. In

Next we c[o through the linked list looking for marked nodes

"go" means a pointer variable successively <u>points</u> to the nodes in the linked list. If we look at language in general, examples multiply.

The problem we face then is how to define words like "go" in a succinct manner in such a way as to allow this profusion of possible interpretations. Heretofore, the typical approach has been to specify a large number of the word's possible uses, for example, by means of a list of environment - interpretation pairs [cf. Riesbeck and Schank 1976]. This may be feasible for words with a small number of quite distinct meanings, such as the adjective "fair" which can mean "mediocre", "light-complexioned", or "even-handed". But for the most common words, like "go", it is utterly inadequate. First of all, it fails to capture subtle similarities among the various uses. Secondly, there are too many possible uses — the list would never end.

Our approach is rather to define a word in as general a way as possible, with the ideal of having one sense of the word cover all its uses. The method is to define the word in terms of very general predicates. When the word is encountered in context, a semantic operation, called <u>predicate interpretation</u>, seeks a more specific interpretation for the general predicates.

In this method, "go" can be defined as follows:

<u>go(yl/y2>y3)</u>l(yl goeS from y2 tC y3*

Required: $at(y_{1\#}y_2)$

Infer: become (at(y^y^ ratCy^y^).

That is, first from the explicit properties of y. and y_2, try to infer "$at(y_1,y_2)$" — try to discover in what sense y. can be a£ y_2, where "at" is a very general predicate in the sense that it can be inferred from many other relations between two entities. The information in the chain of inference leading to "at(y.,y)" provides the more specific relation between y. and y_2. Then we are free to infer that that relationship between y. and y_2 changes into the same relationship between y, and y_3.

For example, if (1) is being analyzed, predicate interpretation must determine in what sense N could be a£ 1. It does so by accessing an axiom in the definition of "variable":

(Vy,z) (variable(y) + (equal(y,z) -► at(y,z))). (2)

That is, if a variable is equal to a number, we will say it is a£ that point on the number scale. "At" is thus a generalization of "equal", and when we say that a variable comes to be at a number, one of the things it could mean is that the variable comes to equal the number. If this is the first chain of inference found in the ordered heuristic search, then this is the interpretation assumed.

Our approach overcomes the objections to the multiple-sense approach. First of all, the subtle shades of similarity and difference among various uses of a word are captured by the degree to which the general predicates are central or prevalent in the definition of the word, and on the similarities among the chains of inference interpreting the general predicates. Secondly, there is no need to anticipate all the possible uses of a word, for its particular interpretation is determined by means of a search through the collection of world knowledge axioms which is needed anyway in a natural language understanding system.

It may be objected that (2) is ad <u>hoc</u> and listed in the definition of "variable" only to handle the one sentence. But in fact, (2) is an example of one of the basic principles upon which the deepest levels of the Lexicon are organized. Linguists and psychologists have frequently noted that in English and other languages, one often appeals to spatial metaphors when speaking of abstract ideas [Whorf 1956, Clark 1973, Jackendoff 1976]. These permeate language to such an extent that they generally escape notice. For example, in the previous sentence — "permeate", "extent", "escape notice". We speak of "high hopes", "high prices", "deep thought", "being iii politics", "a book on sociology", "getting the idea", "giving an example".

Mow the predicate interpretation method allows us to turn the spatial metaphor into a powerful aid in the analysis of English texts.

At the base of the Lexicon are a small set of cognitive primitives with a highly spatial or visual flavor. Among these are the notion of a scale, which results from conceptualizing a "becoming" from one state to another, and may be thought of roughly as a partial ordering; several notions which give structure to scales, including a point being 0n a scale or a member of the partially ordered set, one point exceeding another on the scale, and three operators, Lr, Md, and Sm, which isolate imprecise overlapping regions near the top, middle, and bottom of a scale, respectively; and the notion of an entity being a a point on a scale, or an entity being at another entity. At, scale, and the Lr, Md, and Sm operators are general predicates capable of a wide variety of possible interpretations in a given context.

The most common words are then defined in terms of these primitives. For example, verbs of motion, like "go", have an underlying at which can get interpreted as physical location or some metaphorical location, as we have seen. Many prepositions also have an underlying "at"; "under" can be defined in terms of the relative positions at_which its arguments are located on a real or metaphorical vertical scale — metaphorical to handle such sentences as

The GNP is under what it was last year.

A word like "greater" is defined in terms of exceed and scale. Depending on the arguments of "greater", scale can get interpreted in various ways. A word like "tall" gets defined usually in terms of the specific physical height scale but also in terms of the "Lr" operator; the interpretation of "Lr" depends on context — whether it is a "tall building" or a "tall glass of beer".

Many interpretations are achieved by defining other relations as instances of "at", as in (2). Moreover, "at" can be related to predication in general. We take our cue from equivalences such as

John is hard at work - John is working hard,

and similar "argument + preposition + predication" combinations, such as

John is in control of the situation * John controls the situation.

The czarina is under the influence of Rasputin ■ Rasputin influences the czarina.

The rule is

$$(\forall y,p)(\exists z)(p(y) \rightarrow at(y,p(z)))$$

That is, if an entity is at a predication, it is possible that the entity is one of the arguments of the predication.

Jackendoff [1976] has classified word senses of verbs of motion into positional, possessional, identificational, and circumstantial modes. Our approach now provides an explanation. Consider

Cinderella's coach turned into a driveway.

Cinderella's coach turned into a pumpkin.

In the first, "turned into" is in the positional mode, or has a positional interpretation. In the second, it has an identificational interpretation; the coach actually becomes a pumpkin. In our approach, "turn" and "into" are defined in terms of, among other things, the predicate "at". One aspect of the meaning of "X turned into Y" may be represented

$$become(\ \ldots, at(X,Y)) \qquad (3)$$

The nature of Y is explored to determine in what sense X could be at Y. If Y is a driveway, the fact that a driveway is intended to be a path leads to the spatial interpretation of "at". If Y is a pumpkin, the spatial interpretation is not as readily available as the identificational, and (3) becomes, by "at a prediction",

become(...,pumpkin(X)).

The predicate interpretation approach also gives us a handle on the analysis of adverbials and adverbial adjectives. "Slow" can be defined as follows:

slow(y):

Required: $y{=}move(z_1)$

Infer: $on(z_2,Sm(z_3))$, $speed(z_2,move(z_1))$, $Speed\text{-}scale(z_3)$

Here "move" is a general predicate. We must find the most salient motion associated with the grammatical argument of "slow". Then we can infer that the speed of this motion is in an imprecisely defined region at the lower end of the speed scale, which is isolated by the operator "Sm". "Sm" may itself receive a more specific interpretation in context. If we encounter "John walks slowly", we use the fact

$$(\forall y)(walk(y) \rightarrow move(y))$$

to interpret the motion as the forward motion of the walking. In "slow watch" we use a fact like

$$(\forall y)(\exists z_1,z_2)(watch(y) \rightarrow (hand(z_1,y)\&hand(z_2,y)\&move(z_x)\&move(z_2)))$$

(A watch has hands that move) to interpret the motion as the motion of the hands. Similarly, in "slow race" the motion is the forward motion of the competitors; in "slow horse" its running at full speed, generally in a race (providing facts about horse racing are sufficiently salient in the Lexicon) ; and in "slow business" the metaphorical motion of the outflow of goods and the inflow of money.

The approach to defining and interpreting words which has been presented here is very compelling on several counts. It allows us to collapse a seemingly large number of senses of a word into a very few. It is appealing as a psychological model since it rests on the close connection between visual imagery and language. It is a method which treats virtually everything as a metaphor to be interpreted in context, and thus provides a framework in which spatial metaphors, as well as much other metaphor, can be treated not as an anomaly but as the natural state of affairs.

3. Coherence Beyond Sentence Boundaries

Where there is an explicit intersentence connective, such as "but", "because", "i.e.", the problem of determining or verifying the relation between the sentences becomes just a special case of predicate interpretation. The connective is viewed as a higher predicate whose arguments are the assertions of the sentences it connects. The requirements associated with the connective are just those conditions which make the sentences stand in contrastive, causal or paraphrastic relation to each other. However, since most intersentence relations are implicit, a separate semantic operation is required to recognize them.

We attempt to determine intersentence relations by seeking to match the current sentence and the previous text against a small number of the most common patterns of coherence that link sentences. The patterns are stated in terms of inferences that can be drawn from the current sentence and a previous sentence, and the modification to be performed on the Text if the pattern is matched. Among the patterns we have studied are Temporal Succession, Cause, Contrast, Violated Expectation, Paraphrase, Example, and Parallel. It is not claimed that this list is exhaustive, but it has been adequate for the quite diverse texts which have been examined so far.

The intersentence operation keeps these patterns on a goal list while processing a sentence. Strengths are associated with each pattern, indicating the urgency with which we want it matched. The strength depends on the type of text being analyzed, the presence of conjunctions and certain adverbs, and the expectations created in the analysis of the previous text. The heuristic search which seeks chains of inference to satisfy the patterns is ordered by the strength of the goal as well as the length of the chain of inference and the salience of the axioms in it.

To see how the patterns are to be stated and how the pattern-matching proceeds, consider the following example:

> Republicans were encouraged about their prospects. The party chairman believed that Dewey would be elected.

Intuitively, this is an instance of the Example pattern. To recognize it computationally requires us to access our knowledge of "Republican", "encourage", and so forth.

More precisely, let the Example pattern be stated as follows:

> The predicate and arguments of the assertion of the current sentence stand in a subset or element-of relation to those of the previous sentence.

Suppose "be encouraged" decomposes into "believe something good will happen". Sentences are processed from the top-down, so that in the second sentence "the party chairman believes" is processed first. Since the party chairman is a member of the set of Republicans and since the "believes" matches the "believe" of "be encouraged", the first part of the Example pattern is matched. Because of this partial match, the strongest goal for the processing of the "that" clause is to show what is believed is that something good for a Republican will happen. This is done by accessing the fact about a political party that one of its purposes is to win elections, and the fact about Dewey that he is a Republican.

The target representation for the analysis of a paragraph is a tree-like structure indicating the relations between sentences in the paragraph. (A brief example of the analysis of a paragraph from Newsweek, together with the target representation produced, is given in Part 4.)

The specifications for the other patterns are as follows:

1. Overlapping Temporal Succession, or "then". There are two varieties of this pattern. (In all the descriptions of the patterns, S- refers to the current sentence and S_1 to an eligible previous sentence.)

   a. S asserts a change whose final state is implicit in S_2.
   b. Implicit in S. is a state which is the initial state of a change asserted by S .

The text

> Walk out the door of this building. Turn left. Walk to the end of the block.

exhibits two varieties of the pattern interlocking. The first sentence asserts a change in location; the final position is implicit in the second sentence and is the beginning of the change of position described in the third. Moreover, the first sentence implies one orientation, the third sentence implies another and the second asserts a change from the former to the latter.

The patterns are exhibited twice in the following text:

> He observed a badly joined connection in the mechanism. He took it to his workshop to fix.

The first sentence describes a change in knowledge, a coming to know that something is broken. The second sentence, via "workshop" and "fix", implies that knowledge.

2. Cause: There are two varieties of this pattern:

   a. Find a causal chain from S. to S_2.
   b. Find a causal chain from S_2 to S..

The second variety is exemplified twice in the text.

> He was in a foul humor. He had slept badly during the night. His electric blanket had worked only intermittantly.

The causality between the third and second sentences can be established by noting that the purpose of a blanket is to enable sleep, so that a blanket failing in its purpose (not working) causes one not to sleep. The causal chain between the second and the first is that sleep enables one to overcome fatigue and fatigue frequently brings on a foul humor.

The role of this relation in narratives has been studied by Schank (1974) and Rumelhart (1975).

3. Contrast: Letting "element" refer to either the predicate or one of the arguments of a statement, the Contrast pattern may be stated as follows:

From S and S infer *S'* and S', respectively, where

1) S' and S' have one corresponding pair of elements which are contradictory or lie at opposite ends of some scale;
2) the other corresponding pairs of elements are identical or belong to the same small set (i.e. are "similar").

Consider the text

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.

SI is the first clause, S2 the second. From s1, we can infer, by axioms associated with "likely", that the probability of hitting the bull's eye (call it p.) is less than .5, or whatever other value counts as "likely". From S , we can infer that p. is greater than the typical probability of hitting those other equal areas (call it p.). Thus we have

SI': p < .5

S2': $p_I > p_2$

The predicates of SI' and S2' — "<" and ">" ~ are contradictory. The first arguments are identical — "p ". The second arguments — ".5" and "p " — are similar in that both are probabilities. Hence, the Contrast pattern is matched.

In the text

Research proper brings into play clockwork-like mechanisms; discovery has a magical essence.

"research" and "discovery" are viewed as similar elements and "mechanistic" and "magical" as lying at opposite ends of some scale.

4. Violated Expectation: This pattern holds if from S. one can <u>normally</u> infer S, but from s_2 one can infer <u>not</u> S. <u>Not</u> S is what we are expected to believe.

In

This paper is weak, but interesting

the first clause SI would lead us to infer that the paper is not suitable for presentation, but the second S2 contradicts this. It is the inference from the second clause we are to believe.

5. Paraphrase: It is frequent for stretches of text to be attempts at the successive approximation of a meaning, or attempts to overcome the possibility of misunderstanding. The pattern may be stated

From the assertions of S_1 and S_2, infer S_1 and S2, respectively, where S1 and S' are the same except that either

1) an argument of S' is more fully or precisely described than the corresponding argument of S', or
2) S' has adverbial modification S' lacks.

Consider the text

This immense tract of time is only sparsely illuminated by human relics. Not enough material has yet been found for us to trace the technical evolution of East Asia.

From our knowledge of "sparse" and "illuminate", we can infer from the first sentence S1 that the relics fail to cause one to know the "contents" of the immense tract of time (call this statement SI'). From "not enough" in the second sentence S2, we can infer that the material fails to cause us to know the "contents" of the technical evolution (call this S2'). S1' and S2 both have the form

fail(cause(A,know(we,B))), contents-of(E,C)

SI' describes A as "relics" while S2' uses "material", near synonyms. C, which in SI' is described as an "immense tract of time", is specified more precisely in S2' as "technical evolution of East Asia". Hence, the Paraphrase pattern is recognized.

6. Parallel: The pattern for this relation is

From the assertions of S1 and S , infer S1 and S', respectively, where the corresponding pairs of elements are identical or members of some small set.

At least one pair must be non-identical. This pettern is frequently keyed by similar syntactic structures.

In

Set stack A to empty, and set link variable P to ROOT.

the predicates and agents of both sentences are the same, A and P are both data structures and A being empty and P pointing to ROOT are both examples of plausible initial values.

4. <u>An Example</u>

In this section we examine one paragraph very briefly to see the discourse structure that is produced. A detailed analysis has been carried out by hand and is available in Hobbs (1976b). Implementation of the analysis is in progress. Unsurprisingly, it assumes state-of-the-art syntactic preprocessing. Semantic analysis requires a Lexicon of at least several hundred axioms encoding knowledge ranging from properties of basic spatial, temporal, and mental concepts to very specific facts about American politics. The chains of inference needed to establish some of the intersentence relations are quite long in some cases and it is not yet clear what can be expected from a computer implementation.

The paragraph comes from <u>Newsweek</u>. The clauses and sentences are labelled for subsequent reference.

[S :] All of this only deepened the despond of the President's own party, [S :] and so intensified his isolation in a time of maximum jeopardy. IS.:] The conviction is widespread among Republicans that [S_D:] time is running out on Operation Candor [S_:] —that Mr. Nixon must clear himself by early in the new year [S_F:] or lose whatever slipping hold he has left on the party. £$^S{}_G$'3 But they commonly doubt that the message is getting through to the President, [S :] and now their discouragement has been compounded by the news that [S :] Mr. Nixon's two sawiest political hands, Melvin Laird and Bryce Harlow, plan to quit as soon as Ford settles in. [^{S}j'I "When they go," mourned one Senate GOP leader, "we'll have no one there to talk to. I$^S{}_K$'3 We have nothing to say to Ron Ziegler, [S :] and Al Haig's never been in politics [S :] —he can't even spell the word 'vote'." (Newsweek, Dec. 17, 1973, p. 25).

Between S and S-, there is an explicit causality, keyed by "so", but it is unlikely that we can verify it. Therefore its verification is placed high on the goal list for the rest of the paragraph. It is frequent for a paragraph to provide verification for its topic sentence. S and S are linked by the "or" that introduces the cause of Nixon's obligation. Together the two are recognized as a Paraphrase of S , as follows: From S we infer that Operation Candor requires time to achieve its goal. Its goal is for Nixon to clear himself, which matches part of $S_£$. The remainder of the match is achieved by detecting the correspondence between "require time" and "must_____by early in the new year." The paraphrase is embedded in S , whose Expectation is Violated by S . The expectation is that normally when one is involved in a situation, he knows it. The stretch of text from S to S_Q matches the initial state of S — it describes the Republicans' state of discouragement. Consequently, the Temporal Succession pattern is high on the goal list for the next sentence, for that would allow a match with all of S . S matches the pattern and the result qualifies as an example of S , since discouragement being compounded matches despond deepening.

S is an Example of S because of the match between "one Senate GOP leader mourned" and "their discouragement". S and S might or might not be recognized as a Paraphrase r>y processes that won't be described here. To recognize this as being in Parallel with S requires us to infer from Haig's not being in politics that Haig does not engage in political activities, in particular, he does not confer with others to achieve common political goals. These sentences in turn constitute an example of the Senate GOP leader having no one there to talk to.

The resulting discourse structure is given in Figure 1

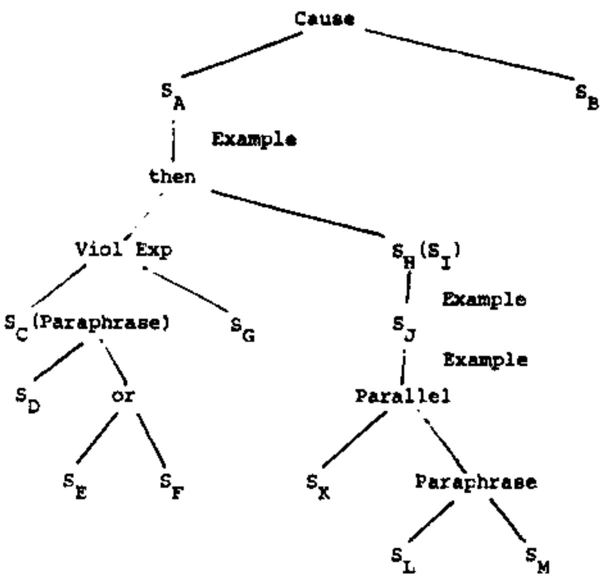

Figure 1

One thing that is apparent from this tree is that it is not complete. We still have not verified the causality between S and S_B. The verification is passed as a goal to the next paragraph, and in fact that is precisely what the next paragraph does.

Ultimately, we will want to be able to determine the message of the text as a whole, or what amounts to the same thing, construct a summary of it. This will require more than the methods of analysis presented in this paper can produce. Among other things it will require information about the redundancies in the paragraph and about the obstacles that are overcome in the paragraph on the way toward the author's or the characters' goals. But the structural information discovered by the intersentence operation and illustrated in Figure 1 will be a very important element in any such construction.

5. Conclusion

What has been described in this paper are several methods by which the structure and complex meaning of a text can be built block by block from the bottom up in a regular and parsimonious manner that can be characterized. They involve close analysis of the ways in which words and sentences interact. Although they assume a great amount of world knowledge, the knowledge is encoded in a rather simple representation and the methods stress the domain-independent ways in which the knowledge is used. Thus, our approach may be viewed as complementary to much current work in discourse analysis, where highly structured knowledge of the situation being described is assumed and guides the analysis of the text. Our approach consequently represents a step toward an accommodation of the richness, subtlety, and flexibility of language that make it interesting.