An Approach to the Structure of Discourse

Jerry R. Hobbs
Artificial Intelligence Center
SRI International
Menlo Park, California

1 Structure from Adjacency

To understand our environment we seek the best explanation for the observable facts we find there. Similarly, to interpret texts we seek the best explanation for the "observable" facts that are presented in the text. This view can be cashed out computationally by taking the interpretation of a text to be the most economic abductive proof of the logical form of the sentences in the text, where "abductive" means that assumptions are allowed at various costs and "economic" means that these costs are minimized (Hobbs et al., 1993).

Among the observable features of our environment that we seek to explain is the adjacency or proximity of objects; this generally escapes our notice except when it is out of the ordinary, as when we see a chair on top of a table or a dog in a classroom. Normally the explanation for something being in the place it is is that that is its place.

A similar situation obtains in language. A text is a string of words, and one of the features of the text that requires explanation is the adjacency of pairs of words or larger segments of text.

The simplest example of this is provided by compound nominals. When we see the phrase "turpentine jar" in a text, the interpretation problem we face is finding the most reasonable relationship in the context between turpentine and jars, using what we know about turpentine and jars. In many cases, the relationship is one arising out of one of the nouns itself, as in "oil sample", where the relation between oil and the sample is precisely the "sample of" relation. "Oil" provides one of the arguments of the predicate "sample".

Syntax and compositional semantics can be seen as arising out of the same need to explain adjacency. When we see the pair of words "men work", we need to find some relation between them. The hypothesis that sentences have syntactic structure amounts to the acceptance of a set of constraints concerning the relation that can obtain between two words or larger stretches of text. In this case, the constraint is that the second word itself provides the relationship. The men have to be the agent of the working. Whereas in the case of "oil sample" "sample" provides a possible relationship, in the case of "men work" "work" provides the obligatory relationship. (This is not quite true; metonymy is possible, so that the second word need only provide a relationship between the event and something functionally related to the the first word.)[1]

[1] This view of syntax emerged from a conversation with Mark Johnson.

The tree structure of sentences arises from the fact that the adjacency relation can be between larger segments of text than simply single words, where the segments have their own internal tree structure resulting from adjacencies. For example, in

John believes men work.

we don't seek to explain the adjacency between "believes" and "men". Rather we first explain the adjacency between "men" and "work", and only then the adjacency between "believes" and "men work" (or the adjacency among "John", "believes", and "men work", depending on your view of the structure of the clause.) This grouping occurs even in the absence of syntactic constraints. Consider the two compound nominals, "Stanford Research Institute" and "Cancer Research Institute". In the latter, we must first find the relationship between cancer and research, and then find the relationship between cancer research and the institute, whereas in the former, we group "Research" with "Institute" and then "Stanford" with "Research Institute".

In order to explain adjacencies between segments of text larger than one word, we need to have an idea of the principal information conveyed by the segments. For example, in "research institute" the reference to the institute is, in a sense, more primary than the reference to research. A research institute is a kind of institute, rather than a kind of research. Similarly, in "men work" the information about some entities working is more primary than the information about the entities being men. Therefore, as we compose larger and larger segments of text, we must have some sense of the primary information conveyed by the composite segments.

The rules of syntax and compositional semantics are a set of constraints on how segments of text can be grouped together and on what the primary information conveyed by the composite segment is. At the level of the main clause, the primary information is often what is conveyed by the main verb and/or by adverbials, although this can be overridden by such factors as intonation, newness, topicalization, and so on. We may call this primary information, however it is determined, the *assertion* of the clause. It is, in a sense, a summary of what the sentence conveys.

The source of discourse structure is precisely the same as the source of intrasentential syntactic structure. Two phrases, or clauses, or sentences, or larger stretches of discourse are adjacent in the discourse, and this fact requires explanation. A relation between them must be found to explain that adjacency. Whereas in the syntactic structure of sentences the relation among adjacent elements is most commonly a predicate-argument relation, the case of larger stretches of unrestricted text is more like the case of compound nominals, in that the relation that explains the adjacency can in principle be any plausible relation between the situations described by the segments of text.

In fact, although I am arguing for a seamless transition from syntax to discourse, if one were interested in defining the minimal discourse unit (Polanyi, 1988), a good choice would be that unit above which predicate-argument relations are no longer the dominant interpretation of adjacency. In written text this tends to be clauses or sentences; in spoken discourse it is often phrasal or smaller elements.

As in the case of syntax, while we compose larger and larger segments of text, we must have some sense of the primary information conveyed by the composite segment. We must be able to specify the "assertion" or "summary" of a supraclausal segment of text. This is both harder and easier than in the case of compound nominals. It is harder in that whereas it is almost always the second noun of a noun-noun pair that is primary, with two supraclausal segments it may be the first or the second, or the primary information may arise equally out of both.

It is easier in that a smaller number of relations can typically obtain between two situations or eventualities described in supraclausal segments than can obtain between two nouns. For compound nominals, Downing (1977) and others have convincingly argued that the relation between the two nouns can be virtually anything, given the right context. On the other hand, Levi (1978) argues convincingly that the vast majority of the relations can be viewed as instances of no more than a dozen or so different abstract relations, such as predicate-argument, function, containment, and so on.

Similarly, it is possible that the relation between two adjacent supraclausal segments of text can be anything at all in the right context. The hearer must simply figure out the most plausible relation between the situations described and the most plausible assertion or summary of the composite segment. Such a view of discourse coherence, unfortunately, gives no guidance as to what the assertion or summary of the composite segment is.

Overwhelmingly, however, the relation between supraclausal segments can be viewed as an instance of one of three broadly construed abstract relations—causality, the figure-ground relation, and similarity. I will refer to these as *coherence relations*. A theory of discourse coherence and discourse structure that recognizes this fact about discourse must develop characterizations of each of these relations, explicate the various classes of instances of each relation, and for each of these classes, define the assertion or summary of the composite segment. This is what I have tried to do in previous work (e.g., Hobbs, 1978, 1985) and what I will try to recast into an abductive framework in this paper.

While I will focus on the way these three relations relate supraclausal segments of text, they can also relate material within single clauses. For example, elements of a list exhibit the similarity required of parallelism (cf. Polanyi, 1988). In the sentence,

A car hit a jogger in Palo Alto last night.

there is an implicit causal relation between the jogging and the hitting. These relations go beyond what is given to us by compositional semantics.

While I have focused, and will continue to, on the *interpretation* problem, it is important to keep in mind that interpretation and generation are intricately interrelated. A speaker seeks to generate utterances that will be understood. When two segments of coherent discourse are uttered in sequence, it is because the speaker expects the hearer to recover the relation that is intended to be conveyed by the adjacency. Conversely, the hearer must often reason about what the speaker is trying to achieve and the other ways in which the speaker might have chosen to achieve it, in order to determine the best interpretation for a stretch of discourse.

Throughout this paper, I adopt the informational perspective on discourse, in which the information conveyed by the discourse is for the most part taken at face value. In fact, interpretations derived as described here are *entertained* rather than *believed*; belief revision can be thought of as a separate process, logically after interpretation. Moreover, the conveying of information in discourse is intentional action. In Hobbs (1995a) I show how the informational perspective can be embedded in an abductive intentional perspective and how the two perspectives can interact in interpretation.

2 Background: Interpretation as Abduction

2.1 Solving Local Pragmatics Problems by Abduction

Abduction is inference to the best explanation. The process of interpreting sentences in discourse can be viewed as the process of providing the best explanation of the information conveyed by the sentences, that is, the best explanation of why the sentences would be true. This idea can be cashed out procedurally as follows:

To interpret a sentence:

(1) Prove the logical form of the sentence,
together with the constraints that predicates impose on their arguments,
allowing for coercions,
Merging redundancies where possible,
Making assumptions where necessary.

By the first line we mean "prove, or derive in the logical sense, from the predicate calculus axioms in the knowledge base, the logical form that has been produced by syntactic analysis and semantic translation of the sentence."

In a discourse situation, the speaker and hearer both have their sets of private beliefs, and there is a large overlapping set of mutual beliefs. An utterance lives on the boundary between mutual belief and the speaker's private beliefs. It is a bid to extend the area of mutual belief to include some private beliefs of the speaker's. It is anchored referentially in mutual belief, and when we succeed in proving the logical form and the constraints, we are recognizing this referential anchor. This is the given information, the definite, the presupposed. Where it is necessary to make assumptions, the information comes from the speaker's private beliefs, and hence is the new information, the indefinite, the asserted. Merging redundancies is a way of getting a minimal, and hence a best, interpretation.

Consider a simple example.

(2) The Boston office called.

This sentence poses at least three problems in local pragmatics, the problems of resolving the reference of "the Boston office", expanding the metonymy to "[Some person at] the Boston office called", and determining the implicit relation between Boston and the office. Let us put these problems aside for the moment, however, and interpret the sentence according to characterization (1). We must prove abductively the logical form of the sentence together with the constraint "call" imposes on its agent, allowing for a coercion. That is, we must prove abductively the expression (ignoring tense and some other complexities)

(3) $(\exists\, x, y, z, e) call'(e, x) \wedge person(x) \wedge rel(x, y) \wedge office(y) \wedge Boston(z)$
$\wedge\, nn(z, y)$

That is, there is a calling event e by x where x is a person. x may or may not be the same as the explicit subject of the sentence, but it is at least related to it, or coercible from it, represented by $rel(x, y)$. y is an office and it bears some unspecified relation nn to z which is Boston. $person(x)$ is the requirement that $call'$ imposes on its agent x.

The sentence can be interpreted with respect to a knowledge base of mutual knowledge[2] that contains the following facts:

$Boston(B_1)$

that is, B_1 is the city of Boston.

$office(O_1) \wedge in(O_1, B_1)$

that is, O_1 is an office and is in Boston.

$person(J_1)$

that is, John J_1 is a person.

$work\text{-}for(J_1, O_1)$

that is, John J_1 works for the office O_1.

$(\forall\, y, z) in(y, z) \supset nn(z, y)$

that is, if y is in z, then z and y are in a possible compound nominal relation.

$(\forall\, x, y) work\text{-}for(x, y) \supset rel(x, y)$

that is, if x works for y, then y can be coerced into x.

The proof of all of (3) is straightforward except for the conjunct $call'(x)$. Hence, we assume that; it is the new information conveyed by the sentence.

This interpretation is illustrated in the proof graph of Figure 1, where a rectangle is drawn around the assumed literal $call'(e, x)$. Such proof graphs play the same role in interpretation as parse trees play in syntactic analysis. They are pictures of the interpretations, and we will see a number of such diagrams in this paper.

Now notice that the three local pragmatics problems have been solved as a by-product. We have resolved "the Boston office" to O_1. We have determined the implicit relation in the compound nominal to be in. And we have expanded the metonymy to "John, who works for the Boston office, called."

Other proofs are possible for (3). For example, we could have assumed $person(x)$, rather than proving it by unifying it with $person(J_1)$. This proof corresponds to the less specific interpretation that *someone* who works at the Boston office called.

[2] Throughout this article it will be assumed that all axioms are mutually known by the speaker and hearer, that they are part of the common cultural background

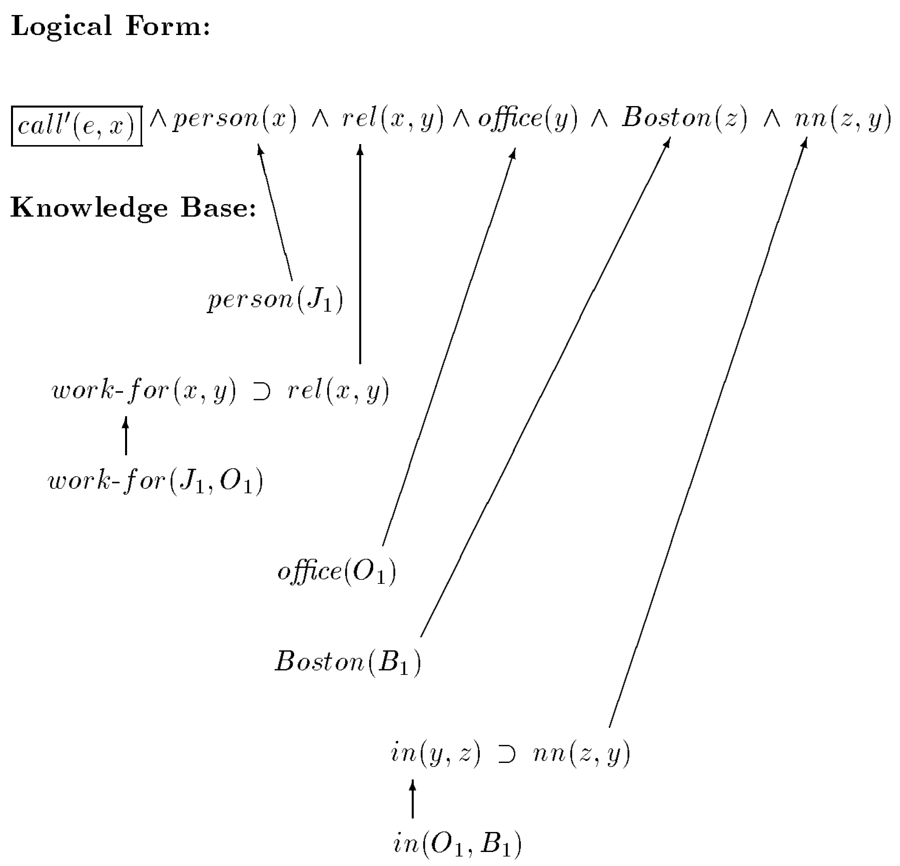

Figure 1: Interpretation of "The Boston office called."

In general, there must be a method of choosing among possible interpretatons. A cost-based scheme is presented in Hobbs et al. (1993), in which every proof is given a cost and the best interpretation is that given by the lowest cost proof. No more will be said about this method in this paper. We will be content to show that the desired interpretations are in the space of possible interpretations, without arguing that they are the best interpretations in that space.

2.2 Integration of Syntax, Compositional Semantics, and Local Pragmatics

By combining the idea of interpretation as abduction with the older idea of parsing as deduction (Kowalski, 1980, pp. 52-53; Pereira and Warren, 1983), it becomes possible to integrate syntax, semantics, and pragmatics in a very thorough and elegant way.[3]

We will present this in terms of example (2), repeated here for convenience.

[3] This idea is due to Stuart Shieber.

(2) The Boston office called.

Recall that to interpret this we must prove the expression

(3a) $(\exists\, x, y, z, e) call'(e, x) \wedge person(x) \wedge rel(x, y)$
(3b) $\wedge\, office(y) \wedge Boston(z) \wedge nn(z, y)$

Consider now a simple grammar, adequate for parsing this sentence, written in Prolog style:

$$(\forall\, w_1, w_2) np(w_1) \wedge verb(w_2) \supset s(w_1\ w_2)$$
$$(\forall\, w_1, w_2) det(the) \wedge noun(w_1) \wedge noun(w_2) \supset np(the\ w_1\ w_2)$$

That is, if string w_1 is a noun phrase and string w_2 is a verb, then the concatenation w_1 w_2 is a sentence. The second rule is interpreted similarly. To parse a sentence W is to prove $s(W)$.

We can integrate syntax, compositional semantics, and local pragmatics by augmenting the axioms of this grammar with portions of the logical form in the appropriate places, as follows:

(4) $(\forall\, w_1, w_2, y, p, e, x) np(w_1, y) \wedge verb(w_2, p) \wedge p'(e, x) \wedge rel(x, y) \wedge Req(p, x)$
$\supset s(w_1\ w_2, e)$

(5) $(\forall\, w_1, w_2, q, r, y, z) det(the) \wedge noun(w_1, r) \wedge noun(w_2, q)$
$\wedge\, r(z) \wedge q(y) \wedge nn(z, y) \supset np(the\ w_1\ w_2, y)$

The second arguments of the "lexical" predicates *noun* and *verb* denote the predicates corresponding to the words, such as *Boston*, *office* or *call*. The atomic formula $np(w_1, y)$ means that the string w_1 is a noun phrase referring to y. The atomic formula $Req(p, x)$ stands for the requirements that the predicate p places on its argument x. The specific constraint can then be enforced if there is an axiom

$$(\forall\, x) person(x) \supset Req(call, x)$$

that says that one way for the requirements to be satisfied is for x to be a person. Axiom (4) can then be paraphrased as follows: "If w_1 is a noun phrase referring to y, and w_2 is a verb denoting the predicate p, and p' is true of some eventuality e and some entity x, and x is related to (or coercible from) y, and x satisfies the requirements p' places on its second argument, then the concatenation w_1 w_2 is a sentence describing eventuality e." Axiom (5) can be paraphrased as follows: "If *the* is a determiner, and w_1 is a noun denoting the predicate r, and w_2 is a noun denoting the predicate q, and the predicate r is true of some entity z, and the predicate q is true of some entity y, and there is some implicit relation nn between z and y, then the concatenation *the* w_1 w_2 is a noun phrase referring to the entity y." Note that the conjuncts from line (3a) in the logical form have been incorporated into axiom (4) and the conjuncts from line (3b) into axiom (5).[4]

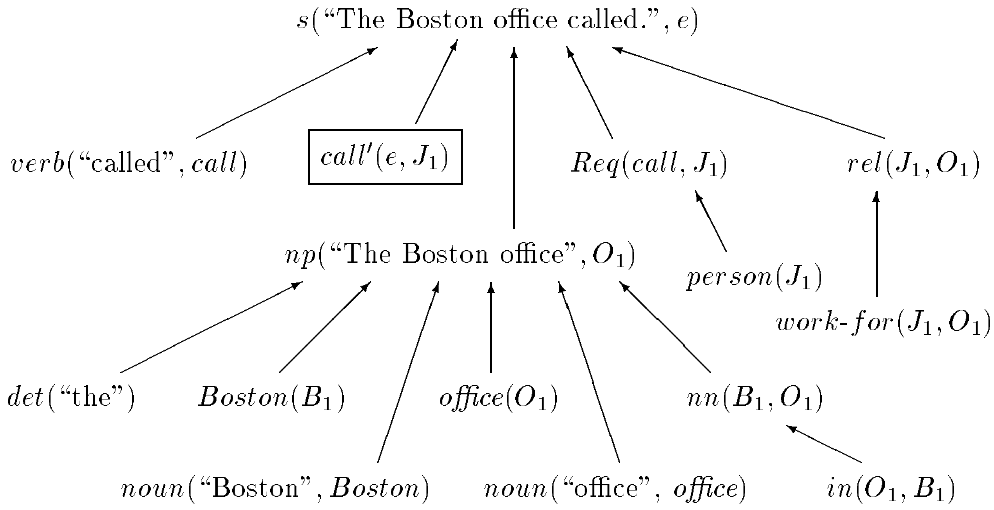

Figure 2: Parse and interpretation of "The Boston office called."

The parse and interpretation of sentence (2) is illustrated in Figure 2.

Before when we proved $s(W)$, we proved that W was a sentence. Now, if we prove $(\exists\, e)s(W,e)$, we prove that W is an *interpretable* sentence and that the eventuality e is its interpretation.

Each axiom in the "grammar" then has a "syntactic" part—the conjuncts like $np(w_1, y)$ and $verb(w_2, p)$—that specifies the syntactic structure, and a "pragmatic" part—the conjuncts like $p'(e,x)$ and $rel(x,y)$—that drives the interpretation. That is, local pragmatics is captured by virtue of the fact that in order to prove $(\exists\, e)s(W,e)$, one must derive the logical form of the sentence together with the constraints predicates impose on their arguments, allowing for metonymy, i.e., characterization (1). The compositional semantics of the sentence is specified by the way the denotations given in the syntactic part are used in the construction of the pragmatic part.

Note that the structure of both axioms (4) and (5) is roughly this: To prove that a whole is a meaningful entity, prove each of its parts are meaningful entities and that there is a meaningful relation between them. In (4) the relation is the coerced predicate-argument relation $p'(e,x) \wedge rel(x,y)$. In (5) it is the nn relation.

In this framework, interpretation and generation differ only in the initial and final conditions of the proof process. To interpret a sentence W is to prove $(\exists\, e)\, s(W,e)$. The resulting e is the new information conveyed. In generation, the eventuality E is known and the sentence must be derived. That is, one must prove $(\exists\, w)\, s(w,E)$. This idea is developed further in Hobbs et al. (1990) and in Hobbs and Kameyama (1990).

A much more extensive and somewhat different account of the syntax of English in an "Interpretation as Abduction" framework is given in Hobbs (1995b); it follows Head-driven

[4] As given, these axioms are second-order, but not seriously so, since the predicate variables only need to be instantiated to predicate constants, never to lambda expressions. It is thus easy to convert them to first-order axioms by having an individual constant corresponding to every predicate constant.

Phrase Structure Grammar (Pollard and Sag, 1994) fairly closely.

2.3 Notational Conventions

Before proceeding, I would like to introduce some notational conventions that are used in this paper.

The first was already slipped in above. We will take $p(x)$ to mean that p is true of x, and $p'(e,x)$ to mean that e is the eventuality or possible situation of p being true of x. This eventuality may or may not exist in the real world. The unprimed and primed predicates are related by the axiom schema

$$(\forall\, x)p(x) \equiv (\exists\, e)p'(e,x) \wedge Rexists(e)$$

where $Rexists(e)$ says that the eventuality e does in fact really exist. This notation, by reifying events and conditions, provides a way of specifying higher-order properties in first-order logic. This Davidsonian reification of eventualities (Davidson, 1967) is a common device in AI. See Hobbs (1985) for further explanation of the specific notation and ontological assumptions.

Often axioms that intuitively ought to be written as

$$(\forall\, x)p(x) \supset q(x)$$

will be written

$$(\forall\, e_1, x)p'(e_1, x) \supset (\exists\, e_2)q'(e_2, x)$$

That is, if e_1 is the eventuality of p being true of x, then there is an eventuality e_2 of q being true of x. It will sometimes be convenient to state this in a stronger form. It is not just that if e_1 exists, then e_2 happens to exist as well. The eventuality e_2 exists *by virtue of the fact that* e_1 exists. Let us express this tight connection by the predicate gen, for "generates". Then the above axiom can be strengthened to

$$(\forall\, e_1, x)p'(e_1, x) \supset (\exists\, e_2)q'(e_2, x) \wedge gen(e_1, e_2)$$

Not only is there an e_2, but there an e_2 by virture of the fact that there is an e_1. The relative existential and modal statuses of e_1 and e_2 can then be axiomatized in terms of the predicate gen. This is the second notational convention.

Third, it might seem in abduction that since we use only backchaining to find a proof and a set of assumptions, we cannot use superset information. However, the fact that we can make assumptions enables us to turn axioms around. In general, an axiom of the form

$$\text{species} \supset \text{genus}$$

can be converted into a biconditional axiom of the form

$$\text{genus} \wedge \text{differentiae} \equiv \text{species}$$

Often we will not be able to prove the differentiae, and in many cases we cannot even spell them out. But in our abductive scheme, this does not matter; they can simply be assumed. In fact, we need not state them explicitly at all. We can simply introduce a predicate, a different one for each axiom, that stands for all the remaining properties. It will never be provable, but it will be assumable. Thus, in addition to having axioms like

$$(\forall\, y) elephant(y) \supset clumsy(y)$$

we may have axioms like

$$(\forall\, x) clumsy(y) \wedge etc_1(y) \supset elephant(y)$$

Then, even though we are strictly backchaining in search for an explanation, the fact that something is clumsy can still be used as (perhaps weak) evidence for its being an elephant, since we can assume the "et cetera" predication $etc_1(x)$ for a certain cost.

This device may seem ad hoc at first blush. But I view the device as implementing a fairly general solution to the problems of nonmonotonicity in commonsense reasoning and vagueness of meaning in natural language, very similar to the use of abnormality predicates in circumscriptive logic (McCarthy, 1987). Whereas, in circumscriptive logic, one typically specificies a partial ordering of abnormality predicates in accordance with which they are minimized, in the weighted abduction framework, one uses a somewhat more flexible system of costs.

There is no particular difficulty in specifying a semantics for the "et cetera" predicates. Formally, etc_1 in the axiom above can be taken to denote the set of all things that are either not clumsy or are clumsy elephants. Intuitively, etc_1 conveys all the information one would need to know beyond clumsiness to conclude something is an elephant. As with nearly every predicate in an axiomatization of commonsense knowledge, it is hopeless to spell out necessary and sufficient conditions for an "et cetera" predicate. In fact, the use of such predicates in general is due largely to a recognition of this fact about commonsense knowledge.

The "et cetera" predicates could be used as the abnormality predicates are in circumscriptive logic, with separate axioms spelling out conditions under which they would hold. However, in the view adopted here, more detailed conditions would be spelled out by expanding axioms of the form

$$(\forall\, x) p_1(x) \wedge etc_1(x) \supset q(x)$$

to axioms of the form

$$(\forall\, x) p_1(x) \wedge p_2(x) \wedge etc_1(x) \supset q(x)$$

An "et cetera" predicate would appear only in the antecedent of a single axiom and never in a consequent. Thus, the "et cetera" predications are only place-holders for assumption costs. They are never proved. They are only assumed.

They constitute one of the principal devices for giving our logic "soft corners". We would expect them to pervade the knowledge base. Virtually any time there is an axiom relating a species to a genus, there should be a corresponding axiom, incorporating an "et cetera" predication, expressing the inverse relation.

Let us summarize at this point the most elaborate form axioms in the knowledge base will have. If we wish to express an implicative relation between concepts p and q, the most natural way to do so is as the axiom

$$(\forall\, x, z) p(x, z) \supset (\exists\, y) q(x, y)$$

where z and y stand for arguments that occur in one predication but not in the other. When we introduce eventualities, this axiom becomes

$$(\forall\, e_1, x, z) p'(e_1, x, z) \supset (\exists\, e_2, y) q'(e_2, x, y)$$

Using the *gen* relation to express the tight connection between the two eventualities, the axiom becomes

$$(\forall\, e_1, x, z) p'(e_1, x, z) \supset (\exists\, e_2, y) q'(e_2, x, y) \wedge gen(e_1, e_2)$$

Next we introduce an "et cetera" proposition into the antecedent to take care of the imprecision of our knowledge of the implicative relation.

$$(\forall\, e_1, x, z) p'(e_1, x, z) \wedge etc_1(x, z) \supset (\exists\, e_2, y) q'(e_2, x, y) \wedge gen(e_1, e_2)$$

Finally we biconditionalize the relation between p and q by writing the converse axiom as well:

$$(\forall\, e_1, x, z) p'(e_1, x, z) \wedge etc_1(x, z) \supset (\exists\, e_2, y) q'(e_2, x, y) \wedge gen(e_1, e_2)$$
$$(\forall\, e_1, x, y) q'(e_2, x, y) \wedge etc_2(x, y) \supset (\exists\, e_1, z) p'(e_1, x, z) \wedge gen(e_2, e_1)$$

This then is the most general formal expression in our abductive logic of what is intuitively felt to be an *association* between the concepts p and q.

In this article, for notational convenience, we will use the simplest form of axiom we can get away with for the example. The reader should keep in mind however that these are only abbreviations for the full, biconditionalized form of the axiom.[5]

3 Axiomatizing the Tree-like Structure of Discourse

The tree-like structure of discourse can be captured with two axioms:

$$(\forall\, w, e) S(w, e) \supset Segment(w, e)$$
$$(\forall\, w_1, e_1, w_2, e_2, e) Segment(w_1, e_1) \wedge Segment(w_2, e_2) \wedge CoherenceRel(e_1, e_2, e) \\ \supset Segment(w_1 w_2, e)$$

[5]The full axioms are non-Horn, but not seriously so. They can be Skolemized and broken into two axioms having the same Skolem functions. This remark holds as well for other axioms in this article that have conjunctions in the consequent.

The first axiom says that a sentence w describing an eventuality e is a coherent discourse segment describing e. The second says that if two segments w_1 and w_2 describe the eventualities e_1 and e_2, respectively, and e_1 and e_2 are related by some coherence relation, then the concatenation w_1w_2 is a coherent discourse segment.

The variable e in the second axiom is the assertion or summary of the composite segment w_1w_2. It is determined by the assertions of the constituent segments, e_1 and e_2, together with the relation that holds between them by virtue of which w_1w_2 is itself a coherent discourse segment.

To prove $CoherenceRel(e_1, e_2, e)$ is to explain the adjacency of w_1 and w_2.

To interpret a text W is then to prove the expression

$$(\exists\, e) Segment(W, e)$$

meaning that W is a coherent segment of discourse conveying or describing the situation or eventuality e. e is the assertion or summary of W.

To explicate a theory of discourse coherence and discourse structure along these lines is to specify the various ways in which

$$CoherenceRel(e_1, e_2, e)$$

can be established, including what e is.

The common distinction between hypotactic coherence relations, with dominant and subordinate component segments, and paratactic coherence relations is easily captured in this framework. If the relation is hypotactic, then e is either e_1 or e_2, corresponding to whether w_1 or w_2 is dominant. If the coherence relation is paratactic, then e must be computed from e_1 and e_2 together. This in fact is what hypotactic and paratactic mean.

This approach has the flavor of discourse grammar approaches. What has always been the problem with discourse grammars is that their terminal symbols (e.g., Introduction) and sometimes their compositions have not been computable. Because in our abductive, inferential approach, we are able to reason about the content of the utterances of the discourse, this problem no longer exists.

4 The Coherence Relations

4.1 The Sources of Coherence

At a sufficiently abstract level, when the relations between successive clauses or sentences in text are not explicitly indicated, there are three relations that obtain overwhelmingly—causality, figure-ground, and similarity. It should not be surprising that these relations are so salient. It is obvious why causality would be of interest to creatures like us, that have to maneuver our way among events beyond our control; prediction promotes survival. Our interest in figure-ground relations and similarity may reduce to causality as well. An entity (the figure) is causally influenced by the environment (the ground) in which it is located, and similar entities behave causally in a similar fashion (and when they don't, it is worthy of note). Thus, knowing these relations aids prediction.

One could argue that this observation is originally due to Hume. In his *Inquiry Concerning Human Understanding* (Section III), he argued that there are general principles of coherent discourse resting upon general principles for the association of ideas. "Were the loosest and freest conversation to be transcribed, there would immediately be observed something, which connected it in all its transitions. Or where this is wanting, the person, who broke the thread of discourse, might still inform you, that there had secretly revolved in his mind a succession of thought, which had gradually led him from the subject of conversation." Moreover, the three principles he proposed are very close to our own principles of causality, figure-ground, and similarity: "To me, there appear to be only three principles of connexion among ideas, namely, *Resemblance*, *Contiguity* in time or place, and *Cause* or *Effect*."

In the remainder of this section, I describe those coherence relations that happen to be required by the fragment of conversation analyzed in Section 6: Causality, Parallelism, Elaboration, and Contrast. Axioms are given defining each coherence relation, and examples, together with the relevant inferences, are given.

In the examples we will see a number of other linguistic interpretation problems, such as pronoun resolution and the interpretation of a metaphor, being solved as a by-product of recognizing coherence relations.

4.2 Causality

Causality can relate two segments of discourse in one of two ways. We can describe an event or situation and then describe a causal consequence; here the cause comes first and the effect second, and either one can be the dominant segment, although generally the effect is. Or we can describe a situation and then describe, as an explanation, the situation that gave rise to it; here the effect comes first and the cause second, and the effect is the dominant member of the pair. I will give a detailed example of the explanation relation.

For one segment of discourse S_2 to function as an explanation of another segment S_1, S_2 must describe a state or event that causes or could cause the state or event described in S_1. The following axiom expresses this (or is at least a first approximation):

$$(\forall\, e_1, e_2) cause(e_2, e_1) \supset \; Explanation(e_1, e_2)$$

That is, if what is asserted by the second segment could cause what is asserted by the first segment, then there is an Explanation relation between the segments. Explanation is generally hypotactic. The explanandum is dominant and thus contributes the assertion or summary; the explanans is subordinate. The following axiom captures this.

$$(\forall\, e_1, e_2) Explanation(e_1, e_2) \supset \; CoherenceRel(e_1, e_2, e_1)$$

As an example of the Explanation relation, consider a variation on the classic example from Winograd (1972):

The police prohibited the women from demonstrating.
They feared violence.

To interpret the text is to prove abductively the expression

$$Segment(\text{“The police ... violence.”},\ e)$$

This involves proving that each sentence is a segment, by proving they are sentences, and proving there is a coherence relation between them. To prove they are sentences, we would tap into an expanded version of the sentence grammar of Section 2.2. This would require us to prove abductively the logical form of the sentences.

One way to prove there is a coherence relation between the sentences is to prove there is an Explanation relation between them, and one way to prove that is to prove a causal relation between their assertions.

After back-chaining in this manner, we are faced with proving the expression

$$\begin{aligned}&(\exists\, e_1, p, d, w, e_2, y, v, z) prohibit'(e_1, p, d) \wedge demonstrate'(d, w) \wedge cause(e_2, e_1)\\ &\qquad \wedge fear'(e_2, y, v) \wedge violent'(v, z)\end{aligned}$$

That is, there is a prohibiting event e_1 by the police p of a demonstrating event d by the women w. There is a fearing event e_2 by someone y (“they”) of violence v by someone z. The fearing event e_2 causes the prohibiting event e_1. This expression is just the logical forms of the two sentences, plus the hypothesized causal relation between them.

Suppose, plausibly enough, we have the following axioms:

$$(\forall\, e_2, y, v) fear'(e_2, y, v) \supset (\exists\, d_2) diswant'(d_2, y, v) \wedge cause(e_2, d_2)$$

That is, if e_2 is a fearing by y of v, then that will cause the state d_2 of y not wanting or “diswanting” v.

$$(\forall\, d, w) demonstrate'(d, w) \supset (\exists\, v, z) cause(d, v) \wedge violent'(v, z)$$

That is, demonstrations cause violence.

$$(\forall\, d, v, d_2, y) cause(d, v) \wedge diswant'(d_2, y, v) \supset (\exists\, d_1) diswant'(d_1, y, d) \wedge cause(d_2, d_1)$$

That is, if someone p diswants v and v is caused by d, then that will cause p to diswant d as well. If you don't want the effect, you don't want the cause.

$$(\forall\, d_1, p, d) diswant'(d_1, p, d) \wedge authority(p) \supset (\exists\, e_1) prohibit'(e_1, p, d) \wedge cause(d_1, e_1)$$

That is, if those in authority diswant something, that will cause them to prohibit it.

$$(\forall\, e_1, e_2, e_3) cause(e_1, e_2) \wedge cause(e_2, e_3) \supset cause(e_1, e_3)$$

That is, *cause* is transitive.

$$(\forall\, p) police(p) \supset authority(p)$$

That is, the police are in authority.

From these axioms, we can prove all of the above logical form except the propositions $police(p)$, $demonstrate'(d, w)$, and $fear'(f, y, v)$, which we assume. This is illustrated in Figure 3. Notice that in the course of doing the proof, we unify y with p, thus resolving

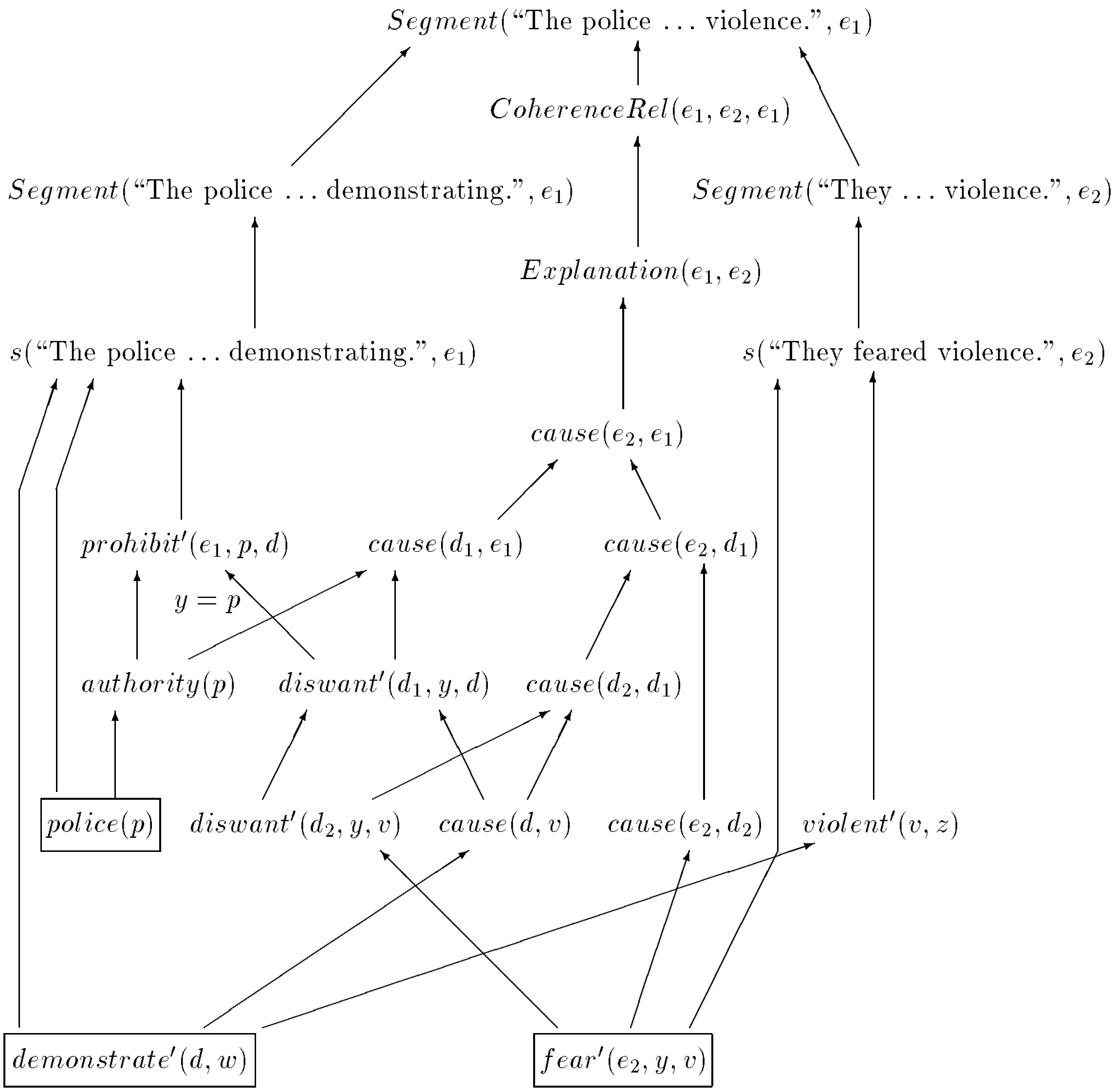

Figure 3: Interpretation of "The police prohibited the women from demonstrating. They feared violence."

	Specific to Specific	Specific to General	General to Specific
Positive:	Parallel	Generalization	Exemplification
Negative:	Contrast	–	–

Figure 4: Coherence Relations Based on Similarity

the problematic pronoun reference that originally motivated this example. "They" refers to the police.

One can imagine a number of variations on this example. If we had not included the axiom that demonstrations cause violence, we would have had to assume the violence and the causal relation between demonstrations and violence. Moreover, other coherence relations might be imagined here by constructing the surrounding context in the right way. It could be followed by the sentence "But since they had never demonstrated before, they did not know that violence might result." In this case, the second sentence would play a subordinate role to the third, forcing the resolution of "they" to the women. Each example, of course, has to be analyzed on its own, and changing the example changes the analysis. In Winograd's original version of this example,

> The police prohibited the women from demonstrating, because they feared violence.

the causality was explicit, thus eliminating the coherence relation as a source of ambiguity. The literal $cause(e_2, e_1)$ would be part of the logical form.

4.3 The Similarity Relations

Another class of coherence relations is based on similarity. These relations, in a sense, expand the discourse in place, rather than carrying it forward or filling in background. They can be classified in terms of moves between specific and general assertions and the interaction of these moves with negation, as illustrated in Figure 4.

I have left two blank spaces in the "Negative" row because such relations would constitute a contradiction. They might be filled in with an "Exception" relation. One states a general truth and then gives a specific exception to it, or vice versa. But I have chosen rather arbitrarily to consider these as examples of Contrast.

There are two important limiting cases. The Elaboration relation is a limiting case of the Parallel relation; the Violated Expectation relation is a limiting case of Contrast.

In the fragment of conversation analyzed in Section 6, only the Parallel, Elaboration, and Contrast relations occur, so only these will be explicated here.

4.4 The Parallel Relation

The definition of the Parallel relation involves recognizing the similarity of entities. The text segments must assert the same properties of similar entities. More precisely, the Parallel relation obtains when from the assertion of the first segment S_1 we can infer $p(x_1, x_2, \ldots)$ and from the assertion of the second segment S_2 we can infer $p(y_1, y_2, \ldots)$, where x_i and y_i are similar, for all i. Two entities are *similar* if they share some (reasonably specific) property. Determinations of similarity are subject to the same fuzziness and considerations of "good-ness" as the coherence relations in general.

A first approximation of an axiom that captures this characterization is the following:

$$(\forall\, p, q, x, y, e_1, e_2, e_3, e_4) p'(e_3, x) \,\wedge\, p'(e_4, y) \,\wedge\, gen(e_1, e_3) \,\wedge\, gen(e_2, e_4) \\ \quad \wedge\, q(x) \,\wedge\, q(y) \,\supset\, Parallel(e_1, e_2, e)$$

That is, a Parallel relation holds between e_1 and e_2 if there is a p such that e_3 is p's being true of some x, where e_1 generates e_3, and e_4 is p's being true of some y, where e_2 generates e_4, and there is some q true of x and y.

The Parallel relation is a coherence relation:

$$(\forall\, e_1, e_2) Parallel(e_1, e_2, e) \,\supset\, CoherenceRel(e_1, e_2, e)$$

The relation is paratactic. I have not specified in the axiom what the assertion of the composite segment would be, but it should be the generalization of which the assertion of each segment is an instance. The generalization may subsume only the x's and y's, or it may include all entities for which q is true.

A simple example of the Parallel relation is this sentence from an algorithm description:

> Set stack `A` empty and set link variable `P` to `T`.

From each of the clauses one can infer (trivially) that a data structure is being set to a value. The predicate p is thus *set*, stack `A` and link variable `P` are similar in that they are both data structures, and the stack's emptiness and `P`'s being equal to `T` are both initial conditions.

The next example is a bit more indirect. It comes from a problem in a physics textbook.

> 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.

Because of the nature of the task, the reader must draw inferences from this sentence about the relevant forces. We might represent the inferences as follows:

$$force(100\,\text{lb}, L, Down, x_1) \,\wedge\, distance(F, x_1, 20\,\text{ft}) \,\wedge\, foot(F, L) \\ force(150\,\text{lb}, x, Down, x_2) \,\wedge\, distance(T, x_2, 10\,\text{ft}) \,\wedge\, top(T, y)$$

Here the predicate p is $force$, the first arguments are similar in that they are both weights, the second and third arguments are both identical (once we identify x with L), hence similar, and the fourth arguments are similar in that they are points on the ladder at certain distances from an end of the ladder (assuming y is L). The assertion of the

BODY MATERIAL	CONTAINS	CONCENTRATION	AGENT
blood	contains	highest concentration	HBV
semen vaginal secretions menstrual blood	contain		agent
saliva	has	lower concentrations	
(saliva of) infected individuals	in	detectable ... no more than half	HBsAg
urine	contains	low concentrations	

Figure 5: The *Parallel* Relation in Example (6).

composite segment is something like "There are downward forces acting on the ladder at some distance from an end of the ladder", although this information alone would not be adequate for solving the problem.

The next example is from a medical textbook on hepatitis:

(6) Blood probably contains the highest concentration of hepatitis B virus of any tissue except liver.
Semen, vaginal secretions, and menstrual blood contain the agent and are infective.
Saliva has lower concentrations than blood, and even hepatitis B surface antigen may be detectable in no more than half of infected individuals.
Urine contains low concentrations at any given time.

The predicate p is *contain*; the diagram in Figure 5 indicates the corresponding similar arguments and the shared properties (the column headings) by virtue of which they are similar.

Note also that the sentences are in order of decreasing concentrations; it is very frequent for particular genres or "microgenres" to be characterized by further constraints imposed on these universal coherence patterns.

4.5 Elaboration

Elaboration is a limiting case of the Parallel relation, where the entities are not merely similar but identical. This amounts to saying that the same proposition can be inferred from the assertions of each of the segments. At some level, both segments say the same thing. In our notation, this can be captured by the relation gen.

$$(\forall\, e_1, e_2, e) Elaboration(e_1, e_2, e) \supset CoherenceRel(e_1, e_2, e)$$

$$(\forall\, e_1, e_2, e) gen(e_1, e) \wedge gen(e_2, e) \supset Elaboration(e_1, e_2, e)$$

That is, if there is an eventuality e that is "generated" by each of the eventualities e_1 and e_2, then there is an Elaboration coherence relation between e_1 and e_2, and the assertion of the composed segment will be e.

Frequently the second segment adds crucial information and this is why I have called the relation Elaboration, but it is not specified in the definition since it is desirable to include pure repetitions under this heading.

A simple example of the Elaboration relation is the following:

> Go down First Street.
> Just follow First Street three blocks to A Street.

Note that it is important to recognize that this is an Elaboration, rather than two temporally successive instructions.

From the first sentence we can infer

go(`Agent`: you, `Goal`: x, `Path`: First St., `Measure`: y)

for some x and y. From the second we can infer

go(`Agent`: you, `Goal`: A St., `Path`: First St., `Measure`: 3 blks)

If we assume that x is A Street and y is 3 blocks, then the two are identical and serve as the proposition P in the definition.

This is a simple enough example that we can work out the details of the abductive derivation of this relation (ignoring "three blocks"). To interpret the discourse we must prove abductively the expression

$Segment$("Go ... A Street.", e)

To prove the text is a segment, we need to prove each sentence is a segment, by proving it is a sentence. This taps us into an expanded version of the sentence grammar of Section 2.2, which requires us to prove the logical form of the sentences. We also need to prove there is a coherence relation between the two sentences. Thus, we need to prove (simplifying somewhat),

$$(\exists\, g, u, x, y, f, f_1) go'(g, u, x, y) \wedge down(g, FS) \wedge CoherenceRel(g, f, f_1)$$
$$\wedge\, follow'(f, u, FS, AS)$$

That is, there is a going g by u from x to y and the going is down First Street (FS). There is also a following f by u of First Street to A Street (AS). Finally, there is a coherence relation between the going g and the following f, with the composite assertion f_1.

Suppose we have the following axioms in our knowledge base:

$$(\forall\, f) gen(f, f)$$

That is, the *gen* relation is reflexive.

$$(\forall\, g, u, x, y, z) go'(g, u, x, y) \wedge along(g, z) \supset (\exists\, f) follow'(f, u, z, y) \wedge gen(g, f)$$

That is, if g is a going by u from x to y and is along z, then g generates a following f by u of z to y.

$$(\forall\, g, z) down(g, z) \supset along(g, z)$$

That is, a *down* relation is one kind of *along* relation.

If we assume $go'(g, u, x, y)$ and $down(g, FS)$, then the proof of the logical form of the text is straightforward. It is illustrated in Figure 6.

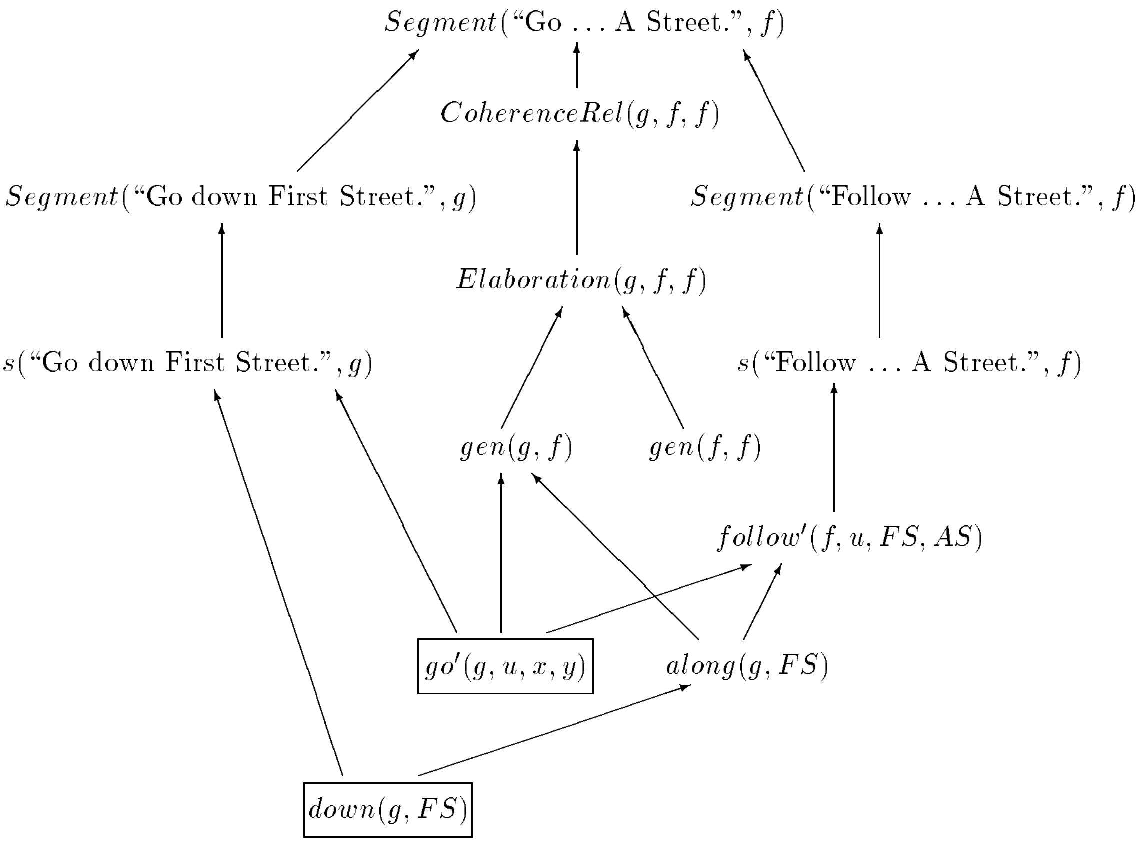

Figure 6: Interpretation of "Go down First Street. Follow First Street to A Street."

4.6 Contrast

There are at least two cases of the Contrast relation. The first holds when we can infer a proposition $p(x)$ from the assertion of one segment S_1 and we can infer $\neg p(y)$ from the assertion of the next segment S_2, where a and b are similar entities.

The following axiom is a first approximation to this characterization:

(7) $(\forall\, p, q, x, y, e_1, e_2, e_3, e_4) p'(e_1, x) \wedge not'(e_3, e_4) \wedge p'(e_4, y)$
$\wedge\, gen(e_3, e_2) \wedge q(x) \wedge q(y) \supset Contrast(e_1, e_2)$

That is, a Contrast relation holds between e_1 and e_2 if there is a p such that e_1 is p's being true of some x, and there is an e_3 that generates e_2 that is the negation of an e_4 which is p's being true of some y, and there is some q true of x and y.

The Contrast relation is a coherence relation:

$$(\forall\, e_1, e_2, e) Contrast(e_1, e_2) \supset CoherenceRel(e_1, e_2, e_2)$$

The relation is generally hypotactic. Usually, the second segment is the dominant one, but this can be overridden.

The second case of the Contrast relation holds when we can infer a proposition $p(x)$ from the assertion of one segment S_1 and we can infer $p(y)$ from the assertion of the next segment S_2, where there is some property q such that $q(x)$ and $\neg q(y)$. The second segment is generally dominant.

The first case is illustrated by an extended example, involving the interpretation of a metaphor.

John is an elephant.

In isolation, this can mean many things, e.g., that John is heavy, or thick-skinned, or has a good memory. In specific contexts it takes on specific meanings. In the following context,

(8) Mary is graceful. John is an elephant.

the most reasonable interpretation is that John is clumsy.[6]

To interpret this text abductively, we first need an axiom relating clumsiness and elephants:

(9) $(\forall\, y) clumsy(y) \wedge etc_2(y) \supset elephant(y)$

This can be read as saying that if something is clumsy and some other unspecified properties hold, then it is an elephant, or in other words, being an elephant is one way of being clumsy.

We now need to introduce a further complication, since we will have to refer explicitly to the properties of clumsiness, elephanthood, and grace. Axiom (9) must be rewritten as follows:

(10) $(\forall\, e_3, y) clumsy'(e_3, y) \wedge etc_2(e_3, y)$
$\supset (\exists\, e_2) elephant'(e_2, y) \wedge gen(e_3, e_2)$

[6]Elephants of course are not clumsy, but according to our conventional stereotype, they are. This property is therefore in our "knowledge" base and hence available in discourse interpretation. Searle (1979) made this point, with respect to gorillas' being "fierce, nasty, prone to violence, and so forth."

That is, if e_3 is the condition of y's being clumsy and some other unspecified things are true of e_3 and y, then there is a condition e_2 of y's being an elephant. Furthermore, there is a very tight relation between e_3 and e_2: y is an elephant *by virtue of* its being clumsy and the other things being true.

Next we need an axiom relating clumsiness and grace.

(11) $(\forall\, e_3, e_4, y) not'(e_3, e_4) \wedge graceful'(e_4, y) \supset clumsy'(e_3, y)$

That is, if e_3 is the condition of e_4 not being true, where e_4 is the condition of y's being graceful, then e_3 is the condition of y's being clumsy.

Suppose we also know that Mary and John are people:

$$person(M),\ person(J)$$

Now we are ready to interpret text (8). Its logical form is

$$(\exists\, e_1, e_2, e) graceful'(e_1, M) \wedge elephant'(e_2, J) \wedge CoherenceRel(e_1, e_2, e)$$

We can backchain on axiom (10) from "elephant" to "clumsy", assume $etc_2(e_3, J)$, backchain on axiom (11) from "clumsy" to "not graceful", and assume $not'(e_3, e_4)$ and $graceful'(e_4, J)$. We also assume $graceful'(e_1, M)$. Then we have a proof of $Contrast(e_1, e_2)$, using axiom (7), with p instantiated to $graceful$ and q instantiated to $person$. This establishes $CoherenceRel(e_1, e_2, e_2)$

As a by-product of recognizing the coherence of the text, we have interpreted the metaphor. Figure 7 illustrates the interpretation of "elephant", although it was the requirement to explain the "but" relation that drove the interpretation.

As always, other coherence relations are theoretically possible in this example. The next sentence might be "Mary can dance on his back," in which case the second sentence would not be in contrast with the first but background for the third, and John would be a real rather than a metaphorical elephant.

5 A Method for Analyzing Discourse

This account of the structure of discourse suggests a method for analyzing discourse. The method consists of four steps, each an order of magnitude more difficult than the one before it. I will illustrate it on the following text:

(12a) I would like now to consider the so-called "innateness hypothesis,"
(12b) to identify some elements in it that are or should be controversial, and
(12c) to sketch some of the problems that arise as we try to resolve the controversy.
(13) Then, we may try to see what can be said about the nature and exercise of the linguistic competence that has been acquired, along with some related matters.

Chomsky, *Reflections on Language,* p. 13.

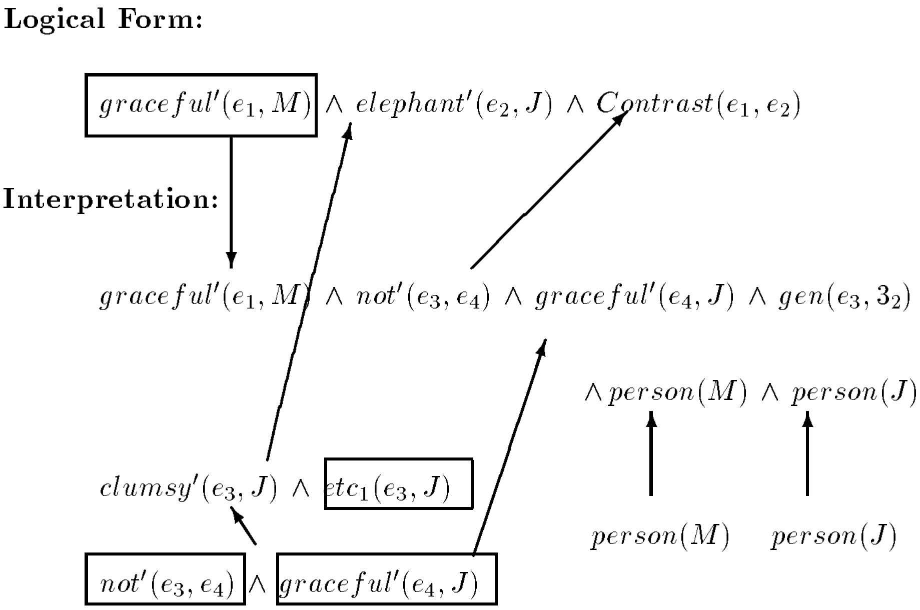

Figure 7: Interpretation of Contrast Example.

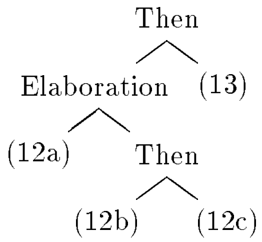

Figure 8: Structure of sentences (12)–(13).

The method is as follows:

1. One identifies the one or two major breaks in the text and cuts it there. That is, one chooses the most natural way to divide the text into two or three segments. This can be done on a strictly intuitive basis by anyone who has understood the text, and among those who have understood it in the same way, there will be a large measure of agreement. This process is then repeated for each of the segments, dividing *them* in the most natural places. The process is continued until reaching the level of single clauses. This yields a tree structure for the text as a whole.

In text (12), for example, the major break comes between sentences (12) and (13). Within sentence (12) there is a break between the first clause and the last two, and of course a final break betwen the second and third clauses of the first sentence. This yields the tree of Figure 8.

2. One labels the nonterminal nodes of the tree with coherence relations. Proceeding from the bottom up, one devises rough accounts of what is asserted by each composed

segment. Thus, in the Chomsky example, we label the node linking (12b) and (12c) with the "then" (or *occasion*) relation. We label the node linking the resulting segment and (12a) with the *elaboration* relation. Finally, we label the node linking (12) and (13) with the "then" relation. This process yields the labels on the nodes of the tree in Figure 8.

In this step the method becomes theory-specific, as one must know what the relations are and have at least rough characterizations of them. One aid in this step is to determine what conjunctions or sentential adverbs it would be appropriate to insert. If we can insert "then" between S_0 and S_1, and the sense would be changed if we reversed the segments, then the *occasion* relation is an excellent candidate. If we can insert "because", the *explanation* relation becomes a strong possibility. "That is" or "i.e." suggests *elaboration*, "similarly" suggests *parallel*, "for example" suggests *exemplification*, and "but" suggests *contrast* or *violated expectation*. It should be emphasized, however, that these tests are informal. They do not define the relations. Conjunctions and sentential adverbials impose constraints on the propositional content of the clauses they link or modify, and in many cases these constraints are almost the same as those imposed by some coherence relation. In the best of cases there is sufficient overlap for the conjunction to tell us what the coherence relation is.

3. One makes (more or less) precise the knowledge or beliefs that support this assignment of coherence relations to the nodes. Each of the coherence relations has been defined in terms of the inferences that must be drawn from the listener's knowledge base in order to recognize the relation. When we say, for example, that a "then" relation occurs between (12b) and (12c), we have to specify the change asserted in (12b) (namely, a change in mutual knowledge about where the controversy lies, from the word "identify") that is presupposed in the event described in (12c), (the effort to resolve the controversy). Thus, we need knowledge about what change is effected by the action of identifying, and we need to know the meanings of "controversy" and "resolution" that allow us to talk about controversies being resolved.

The precision with which we specify the knowledge really can be "more or less", depending on our aims. We might be satisfied with a careful statement in English, or we might demand formulation in terms of some logical language, embedded within a larger formal theory of the commonsense world.

4. One validates the hypotheses made in step 3 about what knowledge underlies the discourse. Mike Agar and I (Agar and Hobbs, 1982) have discussed at length how this should proceed. Briefly, one looks at the larger corpus to which the text belongs, a corpus by the same speaker or from the same culture that assumes the same audience. One attempts to construct a knowledge base or system of mutual beliefs that would support the analyses of all of the texts in the corpus. If step 1 is a matter of minutes for a text of paragraph length, step 2 a matter of an hour or two, and step 3 a matter of days, then step 4 is a matter of months or years.

In each of these steps difficulties may arise, but these difficulties in analysis will usually reveal problematic aspects of the text. In step 1, we might find it difficult to segment the text in certain places, but this probably reflects a genuine area of incoherence in the text itself. We might find it easy to segment the text because the segments are about clearly different topics, but be unable to think of a coherence relation that links the segments.

When this happens, it may be that we have found two consecutive texts rather than a single text. At times the knowledge that underlies a composed segment is not obvious, but this often leads us to very interesting nonstandard assumptions about the belief systems of the participants. Finally, we often cannot be sure the knowledge we have assumed to be operative really is operative; looking at further data forces revisions in our assumptions.

There is a structural feature of some discourses, especially spoken conversation, that must be considered in segmentation—discourse pivots. A discourse pivot is a segment S_2 such that in $S_1 S_2 S_3$, S_1 and S_2 are related to each other by virtue of one part of the content of S_2, and S_2 and S_3 are related to each other by virtue of another part of the content of S_2. The effect on the overall structure of the text is that one tree subsumes $S_1 S_2$, and another subsumes $S_2 S_3$. Segment (14h) below is an example of a discourse pivot.

6 Analysis of a Fragment of Conversation

Texts of all genres can be illuminated by a coherence analysis. For the extended example for this paper, I have chosen a particularly challenging variety—a fragment of a three-person conversation in a decision-making meeting (although there are only two speakers in this fragment). The participants are collaborating to make up a schedule for the visit of a contracting officer, making sure several constraints are satisfied as best as possible. The fragment is as follows:

(14a)	A:	So, um, if I went first, let's say with for um, see I,
(14b)		as I said, I need about an hour and fifteen minutes
(14c)		I could do the, my reporting on the ongoing project, ah, for that first hour.
(14d)		See if we total up all the time we need,
(14e)		let's see an hour for Brian,
(14f)	A:	an hour and fifteen minute for me,
	B:	So it's
(14g)	A:	thirty five minutes,
	B:	almost exactly...
(14h)	A:	it's almost exactly correct. Three hours.
(14i)	B:	But we've got to take into account that they're typically late on these things.
(14j)	B:	All right, so we're gonna get squeezed someplace.
	A:	Okay, right, right, okay. Um,
(14k)	A:	I think what I'd be willing to do is if we get squeezed on the, uh if I go first and if we get squeezed I'll I'll eat the ah the time that we lose.

Overlapping utterances are aligned under the same number.

The coherence structure is as shown in Figure 9.

This fragment breaks into three pieces. The first includes segments (14a) – (14c), in which A says how much time he will need and when it should occur. Then in (14d) – (14h) he backs away to describe the overall schedule of which his presentation would be a part; one part of this, segment (14f), repeats the content of segment (14c). Segment

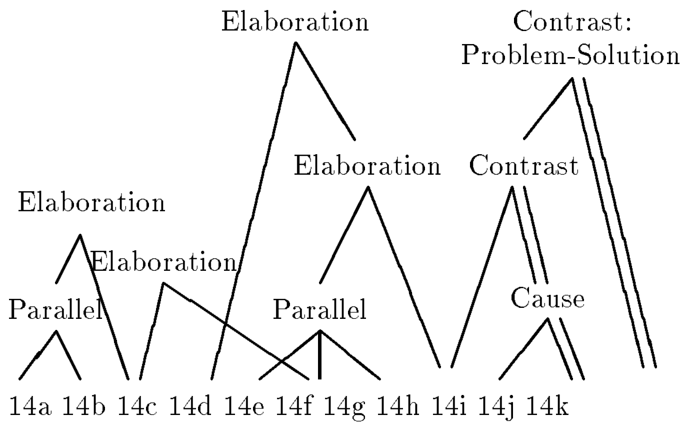

Figure 9: Coherence Structure of Fragment of Conversation

(14h) constitutes a discourse pivot, and segment (14h) – (14k) discusses a problem and its solution. We will examine the fine structure of each of these three segments in turn.

To construct a schedule, the participants must allocate a certain amount of time to each activity and specify the order in which they will occur. In (14a) and (14b) A attends to these two parallel aspects of scheduling for his own presentation. In (14a) he says that he will go first. In (14b) he says that he will take one hour and fifteen minutes. In (14c) he elaborates on this by describing the activity "for that first hour".

At this point he pops up to a more global concern—the total time required by all three participants, in which his hour and fifteen minutes will be a part. He begins by giving an abstract of what he is going to do—total up the time they all need. This is followed by a breakdown into the parallel requirements for each participant, one hour for B in (14e), an hour and fifteen minutes for A in (14f), and thirty-five minutes for C in (14g). This is then summarized or, more precisely, added up in (14h). The segment (14e) – (14h) then constitutes the promised elaboration on (14d) by providing the total time needed.

B has been an active participant in this segment, but only to reinforce or otherwise aid A's development.

In (14i) B keys off of (14h) in a different manner. A uttered it as a summary of the time needed. But in (14i) B questions the correctness of their having a full three hours. Sometimes visitors are late. The lateness causes the available time to become smaller, as B says in (14j). Their having less time is in contrast to the statement in (14h) that they have exactly three hours.

Segment (14h) is thus a discourse pivot; it functions with the segment before it as an elaboration and with the segment after it as the suborinate element in a contrast.

A problem is a situation which is not in accord with one's goals. A solution to the problem is a modification of that situation, generally through one's own actions, that *is* in accord with the goals. Thus, the statement of a problem and its solution is a variety of the contrast pattern. Here segment (14h) – (14j) describes the problem, with the pricipal statement of the problem being in (14j). Segment (14k) describes the solution. If any time is lost, A will subtract that time from his own presentation.

7 Conclusion

The theory of local coherence in discourse that I have sketched in this paper is part of a larger theory that seeks to make explicit the connection between the interpretation of a text and the knowledge or belief system that underlies the text. The coherence relations that give structure to a text are part of what an interpretation is; they are recognized by drawing inferences, and thus specify one connection that must exist between interpretations and knowledge. The method outlined in this paper can be used to exploit that connection in several ways.

Where, as in ethnography, our interest is in the belief systems, or the culture, shared by the participants, the method acts as a "forcing function". It does not tell us what the underlying beliefs are, but it forces us to hypothesize beliefs we might otherwise overlook, and it places tight constraints on what the beliefs can be.

Where our interest is primarily in the interpretation of the text, as in literary criticism, the method gives us a technique for finding the structure of the text, an important aspect of the interpretation. In placing constraints on the ideal structure of a text, it can point us toward problematic areas of the text where the ideal of coherence proposed here does not seem to be satisfied. We might ultimately decide in such cases that the ideal is in fact not satisfied, but many times we will find that the attempt to satisfy the ideal leads us to interesting reinterpretions of the whole text.

References

[1] Agar, Michael, and Jerry R. Hobbs, 1982. "Interpreting Discourse: Coherence and the Analysis of Ethnographic Interviews", *Discourse Processes*, Vol. 5, No. 1, pp. 1-32.

[2] Davidson, Donald, 1967. "The Logical Form of Action Sentences", in N. Rescher, ed., *The Logic of Decision and Action*, pp. 81-95, University of Pittsburgh Press, Pittsburgh, Pennsylvania.

[3] Downing, Pamela, 1977. "On the Creation and Use of English Compound Nouns", *Language*, vol. 53, no. 4, pp. 810-842.

[4] Hobbs, Jerry R. 1978. "Why is Discourse Coherent?" SRI Technical Note 176, SRI International, Menlo Park, California. November 1978. Also in Fritz Neubauer, Ed., *Coherence in Natural-Language Texts*, pp. 29-69. Hamburg: Helmut Buske Verlag. 1983.

[5] Hobbs, Jerry R. 1985. "Ontological Promiscuity." *Proceedings, 23rd Annual Meeting of the Association for Computational Linguistics*, pp. 61-69. Chicago, Illinois, July 1985.

[6] Hobbs, Jerry R., 1985. "On the Coherence and Structure of Discourse", Report No. CSLI-85-37, Center for the Study of Language and Information, Stanford University.

[7] Hobbs, Jerry R., 1995a. "The Roles of Intention and Information in the Interpretation of Utterances", to be published in D. Scott and E. Hovy, eds., *Proceedings*, NATO Workshop on Burning Issues in Discourse, Maratea, Italy, April 1993.

[8] Hobbs, Jerry R., 1995b. "The Syntax of English in an Abductive Framework", manuscript.

[9] Hobbs, Jerry R., and Megumi Kameyama, 1990. "Translation by Abduction", in H. Karlgren, ed., *Proceedings*, Thirteenth International Conference on Computational Linguistics, Helsinki, Finland, Vol. 3, pp. 155-161, August, 1990.

[10] Hobbs, Jerry R., Mark Stickel, Douglas Appelt, and Paul Martin, 1993. "Interpretation as Abduction", SRI Technical Note 499, SRI International, Menlo Park, California. December 1990.

[11] Hobbs, Jerry R., Mark Stickel, Douglas Appelt, and Paul Martin, 1993. "Interpretation as Abduction", *Artificial Intelligence*, Vol. 63, Nos. 1-2, pp. 69-142.

[12] Kowalski, Robert, 1980. *Logic for Problem Solving*, North Holland, New York.

[13] Levi, Judith, 1978. *The Syntax and Semantics of Complex Nominals*, Academic Press, New York.

[14] McCarthy, John, 1987. "Circumscription: A Form of Nonmonotonic Reasoning", in M. Ginsberg, ed., *Readings in Nonmonotonic Reasoning*, pp. 145-152, Morgan Kaufmann Publishers, Inc., Los Altos, California.

[15] Pereira, Fernando C. N., and David Warren, 1983. "Parsing as Deduction", *Proceedings of the 21st Annual Meeting, Association for Computational Linguistics*, pp. 137-144. Cambridge, Massachusetts, June 1983.

[16] Polanyi, Livia, 1988. "A Formal Model of Discourse Structure", *Journal of Pragmatics*, Vol. 12, pp. 601-638.

[17] Pollard, Carl, and Ivan A. Sag, 1994. *Head-Driven Phrase Structure Grammar*, University of Chicago Press and CSLI Publications.

[18] Searle, John, 1979. "Metaphor", In A. Ortony (Ed.), *Metaphor and Thought*, pp. 92-123. Cambridge, England: Cambridge University Press.

[19] Winograd, Terry, 1972. *Understanding Natural Language*, Academic Press, New York.