Intention, Information, and Structure in Discourse: A First Draft

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

1 Introduction

In the paper "Interpretation as Abduction" (hereafter IA) Hobbs et al. (1992) present and elaborate the view that to interpret an utterance is to find the best explanation of why it would be true. We may call this the "Informational Perspective" on discourse interpretation. The only thing to be explained is the information explicitly conveyed by the utterance, and the explanation does not involve any knowledge of the specific goals of the speaker.

Norvig and Wilensky (1990) raise the objection to this approach that what really needs to be explained is what the speaker was trying to accomplish with the utterance. Under this view, to interpret an utterance is to find the best explanation of why it was said. We may call this the "Intentional Perspective" on discourse interpretation.

The Intentional Perspective has been the canonical view in natural language processing since the middle 1970s. It originated with Power (1974), Bruce (1975), and Schmidt et al. (1978), and is the view adopted in Cohen and Perrault (1979), Grosz (1979), Allen and Perrault (1980), Perrault and Allen (1980), Hobbs and Evans (1980), Grosz and Sidner (1986) and many others since that time. The view taken in all of this work is that the speaker is executing a plan, the utterance is an action in that plan, and the job of the hearer is to discover the plan and the role that the utterance plays in the plan. This is an especially useful, indeed essential, perspective when the discourse is a dialogue in which most turns are a sentence or less in length and the participants' plans are being modified continuously by the interaction.

It is clear why the Intentional Perspective is the correct one when we look at things from the broadest possible point of view. An intelligent agent is embedded in the world and must, at each instant, understand the current situation. The agent does so by finding an explanation for what is perceived. Put differently, the agent must explain why the complete set of observables encountered constitutes a coherent situation. Other agents in the environment are viewed as intentional, that is, as planning mechanisms, and this means that the best explanation of their observable actions is most likely to be that they are steps in a coherent plan. Thus, making sense of an environment that includes other agents entails making sense of the other agents' actions in terms of what they are intended to achieve. When those actions are utterances, the utterances must be understood as actions in a plan the agents are trying to effect. That is, the speaker's plan must be recognized—the Intentional Perspective.

But there are several serious problems with the Intentional Perspective. First, the speaker's plan can play at best an indirect role in the interpretation process. The hearer has no direct access to it. It plays a causal role in some observable actions, in particular the utterance, which the hearer can then use, along with background knowledge, to form a belief about exactly what the plan is. Only this belief can play a direct role in interpretation. How is the hearer to arrive at this belief? How can the hearer go from utterance to intention, in those cases where there is no prior knowledge of the intention?

There is a further problem, that occurs especially in extended, one-speaker discourse, such as written text. There is a level of detail that is eventually reached at which the Intentional Perspective tells us little. It tells us that the proper interpretation of a compound nominal like "coin copier" means what the speaker intends it to mean, but it offers us virtually no assistance in determining what it really *does* mean. Frequently what the speaker intends an utterance to mean is just what it would mean if spoken by almost anyone else in almost any other circumstance. We need a notion of interpretation that is independent of and goes beyond speaker's intention. It must, for example, give us access to plausible relations between coins and copiers.

A third problem with the Intentional Perspective is that there are many situations in which the speaker's plan is of little interest to the hearer. Someone in a group conversation may use a speaker's utterance solely as an excuse for a joke, or as a means of introducing a topic *he* or *she* wants to talk about. Very often two speakers in a discussion will try to understand each other's utterances in terms of their own frameworks, rather than attempt to acquire each other's framework. A medical patient, for example, may describe symptoms according to some narrative scheme, while the doctor tries to map the details into a diagnostic framework. A spy learning a crucial technical detail from the offhand remark of a low-level technician doesn't care about the speaker's intention in making the utterance, but only about how the information fits into his own prior global picture. A historian examining a document often adopts a similar stance. In all these cases, the hearer has his or her own set of interests, unrelated to the speaker's plan, and interpretation involves primarily relating the utterance to those interests.

In brief, the role of the speaker's intention is indirect, it is often uninformative, and it is frequently not very important. It cannot be the whole story. We need to have an intention-independent notion of interpretation.

Our first guess might be that we simply need the literal meaning of the utterance. But an utterance does not wear its meaning on its sleeve. Anaphora and ambiguities must be resolved. Metonymies and ellipsis must be expanded. Vague predications, including those conveyed by the mere adjacency of words or larger portions of text, must be made specific. In short, the utterance must be interpreted. The notion of literal meaning gets us nowhere.

The canonical use of language is to present the facts about a situation. To understand a situation that we perceive we have to find an explanation for the observable facts in that situation. Similarly, to understand a situation that is described to us we must find an explanation for the facts we are told. But this is exactly the account of what an interpretation of an utterance is under the Informational Perspective. The "informational interpretation" gives us an analogue of literal meaning that is adequate to the task. As shown in IA, interpreting an utterance by finding the best explanation for the information it conveys solves as a by-product the problems listed above—resolving anaphora and ambiguities, expanding metonymies and ellipsis, and determining specific meanings for vague predicates.

The informational interpretation is, to be sure, relative to an assumed background knowledge. Conversation is possible only between people who share some background knowledge, and interpretation is always with respect to some background knowledge that the hearer presumes to be shared. The explanation that constitutes the interpretaton has to come from somewhere. But conversation, and hence interpretation, *is* possible in the absence of information about the other's specific goals. We have conversations with strangers all the time.

The picture that emerges is this. Humans have constructed, in language, a tool that is primarily for conveying information about situations, relying on shared background knowledge. Like all tools, however, it can be put to uses other than its primary one. We can describe situations for purposes other than having the hearer know about them. The Informational Perspective on discourse interpretation tells us how to understand the situations described in a discourse. The Intentional Perspective tells us how to discover the uses to which this information is being put.

The Intentional Perspective on interpretation is certainly correct. To understand what's going on in a given communicative situation, we need to figure out why the speaker is making this particular utterance. But the Informational Perspective is a necessary component of this. We often need to understand what information the utterance would convey independent of the speaker's intentions. Another way to put it is this. We need to figure out why the speaker uttered a sequence of words conveying a particular content. This involves two parts, the informational aspect of figuring out what the particular content is, and the intentional aspect of figuring out why the speaker wished to convey it.

It should not be concluded from all of this that we first compute an informational interpretation and then as a subsequent process compute the speaker's intention. The two intimately influence each other. Sometimes, especially in the case of long written texts and monologues, the informational aspect completely overshadows considerations of intention. Other times, our knowledge of the speaker's intention completely masks out more conventional readings of an utterance. We consequently need a framework that will give us the conventional meaning, relative to a shared knowledge base, but will also allow us to override or to completely ignore this meaning when more is known about the speaker's aims. This paper is a preliminary effort to provide such a framework.

In Sections 2, 6, and 7 of this paper and more completely in IA, a framework is presented in which a number of discourse phenomena can be handled in a unified framework using abductive inference to construct the best explanation for the information conveyed explicitly in a text. The logical forms of the sentences in the text are proven abductively, and the solution to the discourse problems simply fall out. These phenomena are all basically informational in character. There is no essential appeal to speaker's intention.

The phenomena are

- Local pragmatics, that is, those pragmatics problems that arise within the scope of single sentences, such as resolving anaphora and ambiguities, expanding metonymies and ellipsis, and determining specific meanings for vague predicates.
- Syntactic structure and compositional semantics, in particular, recognizing the predicate-argument relations encoded in the text.
- Local coherence (a term introduced by Agar and Hobbs, 1982), or the recognition of the coherence relations, that is, the relations conveyed by the mere adjacency of segments of text, which give structure to a discourse.

What is left out of that integrated framework was what Agar and Hobbs called "global coherence", namely, the recognition of the relation between parts of the discourse and the speaker's plan—the Intentional Perspective.

Recognizing the speaker's plan is also a problem of abduction. If we encode as axioms beliefs about what kinds of actions cause and enable what kinds of events and conditions, then in the presence of complete knowledge of the speaker's goals and beliefs, it is a matter of deduction to prove that the speaker believes a sequence or more complex arrangement of actions will achieve the goals. Unfortunately, we rarely have complete knowledge. We will almost always have to make assumptions. That is, abduction will be called for. We must prove abductively that the utterance contributes to the achievement of a goal of the speaker, within the context of a coherent plan. In the process we ought to find ourselves making many of the assumptions that hearers make when they are trying to "psych out" what the speaker is doing by means of his or her utterance. (Appelt and Pollack (1990) have also examined how weighted abduction of the sort presented in IA can be used for the plan ascription problem.) One might think that this requirement from the Intentional Perspective is an addition to the informational requirement of proving the logical form. But in this paper it is shown that the former subsumes the latter.

2 Interpretation as Abduction

Abductive inference 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 why the sentences would be true. Cashing in this idea procedurally, we can give the following characterization of the interpretation of a sentence.

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 local pragmatics problems, 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[1] 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.

$$person(J_1)$$

[1]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

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.

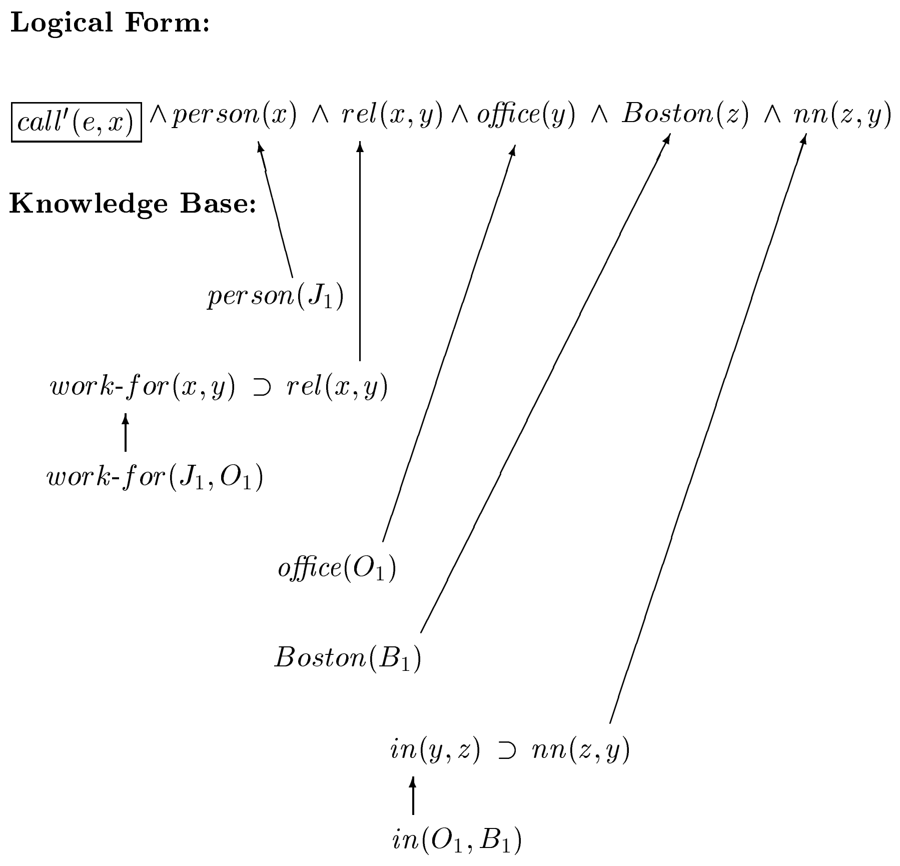

Figure 1: Interpretation of "The Boston office called."

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."

In IA, we show how a number of discourse interpretation problems can be solved in a natural way by this method. However, this was entirely from an Informational Perspective. We did not attempt to recognize the speaker's plan.

3 An Example of Plan Recognition

Let us analyze an example from a set of dialogues collected by Barbara Grosz (1977) between an expert and an apprentice engaged in fixing an air compressor. They are in different rooms, communicating by terminals. The apprentice A is doing the actual repairs, after receiving instructions from the expert B. At one point, the following exchange takes place:

B: Tighten the bolt with a ratchet wrench.
A: What's a ratchet wrench?
B: It's between the wheel puller and the box wrenches.

A seems to be asking for a definition of a ratchet wrench. But that is not what B gives her. He does not say

> A ratchet wrench is a wrench with a pawl, or hinged catch, that engages the sloping teeth of a gear, permitting motion in one direction only.

Instead he tells her where it is.

According to a plausible analysis, B has interpreted A's utterance by relating it to A's overall plan. B knows that A wants to use the ratchet wrench. To use a ratchet wrench, you have to know where it is. To know where it is, you have to know what it is. B responds to A's question, not by answering it directly, but by answering to a higher goal in A's presumed overall plan, by telling A where it is.

B has therefore recognized the relationship between A's utterance and her overall plan. I will give two accounts of how this recognition could have taken place. The first account is informational. It is derived in the process of proving the logical form. The second account is intentional and subsumes the first. It is derived in the process of explaining, or proving abductively, the fact that A's utterance occurred.

4 The Informational Solution

For this solution we will need two axioms encoding the planning process:

(4) $(\forall a, e_0, e_1) goal(a, e_1) \wedge enable(e_0, e_1) \supset goal(a, e_0)$

or if an agent a has e_1 as a goal and e_0 enables, or is a prerequisite for, e_1, then a has e_0 as a goal as well.

(5) $(\forall\, a, e_0, e_1)goal(a, e_1) \wedge cause(e_0, e_1) \wedge etc_1(a, e_0, e_1) \supset goal(a, e_0)$

or if an agent a has e_1 as a goal and e_0 causes, or is one way to accomplish, e_1, then a may have e_0 as a goal as well. The etc_1 literal encodes the uncertainty as to whether e_0 will be chosen as the way to bring about e_1 rather than some other action that causes e_1.

In terms of STRIPS operatiors (Fikes and Nilsson, 1971), the first axiom says that prerequisites for an action must be satisfied, while the second axiom says essentially that to achieve a goal, an operator needs to be chosen and its body (e_0) needs to be executed.

Next we need two domain axioms of a rather general character.

(6) $(\forall\, e_2, a, x)use'(e_2, a, x) \supset (\exists\, e_3, e_4, y)enable(e_3, e_2) \wedge know'(e_3, a, e_4)$
$\wedge\, at'(e_4, x, y)$

or an agent a's use e_2 of a thing x has as a prerequisite a's knowing e_3 the fact e_4 that x is at someplace y. To use something, you have to know where it is.

(7) $(\forall\, e_3, a, e_4, x, y)know'(e_3, a, e_4) \wedge at'(e_4, x, y) \supset (\exists\, e_5, e_6)enable(e_5, e_3)$
$\wedge\, know'(e_5, a, e_6) \wedge wh'(e_6, x)$

or an agent a's knowing e_3 the fact e_4 that a thing x is at someplace y has as a prerequisite a's knowing e_5 what x is (e_6). To know where something is, you have to know what it is. We dodge the complex problem of specifying what constitutes knowing what something is by encoding it in the predicate wh, which represents the relevant context-dependent essential property.

Let us suppose that the logical form of

What's a ratchet wrench?

is

(8) $(\exists\, a, e_5, e_6)goal(a, e_5) \wedge know'(e_5, a, e_6) \wedge wh'(e_6, RW)$

That is, the speaker a has the goal e_5 of knowing the essential property e_6 of the ratchet wrench RW.

Suppose also that in B's knowledge of the context is the following fact:

(9) $goal(A, E_2) \wedge use'(E_2, A, RW)$

That is, the apprentice A has the goal E_2 of using the ratchet wrench RW.

The proof of the logical form (8) follows from axioms (4) through (7) together with fact (6), as indicated in Figure 2. Axiom (4) is used twice, first in conjunction with axiom

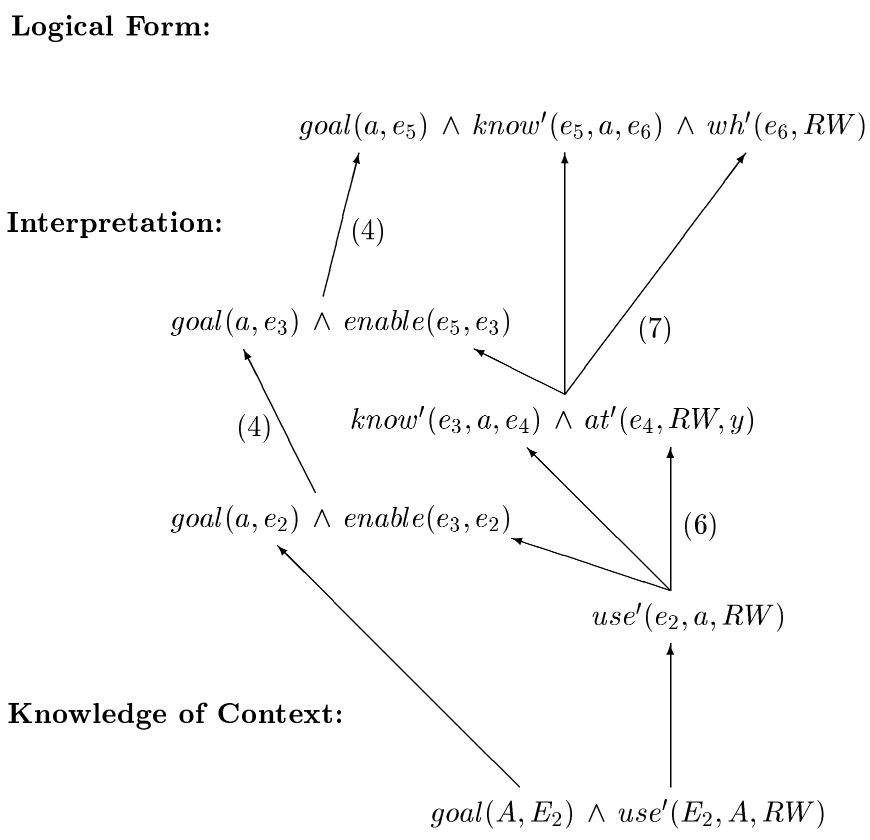

Figure 2: Informational Interpretation of "What's a ratchet wrench?"

(7) and then with axiom (6), to move up the planning tree. The apprentice wants to know what a ratchet wrench is because she wants to know where it is, and she wants to know where it is because she wants to use it. The proof then bottoms out in fact (9).

To summarize, if we take the logical form of a question to be the expression of a desire to know something, then the proof of that logical form very often involves the recognition of the ultimate aims of the speaker in asking it.

5 The Intentional Solution

According to the Informational Perspective, it is the logical form of the utterance that needs to be explained, or proven abductively. We will now take a broader view in which it is the occurrence of an event in the world that has to be explained. It is not the content of the utterance that we have to explain, but rather the very fact that the utterance occurred. Frequently, the best explanation of an event is that it is an intentional action on the part of some agent, that is, it is an action in the service of some goal. This is especially true of utterances—they are generally intentional acts. Thus, we will be interpreting the utterance from an Intentional Perspective. We will ask why the speaker said what she did. We will see how this in turn encompasses the Informational Perspective.

We need several more axioms. First we need some axioms about speaking.

(10) $(\forall\, e_7, a, b, e_8) say'(e_7, a, b, e_8) \supset (\exists\, e_9) cause(e_7, e_9) \wedge know'(e_9, b, e_8)$

That is, if e_7 is a's saying e_8 to b, then that will cause the condition e_9 of b's knowing e_8. Saying causes knowing. The next axiom is the converse of this.

(11) $(\forall\, e_k, y, e) know'(e_k, y, e) \wedge etc_2(e_k, y, e)$
$\supset (\exists\, e_s, x) cause(e_s, e_k) \wedge say'(e_s, x, y, e)$

That is, if e_k is y's knowing the fact e, then it may be (etc_2) that this knowing was caused by the event e_s of x's saying e to y. Knowing is sometimes caused by saying. In the interpretation of the utterance we need only the second of these axioms.

Next we need some axioms (or axiom schemas) of cooperation.

(12) $(\forall\, e_5, e_8, e_9, e_{10}, a, b) know'(e_9, b, e_8) \wedge goal'(e_8, a, e_5) \wedge cause(e_{10}, e_5)$
$\wedge\, p'(e_{10}, b) \wedge etc_3(e_5, e_8, e_9, e_{10}, a, b) \supset cause(e_9, e_{10})$

That is, if e_9 is b's knowing the fact e_8 that a has goal e_5 and there is some action e_{10} by b doing p that causes e_5, then it may be (etc_3) that that knowing will cause e_{10} to actually occur. If I know your goals, maybe I'll help you achieve them. The next axiom schema is the converse of this. It is a kind of attribution of cooperation.

(13) $(\forall\, e_5, e_{10}, b) p'(e_{10}, b) \wedge cause(e_{10}, e_5) \wedge etc_4(e_5, e_{10}, b)$
$\supset (\exists\, e_8, e_9, a) cause(e_9, e_{10}) \wedge know'(e_9, b, e_8) \wedge goal'(e_8, a, e_5)$

That is, if an action e_{10} by b occurs, where e_{10} can cause e_5, then it may be (etc_4) that it was caused by the condition e_9 of b's knowing the fact e_8 that a has the goal e_5. Sometimes I do things because I know it will help you. In the example we will only need the axiom in this direction.

Finally, we need an axiom schema that says that people do what they want to do.

(14) $(\forall\, a, e_7) goal(a, e_7) \wedge p'(e_7, a) \wedge etc_5(a, e_7) \supset Rexists(e_7)$

That is, if a has as a goal some action e_7 that a can perform, then it could be (etc_5) that e_7 will actually occur. This axiom, used in backward chaining, allows us to attribute intention to events.

Now the problem we set for ourselves is not to prove the logical form of the utterance, but rather to explain, or prove abductively, the occurrence of an utterance with that particular content. We need to prove

(15) $(\exists\, e_7, a, b, e_8, e_5, e_6) Rexists(e_7) \wedge say'(e_7, a, b, e_8) \wedge goal'(e_8, a, e_5)$
$\wedge\, know'(e_5, a, e_6) \wedge wh'(e_6, RW)$

That is, we need to explain the existence in the real world of the event e_7 of someone a saying to someone b the proposition e_8 that a has the goal e_5 of knowing the essential property e_6 of a ratchet wrench.

The proof of this is illustrated in Figure 3. The boxes around the "et cetera" literals indicate that they have to be assumed. By axiom (14) we attribute intention to explain the occurrence of the utterance act e_7; it's not like a sneeze. Using axiom (5), we hypothesize that this intention or goal is a subgoal of some other goal e_9. Using axiom (11), we hypothesize that this other goal is b's knowing the content e_8 of the utterance. A uttered the sentence so that B would know its content. Using axiom (5) again, we hypothesize that e_9 is a subgoal of some other goal e_{10}, and using axiom (13) we hypothesize that e_{10} is b's saying e_6 to a. A told B A's goal so that B would satisfy it. Using axiom (5) and (11) again, we hypothesize that e_{10} is a subgoal of e_5, which is a's knowing e_6, the essential property of a ratchet wrench. A wants B to tell her what a ratchet wrench is so she will know it.

The desired causal chain is this: A tells B she wants to know what a ratchet wrench is, so B will know that she wants to know what a ratchet wrench is, so B will tell her what a ratchet wrench is, so she will know what a ratchet wrench is. Causal chains are reversed in planning; if X causes Y, then our wanting Y causes us to want X. Hence, the causal chain is found by following the arrows in the diagram in the reverse direction.

At this point all that remains to prove is

$$(\exists\, a, e_5, e_6) goal(a, e_5) \wedge know'(e_5, a, e_6) \wedge wh'(e_6, RW)$$

But this is exactly the logical form whose proof is illustrated in Figure 2. We have reduced the problem of explaining the occurrence of an utterance to the problem of discovering its intention, and then reduced that to the problem of explaining the content of the utterance. Interpetation from the Intentional Perspective includes as a subpart the interpretation of the utterance from the Informational Perspective.

Observable to be Explained:

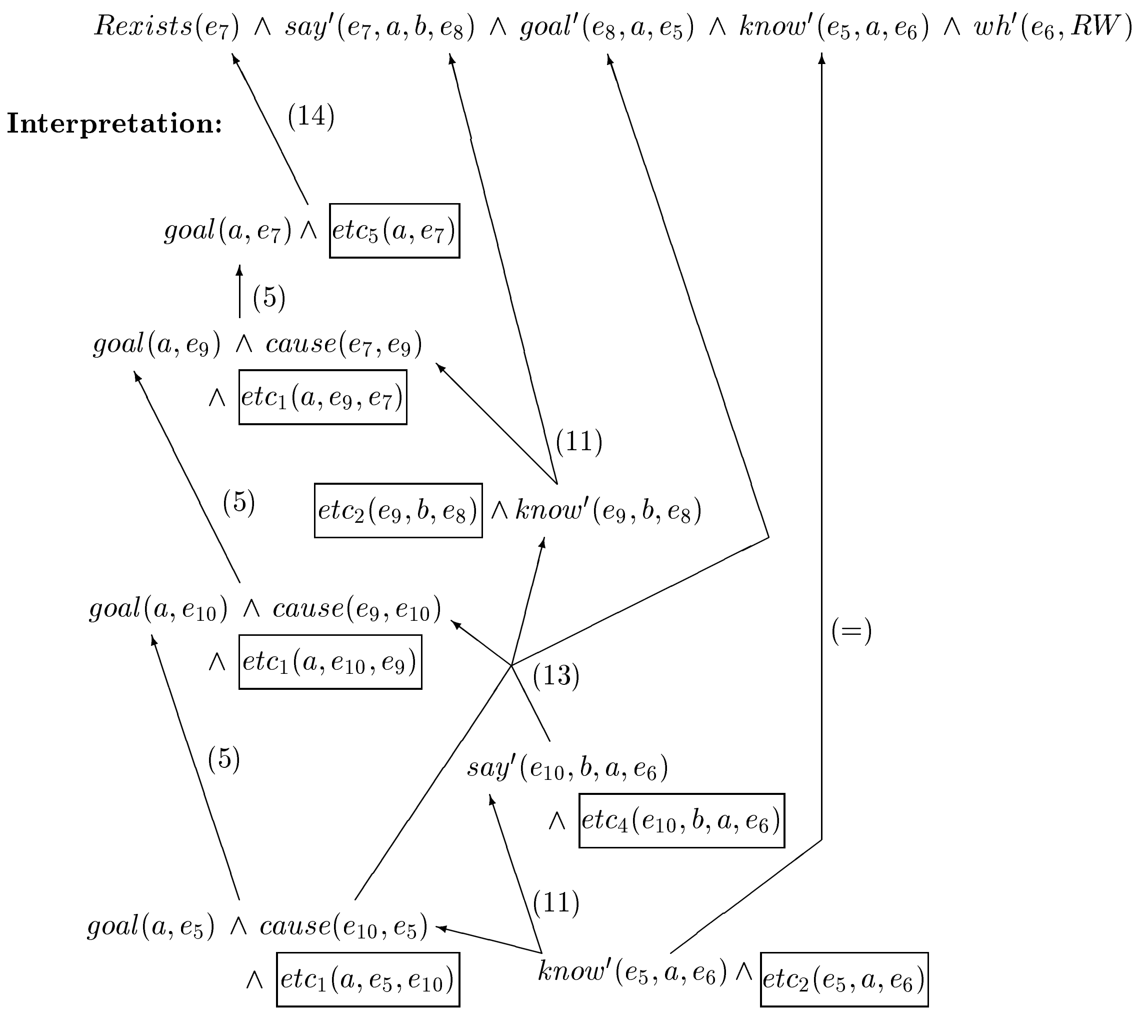

Figure 3: Intentional Interpretation of "What's a ratchet wrench?"

6 Integrating Syntax, Compositional Semantics, and Local Pragmatics

At this point, it is convenient to consider integrating the approach we have developed to local pragmatics with syntactic and compositional semantic processing. We will then extend this to the recognition of discourse structure, and finally return to the problem of integrating the Informational and Intentional Perspectives.

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.[2]

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

(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) $\quad \wedge\, office(y) \wedge Boston(z) \wedge nn(z, y)$

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

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

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

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

[2]This idea is due to Stuart Shieber.

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

that says that one way for the requirements to be satisfied is for x to be a person. Axiom (16) 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 (17) 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 (16) and the conjuncts from line (3b) into axiom (17).

The parse and interpretation of sentence (2) is illustrated in Figure 4.

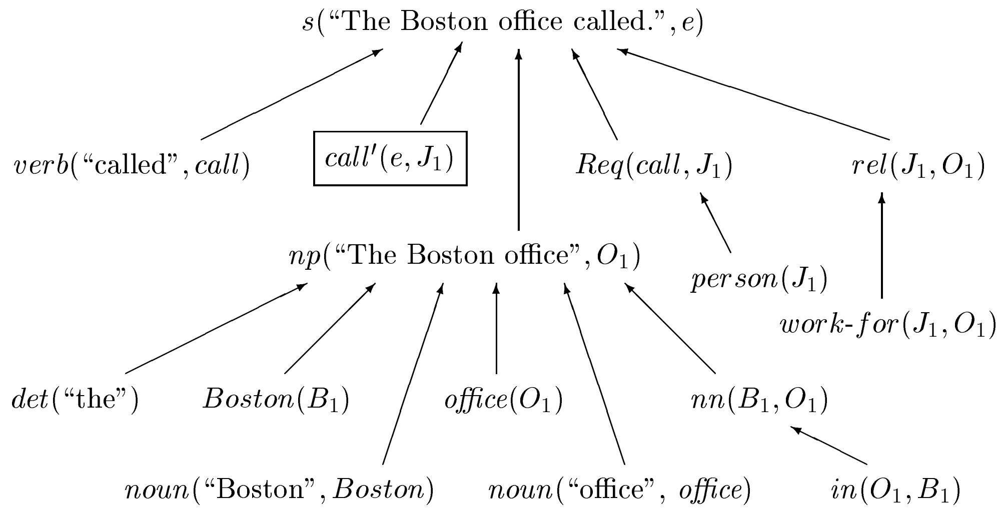

Figure 4: Parse and interpretation of "The Boston office called."

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. 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.

7 Recognizing Local Coherence

The "Interpretation as Abduction" framework tells us that to interpret a text, we need to find the best explanation for the information it conveys. Among the most important information conveyed by the text is that conveyed by the mere adjacency of sentences or larger segments of discourse. The relations conveyed by adjacency are almost always some form of causality, broadly construed, a figure-ground relation (or more commonly, ground-figure), and relations dependent on similarity and identity.

In Hobbs (1985) a theory of discourse structure is outlined in which these coherence relations—among them Parallel, Elaboration, and Explanation—can hold between successive segments of a discourse and when they hold, the two segments compose into a larger segment, giving the discourse as a whole a hierarchical structure. The coherence relations can be defined in terms of the information conveyed by the segments.

Insofar as the coherence relations can be defined precisely, it is relatively straightforward to incorporate the theory into our method of interpretation as abduction. The hierarchical structure can be captured by the axiom

(18) $(\forall w, e) s(w, e) \supset Segment(w, e)$

specifying that a sentence is a discourse segment, and the axiom

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

saying that if w_1 is a segment whose assertion or topic is e_1, and w_2 is a segment asserting e_2, and a coherence relation holds between the content of w_1 and the content of w_2, then $w_1\ w_2$ is also a segment. The third argument e of *CoherenceRel* is the assertion or topic of the composed segment, as determined by the definition of the particular coherence relation.

To interpret a text W, one must then prove the expression

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

For example, Explanation is a coherence relation.

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

A first approximation to a definition for Explanation would be the following:

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

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. In explanations, what is explained is the dominant segment, so the assertion of the composed segment is simply the assertion of the first segment. (In fact, this is what "dominant segment" means.) Hence, the third argument of *CoherenceRel* above is e_1.

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

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

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 6. 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

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

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.

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

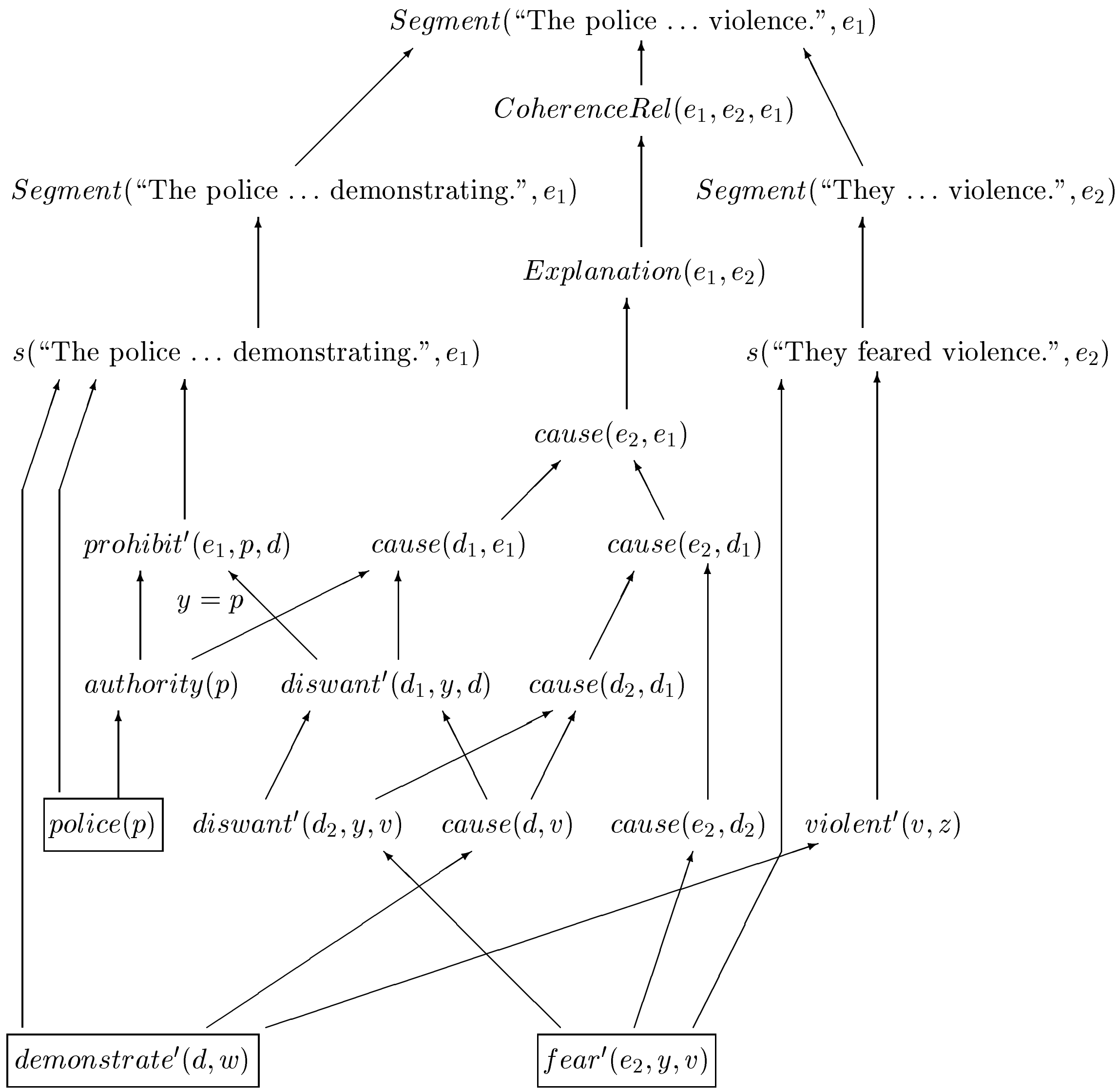

Figure 5: Interpretation of "The police prohibited the women from demonstrating. They feared violence."

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 5. Notice that in the course of doing the proof, we unify y with p, thus resolving 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.

Consider another coherence relation. A first approximation to the Elaboration relation is 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*.

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

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.

Let us consider a simple example:

> Go down First Sreet. Follow First Street to A Street.

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

To interpret the text 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 6, 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.

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.

8 Integrating Syntax, Local Coherence, and Plan Recognition

Let us return to the example of Sections 3-5. Adopting the approach of Section 6, suppose our "grammar" contains the following axiom for the structure and interpretation of wh-questions:

(24) $$(\forall\, w_1, w_2, w_3, x, a, e_5, e_6, e_8) wh\text{-}word(w_1) \wedge copula(w_2) \wedge np(w_3, x)$$
$$\wedge goal'(e_8, a, e_5) \wedge know'(e_5, a, e_6) \wedge wh'(e_6, x) \wedge speaker(a)$$
$$\supset s(w_1\ w_2\ w_3, e_8)$$

That is, if w_1 is a wh-word, w_2 is a copula, w_3 is a noun phrase referring to x, e_8 is the condition of the speaker a having the goal e_5 of knowing the essential property e_6 of x, then the concatenation of w_1, w_2, and w_3 is a sentence whose meaning is e_8.

We also know the following facts:

(25) *wh-word*("what"), *copula*("'s"), *speaker*(A)

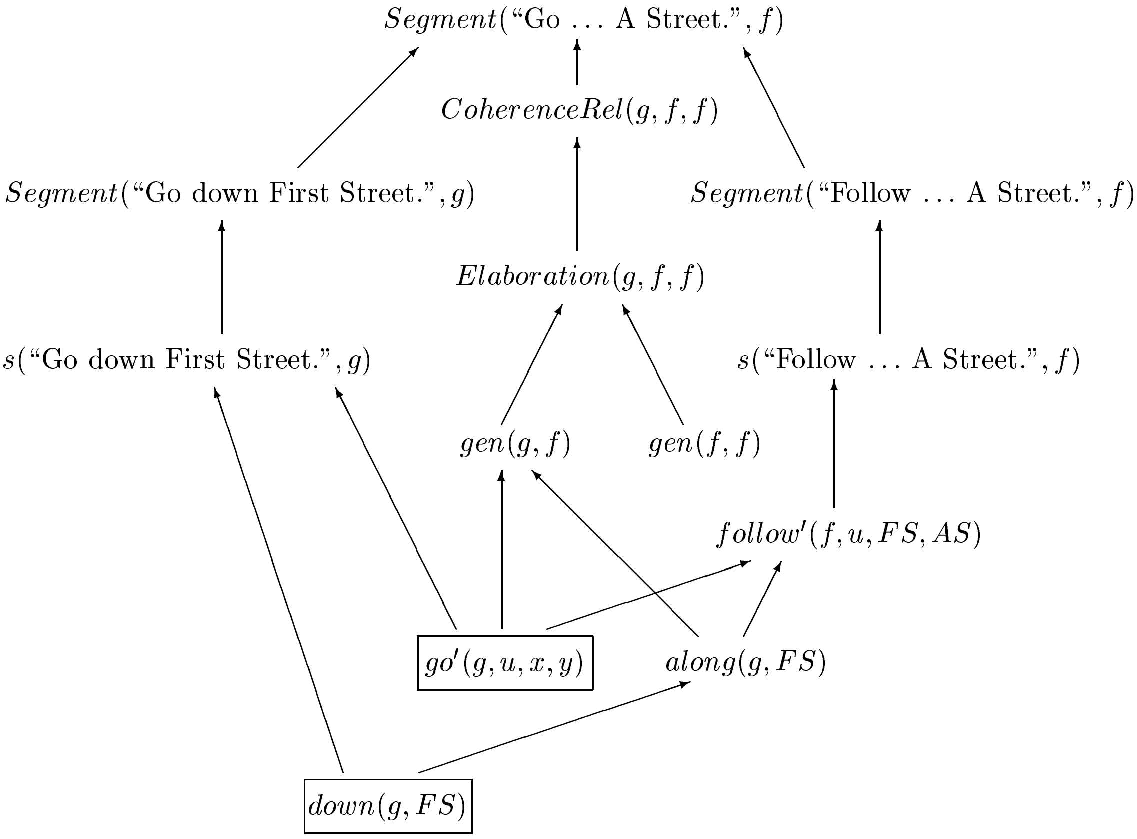

Figure 6: Interpretation of "Go down First Street. Follow First Street to A Street."

That is, "what" is a wh-word, "'s" is a copula, and A is the speaker. For completeness, we will formalize our gimmick for bypassing the reference of "a ratchet wrench" by assuming that the knowledge base also contains the literal

(26) $np(\text{"a ratchet wrench"}, RW)$

That is, the string "a ratchet wrench" is a noun phrase referring to the abstract object RW.

We now need one more axiom. The predicate *say* as used above has the content of the utterance as its final argument. We will not change this. Rather we will next introduce a predicate *utter*, which is like *say* but without the presumption of content or a hearer. Saying a meaningful segment of discourse is one example of uttering something.

(27) $(\forall\, w, e_5, e_7, a, b) Segment(w, e_5) \wedge say'(e_7, a, b, e_5) \supset utter'(e_7, a, w)$

That is, if the string of words w is a discourse segment whose content is e_5 and there is a saying e_7 of e_5 by a to b, then e_7 is an uttering by a of the string of words w. Backchaining on this axiom will allow us to explain the uttering of strings of words as the production of meaningful discourse.

Let us now redo the example. The observable to be explained is now the occurrence of the utterance.

(28) $(\exists\, e_7, a) Rexists(e_7) \wedge utter'(e_7, a,$ "What's a ratchet wrench"$)$

That is, we need to explain the existence in the real world of the event e_7 of someone a uttering the string of words "What's a ratchet wrench".

Figure 7 shows the first few steps of this proof. Using axiom (27), we hypothesize that the utterance is a saying of a contentful segment of discourse. Using axiom (18) we hypothesize that the segment of discourse is a single sentence. Using axiom (24), we unpack this into the syntactic structure and logical form of the sentence. Most of this can then be established by the facts in (25) and (26). What remains to be proved at this point is

(15) $(\exists\, e_7, a, b, e_8, e_5, e_6) Rexists(e_7) \wedge say'(e_7, a, b, e_8) \wedge goal'(e_8, a, e_5)$
$\wedge\, know'(e_5, a, e_6) \wedge wh'(e_6, RW)$

But this is just what we proved in Section 5, as illustrated in Figure 3.

9 Tautology

The framework that has been presented here gives us a handle on some of the more complex things speakers do with their utterances. Let us see how we could deal with one example—tautology.

Imagine two mothers, A and B, sitting in the playground and talking.

A: Your Johnny is certainly acting up today, isn't he?
B: Boys will be boys.

From the Informational Perspective the interpretation of B's utterance might go something like this. The sentence expresses an implicative relation between two general propositions—$boy(x)$ and $boy(x)$. This implicative relation can be proved from the reflexive property of implication. Hence, the sentence tells us nothing new.

But from a global perspective this is not the best explanation, because it leaves too much unaccounted for. There is no explanation of why B would utter this or of how it is a response to A's utterance. We may have a good explanation for the content of the sentence, but we do not have a good explanation for the saying of a sentence with that content.

This forces us into an interpretation of the content that, while not optimal locally, contributes to a global interpretation that *is* optimal. In particular, we interpret the first occurrence of "boys" extensionally as a set that includes Johnny, and we interpret the second occurrence of "boys" intensionally, as entailing the property of always acting up. So the interpretation of the sentence becomes "Members of the class that Johnny belongs

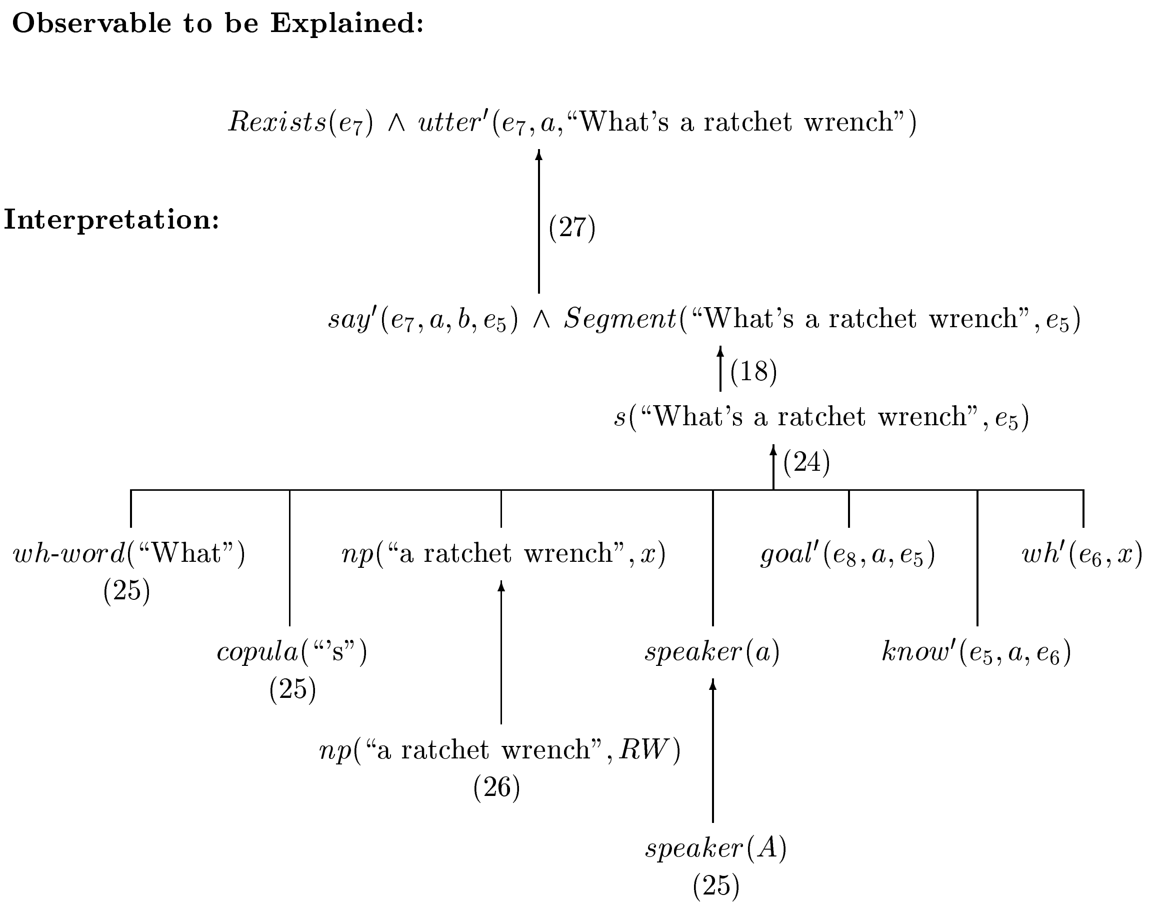

Figure 7: Syntactic Analysis and Compositional Semantics of "What's a ratchet wrench?"

to always behave in this fashion." It thus defends B against the implied accusation that she is not a good mother.

10 Plan Recognition and Local Coherence

To be written:

Using Moore and Pollack's example,

> George Bush supports big business.
> He's sure to veto House bill 1711.

give account of how *both* the Cause coherence relation (at the Informational level) and the Evidence relation (at the Intentional level) are recognized, the former in terms of knowledge of what support entails (i.e., support as a disposition to engage in certain actions) and the knowledge of the content of the bill, and the latter in terms of the speaker's knowledge that the hearer will use Modus Ponens. Then show that in the absence of knowledge of the content of bill 1711, the Cause coherence relation can be recognized by making assumptions necessary for recognizing the Evidence relation, and that in the absence of knowledge of the speaker's discourse plan, the Evidence relation can be recognized via recognition of the Cause coherence relation.

11 Discussion

Utterings of sentences are actions. As actions, utterances will generally take their place as steps in a coherent plan. This plan gives structure to the conversation as a whole. Very often, higher-level actions will correspond to larger segments of discourse, whose realizations involve the uttering of individual sentences. Following Agar and Hobbs (1982), we can call this structure the Global Coherence of the discourse. It is the structure we see from the Intentional Perspective.

But utterances are actions of particular kinds. The sentences uttered convey information; they describe situations. The situations described are also related in various ways, and, indeed, the very adjacency of the descriptions conveys that they are related. When adjacent sentences or larger segments of discourse are related in some way, the composite of the two itself constitutes a segment of the discourse, and this gives rise to a hierarchical structure for the entire discourse. This structure is what Agar and Hobbs (1982) called Local Coherence. It is the structure we see from the Informational Perspective.

The two varieties of structure are often very closely related. Indeed, quite often, especially in extended, single-speaker discourse, knowledge of the Local Coherence is the chief evidence we have for the Global Coherence. Conversely, quite often, especially in conversation, knowledge of the Global Coherence is the chief evidence we have for the Local Coherence.

There are therefore both intentional and semantic intersegment relations, and both impart their structure on the discourse. Very often but not always, these structures segment the discourse in precisely the same way.

Lumpers and spliter will find different sets of possible relations of both kinds. Lumpers like Grosz and Sidner (1986) find the Intentional relations of Dominance and Satisfaction-Precedence. A lumper like me finds the Informational relations of causality, figure-ground, and similarity. Spliters like Mann and Thompson (1986) find more fine-grained Intentional relations like Evidence and more fine-grained Informational relations like Volitional Cause. The lumper accounts are more elegant; the splitter accounts are more informative. A good review of the literature would show how the fine-grained relations are subsumed by the coarse-grained relations.

For a particular set of relations to be validated, they would have to be defined precisely. All the relevant accounts have taken substantial steps in this direction, and have substantial steps yet to be taken. Since all the relations depend on commonsense knowledge of the world in general and the behavior of people in it in particular, large amounts of commonsense knowledge must be specified formally. To show that a set of relations is complete would then require showing that that set of relations, as defined precisely in terms of the commonsense knowledge, was adequate for the interpretation of large amounts of text.

It is possible to construct grammars of discourse insofar as we can produce formal definitions of sets of relations, either Intentional or Informational, the cover the discourses we encounter.

12 Summary

The problem of interpreting discourse has been subsumed under the general problem faced by intelligent agents of interpreting the situation they are in by explaining the observable facts. The possibility of interpreting an event as the saying by an intelligent agent of a meaningful stretch of discourse is given by an axiom—axiom (27). The ways in which a stretch of discourse can be analyzed into its parts are given by axioms—axioms (18) and (19) and the axioms defining coherence relations, two of which are given in Section 7. These axioms tell us what kinds of information can be conveyed by the adjacency of discourse segments. This analysis takes us down to the level of sentences. Then the ways in which a string of words can be analyzed as a sentence are given in axioms—axioms like (16), (17) and (24). The antecedents of these axioms specify the predicate-argument relations encoded in the syntactic structures and require us to explain the propositional content of the sentence, using the background knowledge that is shared with the speaker. Meanwhile, the saying of this stretch of discourse can be related to the speaker's plan by using axioms (4) and (5), together with axioms stating what sorts of things cause and enable what other sorts of things, to see the saying event as a subgoal of some other goal, and that as the subgoal of another goal, and so on, until a link with the speaker's presumed goals is achieved. Many of these causal axioms, including axioms (10) and (11), specify the relations between communicative acts and the speaker's and hearer's mental states, which as been the focus in research on planning speech acts.

All of these axioms are expressed in a uniform fashion and used by a single process—abductive inference. Therefore, there is no problem of one "module" of the "discourse comprehension engine" communicating or interacting with another "module". Different branches of a proof graph can share variables. Thus, what is a good proof in one subgraph may not be part of a good proof of the whole. It is in this way that influence is communicated from one "module" to another. This is what happened in our analysis of the tautology and in the analysis of the example from Moore and Pollack.

We can certainly continue to *think* of, say, syntax and speaker's plan as different modules. But the distinction is entirely in our comments, not in our code.

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] Allen, James F., and C. Raymond Perrault, 1980. "Analyzing Intention in Utterances", *Artificial Intelligence*, Vol. 15, pp. 143-178.

[3] Appelt, Douglas E., and Martha E. Pollack, 1990. "Weighted Abduction for Plan Ascription", Technical Note 491, SRI International, Menlo Park, California, May 1990.

[4] Bruce, Bertram C., 1975. "Belief Systems and Language Understanding", Technical Report 2973, Bolt, Beranek, and Newman, Inc., Cambridge, Massachusetts.

[5] Cohen, Philip, and C. Raymond Perrault, 1979. "Elements of a Plan-based Theory of Speech Acts", *Cognitive Science*, Vol. 3, No. 3, pp. 177-212.

[6] Fikes, Richard, and Nils J. Nilsson, 1971. "STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving", *Artificial Intelligence*, Vol. 2, pp. 189-208.

[7] Grosz, Barbara, 1977. "The Representation and Use of Focus in Dialogue Understanding". Stanford Research Institute Technical Note 151, Stanford Research Institute, Menlo Park, California, July 1977.

[8] Grosz,Barbara, 1979. "Utterance and Objective: Issues in Natural Language Communication", *Proceedings*, Sixth International Joint Conference on Artificial Intelligence, pp. 1067-1076. Tokyo, Japan. August 1979.

[9] Grosz, Barbara, and Candy Sidner, 1986, "Attention, Intentions, and the Structure of Discourse", *Computational Linguistics*, Vol. 12, No. 3, pp. 175-204.

[10] 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.

[11] Hobbs, Jerry R. and David Andreoff Evans, 1980, "Conversation as Planned Behavior," *Cognitive Science*, Vol. 4, No. 4, pp. 349-377.

[12] Hobbs, Jerry R., Mark Stickel, Douglas Appelt, and Paul Martin, 1993. "Interpretation as Abduction", to appear in *Artificial Intelligence Journal.* Also published as SRI Technical Note 499, SRI International, Menlo Park, California. December 1990.

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

[14] Mann, William, and Sandra Thompson, 1986, "Relational Propositions in Discourse", *Discourse Processes*, Vol. 9, No. 1, pp. 57-90.

[15] Moore, Johanna D., and Martha E. Pollack, 1992. "A Problem for RST: The Need for Multi-Level Discourse Analysis", to appear in *Computational Linguistics.*

[16] Norvig, Peter, and Robert Wilensky, 1990. "A Critical Evaluation of Commensurable Abduction Models for Semantic Interpretation", in H. Karlgren, ed., *Proceedings*, Thirteenth International Conference on Computational Linguistics, Helsinki, Finland, Vol. 3, pp. 225-230, August, 1990.

[17] Pereira, Fernando C.N., and David H.D. Warren, 1980. "Definite Clause Grammars for Language Analysis - A Survey of the Formalism and a Comparison with Augmented Transition Networks", *Artificial Intelligence*, Vol. 13, pp. 231-278. Reprinted in Grosz et al. (1986), *Readings in Natural Language Processing.*

[18] Perrault, C. Raymond, 1990. "An Application of Default Logic to Speech Act Theory", in P. C. Cohen, J. Morgan, and M. E. Pollack (Eds.), *Intentions in Communication*, Bradford Books, MIT Press, Cambridge, Massachusetts, pp. 161-185.

[19] Perrault, C. Raymond, and James F. Allen, 1980. "A Plan-Based Analysis of Indirect Speech Acts", *American Journal of Computational Linguistics*, Vol. 6, No. 3-4, pp. 167-182. (July-December).

[20] Power, Richard, 1974. "A Computer Model of Conversation", Ph. D. thesis, University of Edinburgh, Scotland.

[21] Schmidt, Charles F., N. S. Sridharan, and J. L. Goodson, 1978. "The Plan Recognition Problem: An Intersection of Psychology and Artificial Intelligence", *Artificial Intelligence*, Vol. 11, pp. 45-83.

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