On the Relation between the Informational and Intentional Perspectives on Discourse

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

1 Introduction

In the paper "Interpretation as Abduction" (hereafter IA) Hobbs et al. (1993) presented and elaborated 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 necessarily involve any knowledge of the specific goals of the speaker.

Norvig and Wilensky (1990) raised 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.

A primary 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 Section 2 the IA framework is presented, in just enough detail to allow this paper to stand on its own. The interested reader should consult IA and the other cited papers for a deeper discussion of the framework. In Sections 3 and 4, two examples are given. In the first, the informational reading and the intentional reading are essentially the same. In the second, they are in conflict and the intentional reading wins out. Section 5 summarizes.

2 Background

2.1 Logical Notation

In this paper I will use the ontologically extravagant, first-order notation introduced in Hobbs (1983, 1985, 1995). For the purposes of this paper, the chief feature of this notation is the use of eventualities.

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

$$(\forall\, x)p(x) \equiv (\exists\, e)p'(e, x) \wedge Rexists(e)^1$$

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. Hobbs (1985) provides further explanation of the specific notation and ontological assumptions.

In this notation, the logical form of a sentence is a flat, scope-free, conjunction of positive literals, each of which is a predicate applied to the appropriate number of existentially quantified variables. For example, the logical form of the sentence

A man walked slowly.

is

$$(\exists\, x, e)man(x) \wedge Past(e) \wedge walk'(e, x) \wedge slow(e)$$

2.2 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. This insight can be cashed out procedurally in terms of theorem-proving technology 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.

[1] In the logical formulas in this paper, quantifiers are assumed to scope over logical operators.

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 spans 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 where 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.

Choosing the best or minimal interpretation relies on an algorithm for weighted abduction that levies variable costs for assumptions and for length of proof and reduces costs when redundancies are recognized. In IA the weighted abduction inference procedure, due to Stickel, is described in detail.

Consider a simple example.

The Boston office called.

This sentence poses at least three local pragmatics problems, the problems of resolving the definite 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)

$$(2) \quad (\exists\, x, y, z, e) call'(e, x) \wedge person(x) \wedge rel(x, y) \wedge office(y) \wedge Boston(z) \wedge nn(z, y)$$

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 y of the sentence, but x is at least related to y, or coercible from y, represented by $rel(x, y)$. y is an office and 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 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 (2) 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 $call'(e, x)$.

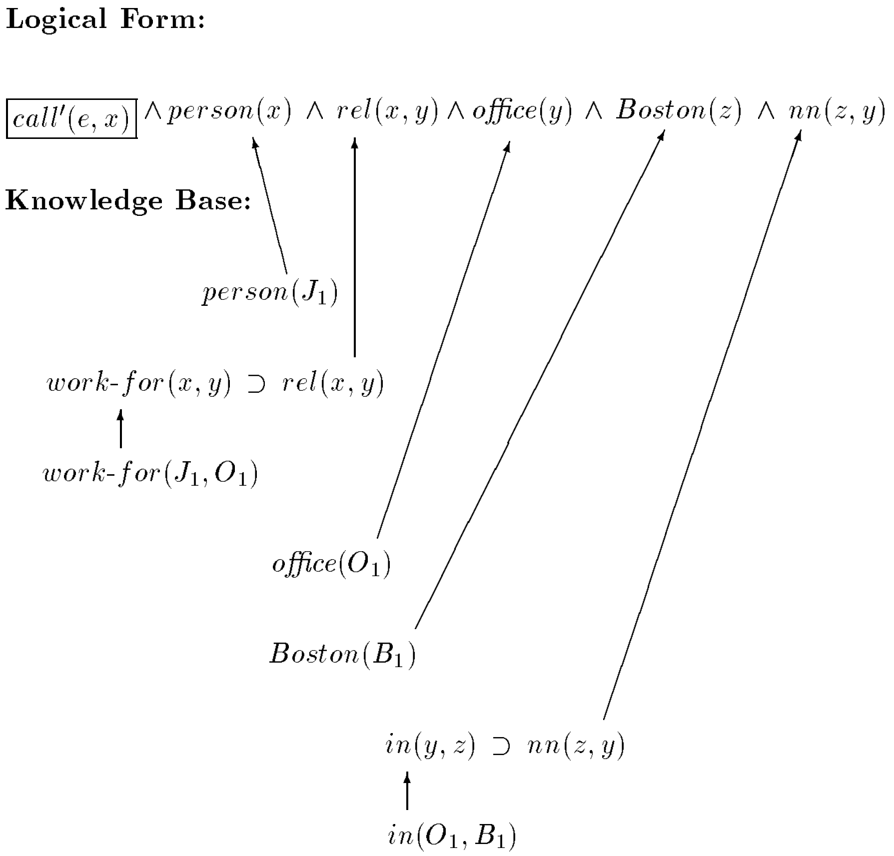

Figure 1: Interpretation of "The Boston office called."

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 a number of other examples are presented showing how this approach yields solutions to problems of syntactic and lexical ambiguity, the resolution of pronouns and implicit arguments, the interpretation of compound nominals, the expansion of metonymies, and schema recognition. Hobbs (1992) uses the abductive approach to deal with metaphor. In IA it is shown how the interpretation as abduction approach can be combined with the parsing as deduction approach to yield a smooth integration of syntax, compositional semantics, and local pragmatics. It also sketches how this approach can be extended to the recognition of the structure of discourse.

The present paper is the beginning of an effort to extend the framework to global pragmatics, that is, to the recognition of the role of the discourse in the participants' ongoing plans.

2.3 The Form of Axioms

Because of the use of eventualities, 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.

In this paper, the predicates *cause*, *enable* and *imply* will sometimes play the role of gen.

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

$$(\forall x) elephant(x) \supset mammal(x)$$

we may have axioms like

$$(\forall x) mammal(x) \wedge etc_1(x) \supset elephant(x)$$

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

We can read this axiom as saying, "One way of being a mammal is being an elephant."

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 mammals or are elephants. Intuitively, etc_1 conveys all the information one would need to know beyond mammalhood 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 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 paper, for notational convenience, I will use the simplest form of axiom I 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.

3 An Example of Plan Recognition

3.1 The Example

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.

3.2 The Informational Solution

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

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

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.

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

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 operators (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.

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

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.

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

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

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

That is, the speaker a has the goal e_5 of knowing the essential property e_6 of the ratchet wrench RW. Most of this logical form comes, of course, from our recognition that the utterance is a question.

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

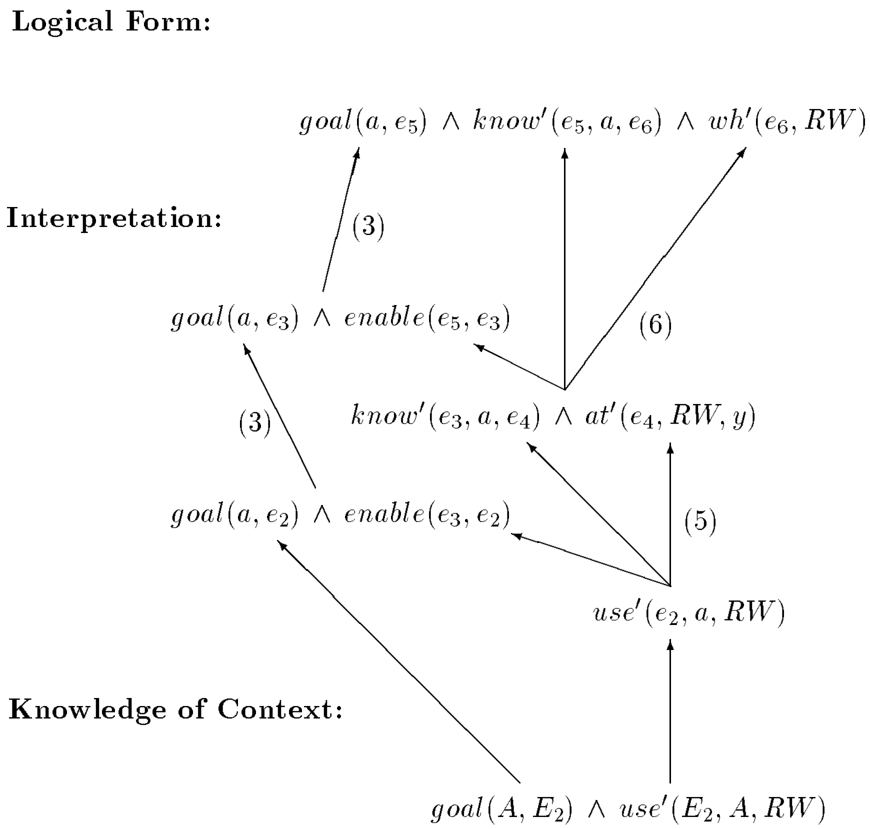

Figure 2: Informational Interpretation of "What's a ratchet wrench?"

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

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

The proof of the logical form (7) follows from axioms (3) through (6) together with fact (8), as indicated in Figure 2. Axiom (3) is used twice, first in conjunction with axiom (6) and then with axiom (5), 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 (8).

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.

3.3 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.

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

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.

(9) $$\begin{aligned}&(\forall\, e_k, y, e) know'(e_k, y, e) \wedge etc_2(e_k, y, e)\\ &\qquad \supset (\exists\, e_s, x) cause(e_s, e_k) \wedge say'(e_s, x, y, e)\end{aligned}$$

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.

$$\begin{aligned}&(\forall\, e_5, e_8, e_9, e_{10}, a, b) know'(e_9, b, e_8) \wedge goal'(e_8, a, e_5) \wedge cause(e_{10}, e_5)\\ &\qquad \wedge p'(e_{10}, b) \wedge etc_3(e_5, e_8, e_9, e_{10}, a, b) \supset cause(e_9, e_{10})\end{aligned}$$

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.[2]

The next axiom schema is the converse of this. It is a kind of attribution of cooperation.

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

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.

$$(11)\quad (\forall\, a, e_7) goal(a, e_7) \wedge p'(e_7, a) \wedge etc_5(a, e_7) \supset Rexists(e_7)$$

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

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

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 (11) we attribute intention to explain the occurrence of the utterance act e_7; it's not like a sneeze. Using axiom (4), we hypothesize that this intention or goal is a subgoal of some other goal e_9. Using axiom (9), 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 (4) again, we hypothesize that e_9 is a subgoal of some other goal e_{10}, and using axiom (10) 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 (4) and (9) 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

[2]More properly, where I have $p'(e_{10}, b)$ I should have $agent(b, e_{10})$, together with a set of axioms of the form

$$(\forall\, e, x) p'(e, x, \ldots) \supset agent(x, e)$$

Observable to be Explained:

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

Interpretation:

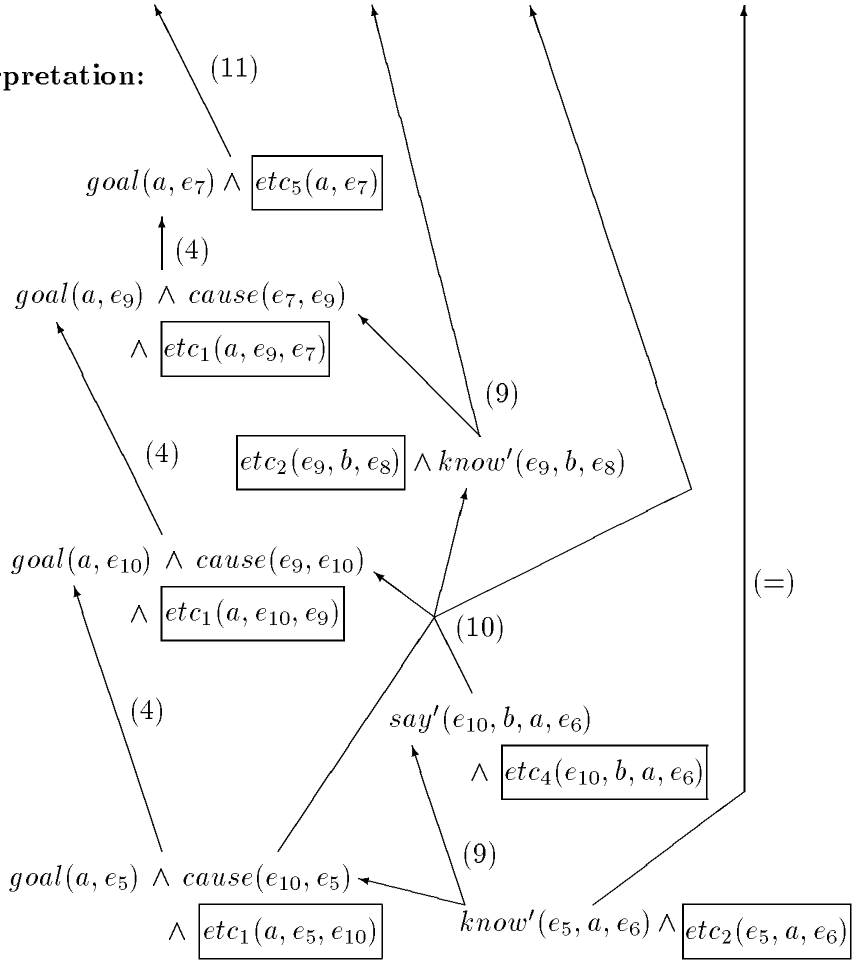

Figure 3: Intentional Interpretation of "What's a ratchet wrench?"

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.

4 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, such as "boys will be boys," "fair is fair," and "a job is a job." Norvig and Wilensky (1990) cite this figure of speech as something that should cause trouble for an abduction approach that seeks minimal explanations, since the minimal explanation is that they just express a known truth. Such an explanation requires no assumptions at all.

In fact, the phenomenon is a good example of why an informational account of discourse interpretation has to be embedded in an intentional account. Let us 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.

In order to avoid dealing with the complications of plurals and tense in this example, let us simplify B's utterance to

B: A boy is a boy.

From the Informational Perspective, several interpretations of this utterance are possible. The first is the Literal Extensional Interpretation. The first "a boy" introduces a specific, previously unidentified boy and the second says about him that he is a boy. The second is the Literal Intensional Interpretation. The sentence expresses a trivial implicative relation between two general propositions—$boy(x)$ and $boy(x)$. The third is the Desired Interpretation. The first "a boy" identifies the typical member of a class which Johnny is a member of and the second conveys a general property, "being a boy", as a way of conveying a specific property, "misbehaving", which is true of members of that class.

More precisely, the logical form of the sentence can be written as follows:

$$(\exists\, e_1, e_2, x, y, z, w) boy'(e_1, x) \;\wedge\; rel(z, x) \;\wedge\; be(z, w) \;\wedge\; rel(w, y) \;\wedge\; boy'(e_2, y)$$

The sentence expresses a *be* relation between two entities, but either or both of its arguments may be subject to coercion. Thus, we have introduced the two *rel* relations. The logical form can be given the tortured paraphrase, "z is w, where z is related to x whose boy-ness is e_1 and w is related to y whose boy-ness is e_2."

The required axioms are as follows:

Everything is itself:

$$(\forall\, x) be(x, x)$$

Implication can be expressed by "to be":

$$(\forall\, e_1, e_2) imply(e_1, e_2) \supset be(e_1, e_2)$$

Implication is reflexive:

$$(\forall\, e) imply(e, e)$$

Boys misbehave:

$$(\forall\, e_1, x) boy'(e_1, x) \supset (\exists\, e_3) misbehave'(e_3, x) \wedge imply(e_1, e_3)$$

Misbehavers are often boys:

$$(\forall\, e_3, x) misbehave'(e_3, x) \wedge etc_1(x) \supset (\exists\, e_2) boy'(e_2, x)$$

Identity is a possible coercion relation:

$$(\forall\, x) rel(x, x)$$

An entity can be coerced into a property of the entity:

$$(\forall\, e, x) boy'(e, x) \supset rel(e, x)$$
$$(\forall\, e, x) misbehave'(e, x) \supset rel(e, x)$$

Now the Literal Extensional Interpretation is established by taking the two coercion relations to be identity, taking *be* to be expressing identity, and assuming $boy(e_1, x)$ (or equivalently, $boy(e_2, y)$).

In the Literal Intensional Interpretation, z is identified with e_1, w is identified with e_2, and $boy'(e_1, x)$ and $boy'(e_2, y)$ are taken to be the two coercion relations. Then e_2 is identified with e_1 and $be(e_1, e_1)$ is interpreted as a consequence of $imply(e_1, e_1)$. Again, $boy(e_1, x)$ is assumed.

In the Desired Interpretation, the first coercion relation is taken to be $boy'(e_1, x)$, identifying z as e_1. The second coercion relation is taken to be $misbehave'(e_3, y)$, identifying w as e_3. If $etc_1(y)$ is assumed, then $misbehave'(e_3, y)$ explains $boy(e_2, y)$. If $boy(e_1, x)$ is assumed, it can explain $misbehave'(e_3, y)$, identifying x and y, and also $imply(e_1, e_3)$. The latter explains $be(e_1, e_3)$.

From the Informational Perspective alone, the Literal Extensional Interpretation is minimal and hence would be favored. The Desired Interpretation is the worst of the three.

But the Literal Extensional and Intensional Interpretations leave the *fact* of the utterance unaccounted for. From the Intentional Perspective, this is what we need to explain. The explanation would run something like this:

B wants A to believe that B is not responsible for Johnny's misbehaving.

Thus, B wants A to believe that Johnny misbehaves necessarily.

Thus, given that Johnny is necessarily a boy, B wants A to believe that Johnny's being a boy implies that he misbehaves.

Thus, B wants to convey to A that being a boy implies misbehaving.

Thus, given that boy-ness implies misbehaving is a possible interpretation of a boy being a boy, B wants to say to A that a boy is a boy.

The content of the utterance under the Literal Extensional and Intensional Interpretations do not lend themselves to explanations for that fact of the utterance in the way that the Desired Interpretation does. The requirement for the *globally* minimal explanation in the Intentional Perspective, that is, the requirement that both the content and the fact of the utterance must be explained, forces us into an interpretation of the content that would not be favored from the Informational Perspective alone.

The two literal interpretations give us good explanations for the content of the sentence but do not give us good explanations for the saying of a sentence with that content. The Desired Interpretation, however, does fit into an explanation of why the utterance was uttered. It can be paraphrased as "Members of the class that Johnny belongs to always behave in this fashion," and it thus defends B against the implied accusation that she is not a good mother.

We are forced into an interpretation of the content that, while not optimal locally, contributes to a global interpretation that *is* optimal.

5 Summary

For a discourse consisting of assertions from a speaker A to a hearer B,[3] the relation between the intentional account of interpretation and the informational account can be summarized succinctly by the following formula:

(12) $$\textbf{intentional-account} = goal(A, believe(B, \textbf{informational-account}))$$

The speaker ostensibly has the goal of changing the beliefs of the hearer to include the content characterized by the informational account. When we reason about the speaker's intention, we are reasoning about how this goal fits into the larger picture of her ongoing plan. We are asking why she seems to be trying to get the hearer to believe this content.

In some cases there is a strong correspondence between the two accounts. The content is something that is reasonable to believe and it is easy to see why the speaker wants the hearer to believe it in the given situation. In other cases, there is very little information about one account or the other. In pragmatically elliptical utterances, the informational account is highly underdetermined and the global interpretation is thus primarily shaped by the intentional account. In beginnings, chance encounters with strangers, and random

[3] Similar schemas could be given for questions and orders.

remarks, the hearer can guess very little about the intention of the speaker except via the informational account, so that is primarily what shapes the global interpretation. Finally, there are cases of genuine conflict between the two accounts. The informational account does not seem to be true, or it seems to run counter to the speaker's goals for the hearer to come to believe it, or it ought to be obvious that the hearer already does believe it. Tautologies are an example of the last of these cases. In these cases, the intentional account may force an alternate, ordinarily nonoptimal informational reading, or the hearer may be forced to reassess the speaker's goals.

The framework presented in IA gives a uniform account of how one determines the informational interpretation of a sentence or larger stretch of text, in terms of abductive inference. The present paper begins to extend that framework to encompass intentional interpretation as well.

Schema (12) points to a third major area of research, beyond the determination of intention and information—the problem of belief revision. Given that the speaker's goal is to get the hearer to believe the informational reading of the utterance, under what circumstances should the hearer actually come to believe it. This, however, is an issue for another paper.

Acknowledgments

The research described here was supported by the National Science Foundation and the Advanced Research Projects Agency under Grant IRI-9314961 (Integrated Techniques for Generation and Interpretation).

References

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

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

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

[9] Hobbs, Jerry R., 1983. "An Improper Treatment of Quantification in Ordinary English", *Proceedings of the 21st Annual Meeting, Association for Computational Linguistics*, pp. 57-63. Cambridge, Massachusetts, June 1983.

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

[11] Hobbs, Jerry R., 1992. "Metaphor and Abduction", in A. Ortony, J. Slack, and O. Stock, eds., *Communication from an Artificial Intelligence Perspective: Theoretical and Applied Issues*, Springer-Verlag, Berlin, pp. 35-58. Also published as SRI Technical Note 508, SRI International, Menlo Park, California. August 1991.

[12] Hobbs, Jerry R., 1995. "Monotone Decreasing Quantifiers in a Scope-Free Logical Form", to appear in K. van Deemter and S. Peters (Eds.), *Semantic Ambiguity and Underspecification*, CSLI Lecture Notes, Stanford, California, 1995.

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

[14] 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.

[15] 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.

[16] Moore, Johanna D., and Martha E. Pollack, 1992. "A Problem for RST: The Need for Multi-Level Discourse Analysis", *Computational Linguistics*, Vol. 18, No. 4, December 1992.

[17] 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.

[18] 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).

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

[20] 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.