Interpretation as Abduction

Jerry R. Hobbs, Mark Stickel,
Douglas Appelt, and Paul Martin

Artificial Intelligence Center
SRI International

Abstract

Abduction is inference to the best explanation. In the TACITUS project at SRI we have developed an approach to abductive inference, called "weighted abduction", that has resulted in a significant simplification of how the problem of interpreting texts is conceptualized. The interpretation of a text is the minimal explanation of why the text would be true. More precisely, to interpret a text, one must prove the logical form of the text from what is already mutually known, allowing for coercions, merging redundancies where possible, and making assumptions where necessary. It is shown how such "local pragmatics" problems as reference resolution, the interpretation of compound nominals, the resolution of syntactic ambiguity and metonymy, and schema recognition can be solved in this manner. Moreover, this approach of "interpretation as abduction" can be combined with the older view of "parsing as deduction" to produce an elegant and thorough integration of syntax, semantics, and pragmatics, one that spans the range of linguistic phenomena from phonology to discourse structure and accommodates both interpretation and generation. Finally, we discuss means for making the abduction process efficient, possibilities for extending the approach to other pragmatics phenomena, and the semantics of the weights and costs in the abduction scheme.

1 Introduction

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. In the TACITUS Project at SRI, we have developed a scheme for abductive inference that yields a significant simplification in the description of such interpretation processes and a significant extension of the range of phenomena that can be captured. It has been implemented in the TACITUS System (Hobbs, 1986; Hobbs and Martin, 1987) and has been or is being used to solve a variety of interpretation problems in several kinds of messages, including equipment failure reports, naval operations reports, and terrorist reports.

It is a commonplace that people understand discourse so well because they know so much. Accordingly, the aim of the TACITUS Project has been to investigate how knowledge is used in the interpretation of discourse. This has involved building a large knowledge base of commonsense and domain knowledge (see Hobbs et al., 1987), and developing procedures for using this knowledge for the interpretation of discourse. In the latter effort, we have concentrated on problems in "local pragmatics", specifically, the problems of reference resolution, the interpretation of compound nominals, the resolution of some kinds of syntactic ambiguity, and metonymy resolution. Our approach to these problems is the focus of the first part of this paper.

In the framework we have developed, what the interpretation of a sentence is can be described very concisely:

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 stands with one foot in mutual belief and one foot in 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.[1] 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.[2]

Consider a simple example.

(2) The Boston office called.

---

[1] This is clearest in the case of assertions. But questions and commands can also be conceived of as primarily conveying information—about the speaker's wishes. In any case, most of what is required to interpret the three sentences,

John called the Boston office.
Did John call the Boston office?
John, call the Boston office.

is the same.

[2] Interpreting indirect speech acts, such as "It's cold in here," meaning "Close the window," is not a counterexample to the principle that the minimal interpretation is the best interpretation, but rather can be seen as a matter of achieving the minimal interpretation coherent with the interests of the speaker. More on this in Section 8.2.

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 that contains the following facts:

$$Boston(B_1)$$

that is, B_1 is the city of Boston.

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

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

$$person(J_1)$$

that is, John J_1 is a person.

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

that is, John J_1 works for the office O_1.

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

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

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

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

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

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 Section 2 of this paper, we give a high-level overview of the TACITUS system, in which this method of interpretation is implemented. In Section 3, we justify the first clause of the above characterization by showing in a more detailed fashion that solving local pragmatics problems is equivalent to proving the logical form plus the constraints. In Section 4, we justify the last two clauses by describing our scheme of abductive inference. In Section 5 we present several examples. In Section 6 we show how the idea of interpretation as abduction can be combined with the older idea of parsing as deduction to yield a thorough and elegant integration of syntax, semantics, and pragmatics, that works for both interpretation and generation. In Section 7 we discuss related work. In Section 8 we discuss three kinds of future directions, improving the efficiency, extending the coverage, and devising a principled semantics for the abduction scheme.

2 The TACITUS System

TACITUS stands for The Abductive Commonsense Inference Text Understanding System. It is intended for processing messages and other texts for a variety of purposes, including message routing and prioritizing, problem monitoring, and database entry and diagnosis on the basis of the information in the texts. It has been used for three applications so far:

1. Equipment failure reports or casualty reports (casreps). These are short, telegraphic messages about breakdowns in machinery. The application is to perform a diagnosis on the basis of the information in the message.

2. Naval operation reports (opreps). These are telegraphic messages about ships attacking other ships, of from one to ten sentences, each of from one to thirty words, generated in the midst of naval exercises. There are frequent misspellings and uses of jargon, and there are more sentence fragments than grammatical sentences. The application is to produce database entries saying who did what to whom, with what instrument, when, where, and with what result.

3. Newspaper articles and similar texts on terrorist activities. The application is again to produce database entries.

To give the reader a concrete sense of these applications, we give an example of the input and output of the system for a relatively simple text. One sentence from the terrorist reports is

> Bombs exploded at the offices of French-owned firms in Catalonia, causing serious damage.

The corresponding database entries are

Incident Type:	Bombing
Incident Country:	Spain
Responsible Organization:	—
Target Nationality:	France
Target Type:	Commercial
Property Damage:	Some Damage

There is an incident of type Bombing. The incident country is Spain, since Catalonia is a part of Spain. There is no information about what organization is responsible. The target type is Commercial, since it was firms that were attacked, and the target nationality was France, since the firms are owned by the French. Finally, there is some level of property damage.

The naval operation reports is the application that has been developed most extensively. The system has been evaluated on a corpus of naval operation reports. Recall is defined as the number of correct items the system enters into the database, divided by the total number of items it should have entered. The recall for TACITUS on the full set of 130 opreps was 47%. Error rate is the percent of incorrect database entries proposed by the system. The error rate was 8%. There is very little that is general that one could say about the nature of the misses and errors. We specifically targeted 20 of the messages and tried to eliminate the bugs that those messages revealed, without attempting to extend the power of the system in any significant way. After we did this, the recall for the 20 messages was 72% and the error rate was 5%. It was our estimate that with several more months of work on the system we could raise the recall for the full corpus to above 80%, keeping the error rate at 5% or below. At that point we would encounter some of the hard problems, where equipping the system with the necessary knowledge would threaten its efficiency, or where phenomena not currently handled, such as semantic parallelism between sentences, would have to be dealt with.

The system, as it is presently constructed, consists of three components: the syntactic analysis and semantic translation component, the pragmatics component, and the task component. How the pragmatics component works is the topic of Sections 3, 4, and 8.1. Here we describe the other two components very briefly.

The syntactic analysis and semantic translation is done by the DIALOGIC system. DIALOGIC includes a large grammar of English that was constructed in 1980 and 1981 essentially by merging the DIAGRAM grammar of Robinson (1982) with the Linguistic String Project grammar of Sager (1981), including semantic translators for all the rules. It has since undergone further development. Its coverage encompasses all of the major syntactic structures of English, including sentential complements, adverbials, relative clauses, and the most common conjunction constructions. Selectional constraints can be encoded and applied in either a hard mode that rejects parses or in a soft mode that orders parses. A list of possible intra- and inter-sentential antecedents for pronouns is produced, ordered by syntactic criteria. There are a number of heuristics for ordering parses on the basis of syntactic criteria (Hobbs and Bear, 1990). Optionally, the system can produce neutral representations for the most common cases of structural ambiguity (Bear and Hobbs, 1988). DIALOGIC produces a logical form for the sentence in an ontologically promiscuous version of first-order predicate calculus (Hobbs, 1985a), encoding everything that can be determined by purely syntactic means, without recourse to the context or to world knowledge.

This initial logical form is passed to the pragmatics component, which works as described below, to produce an elaborated logical form, making explicit the inferences and assumptions required for interpreting the text and the coreference relations that are discovered in interpretation.

On the basis of the information in the elaborated logical form, the task component produces the required output, for example, the diagnosis or the database entries. The task component is generally fairly small because all of the relevant information has been made explicit by the pragmatics component. The task component is programmed in a schema-specification language that is a slight extension of first-order predicate calculus (Tyson and Hobbs, 1990).

TACITUS is intended to be largely domain- and application-independent. The lexicon used by DIALOGIC and the knowledge base used by the pragmatics component must of course vary from domain to domain, but the grammar itself and the pragmatics procedure do not vary from one domain to the next. The task component varies from application to application, but the use of the schema-specification language makes even this component largely domain-independent.

This modular organization of the system into syntax, pragmatics, and task is undercut in Section 5. There we propose a unified framework that incorporates all three modules. The framework has been implemented, however, only in a preliminary experimental manner.

3 Local Pragmatics

The four local pragmatics problems we have concentrated on so far can be illustrated by the following "sentence" from an equipment failure report:

(4) Disengaged compressor after lube-oil alarm.

Identifying the compressor and the alarm are **reference resolution** problems. Determining the implicit relation between "lube-oil" and "alarm" is the problem of **compound nominal interpretation**. Deciding whether "after lube-oil alarm" modifies the compressor or the disengaging is a problem in **syntactic ambiguity resolution**. The preposition "after" requires an event or condition as its object and this forces us to coerce "lube-oil alarm" into "the sounding of the lube-oil alarm"; this is an example of **metonymy resolution**. We wish to show that solving the first three of these problems amounts to deriving the logical form of the sentence. Solving the fourth amounts to deriving the constraints predicates impose on their arguments, allowing for coercions. Thus, to solve all of them is to interpret them according to characterization (1). For each of these problems, our approach is to frame a logical expression whose derivation, or proof, constitutes an interpretation.

Reference: To resolve the reference of "compressor" in sentence (4), we need to prove (constructively) the following logical expression:

(5) $(\exists\, c) compressor(c)$

If, for example, we prove this expression by using axioms that say C_1 is a "starting air compressor",[3] and that a starting air compressor is a compressor, then we have resolved

---

[3] That is, a compressor for the air used to start the ship's gas turbine engines.

the reference of "compressor" to C_1.

In general, we would expect definite noun phrases to refer to entities the hearer already knows about and can identify, and indefinite noun phrases to refer to new entities the speaker is introducing. However, in the casualty reports most noun phrases have no determiners. There are sentences, such as

> Retained oil sample and filter for future analysis.

where "sample" is indefinite, or new information, and "filter" is definite, or already known to the hearer. In this case, we try to prove the existence of both the sample and the filter. When we fail to prove the existence of the sample, we know that it is new, and we simply assume its existence.

Elements in a sentence other than nominals can also function referentially. In

> Alarm sounded.
> Alarm activated during routine start of compressor.

one can argue that the activation is the same as, or at least implicit in, the sounding. Hence, in addition to trying to derive expressions such as (5) for nominal reference, for possible non-nominal reference we try to prove similar expressions.

$$(\exists \ldots e, a, \ldots) \ldots \wedge \; activate'(e, a) \; \wedge \; \ldots^{4}$$

That is, we wish to derive the existence, from background knowledge or the previous text, of some known or implied activation. Most, but certainly not all, information conveyed non-nominally is new, and hence will be assumed by means described in Section 4.

Compound Nominals: To resolve the reference of the noun phrase "lube-oil alarm", we need to find two entities o and a with the appropriate properties. The entity o must be lube oil, a must be an alarm, and there must be some implicit relation between them. If we call that implicit relation nn, then the expression that must be proved is

$$(\exists\, o, a, nn) lube\text{-}oil(o) \;\wedge\; alarm(a) \;\wedge\; nn(o, a)$$

In the proof, instantiating nn amounts to interpreting the implicit relation between the two nouns in the compound nominal. Compound nominal interpretation is thus just a special case of reference resolution.

Treating nn as a predicate variable in this way assumes that the relation between the two nouns can be anything, and there are good reasons for believing this to be the case (e.g., Downing, 1977). In "lube-oil alarm", for example, the relation is

$$\lambda x, y \; [y \text{ sounds when the pressure of } x \text{ drops too low}]$$

---

[4] Read this as "e is the activation of a." This is an example of a notational convention used throughout this article. Very briefly, where $p(x)$ says that p is true of x, $p'(e, x)$ says that e is the eventuality or possible situation of p being true of x. The unprimed and primed predicates are related by the axiom schema $(\forall\, x) p(x) \equiv (\exists\, e) p'(e, x) \wedge Rexists(e)$ where $Rexists(e)$ says that the eventuality e does in fact really exist. See Hobbs (1985a) for further explanation of this notation for events.

However, in our implementation we use a first-order simulation of this approach. The symbol nn is treated as a predicate constant, and the most common possible relations (see Levi, 1978) are encoded in axioms. The axiom

$$(\forall x, y) part(y, x) \supset nn(x, y)$$

allows interpretation of compound nominals of the form "<whole> <part>", such as "filter element". Axioms of the form

$$(\forall x, y) sample(y, x) \supset nn(x, y)$$

handle the very common case in which the head noun is a relational noun and the prenominal noun fills one of its roles, as in "oil sample". Complex relations such as the one in "lube-oil alarm" can sometimes be glossed as "for".

$$(\forall x, y) for(y, x) \supset nn(x, y)$$

Syntactic Ambiguity: Some of the most common types of syntactic ambiguity, including prepositional phrase and other attachment ambiguities and very compound nominal ambiguities[5], can be converted into constrained coreference problems (see Bear and Hobbs, 1988). For example, in (4) the first argument of *after* is taken to be an existentially quantified variable which is equal to either the compressor or the disengaging event. The logical form would thus include

$$(\exists \ldots e, c, y, a, \ldots) \ldots \wedge \mathit{after}(y, a) \wedge y \in \{c, e\} \wedge \ldots$$

That is, no matter how $\mathit{after}(y, a)$ is proved or assumed, y must be equal to either the compressor c or the disengaging e. This kind of ambiguity is often solved as a by-product of the resolution of metonymy or of the merging of redundancies.

Metonymy: Predicates impose constraints on their arguments that are often violated. When they are violated, the arguments must be coerced into something related that satisfies the constraints. This is the process of metonymy resolution.[6] Let us suppose, for example, that in sentence (4), the predicate *after* requires its arguments to be events:

$$\mathit{after}(e_1, e_2) : \mathit{event}(e_1) \wedge \mathit{event}(e_2)$$

To allow for coercions, the logical form of the sentence is altered by replacing the explicit arguments by "coercion variables" which satisfy the constraints and which are related somehow to the explicit arguments. Thus the altered logical form for (4) would include

$$\begin{array}{l}(\exists \ldots k_1, k_2, y, a, rel_1, rel_2, \ldots) \ldots \wedge \mathit{after}(k_1, k_2) \wedge \mathit{event}(k_1) \wedge rel_1(k_1, y) \\ \quad \wedge \mathit{event}(k_2) \wedge rel_2(k_2, a) \wedge \ldots\end{array}$$

---

[5] A very compound nominal is a string of two or more nouns preceding a head noun, as in "Stanford Research Institute". The ambiguity they pose is whether the first noun is taken to modify the second or the third.

[6] There are other interpretive moves in this situation besides metonymic interpretation, such as metaphoric interpretation. For the present article, we will confine ourselves to metonymy, however.

Here, k_1 and k_2 are the coercion variables, and the *after* relation obtains between them, rather than between y and a. k_1 and k_2 are both events, and k_1 and k_2 are coercible from y and a, respectively.

As in the most general approach to compound nominal interpretation, this treatment is second-order, and suggests that any relation at all can hold between the implicit and explicit arguments. Nunberg (1978), among others, has in fact argued just this point. However, in our implementation, we are using a first-order simulation. The predicate constant *rel* is treated as a predicate constant, and there are a number of axioms that specify what the possible coercions are. Identity is one possible relation, since the explicit arguments could in fact satisfy the constraints:

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

In general, where this works, it will lead to the best interpretation. We can also coerce from a whole to a part and from an object to its function. Hence,

$$(\forall x,y) part(x,y) \supset rel(x,y)$$
$$(\forall x,e) function(e,x) \supset rel(e,x)$$

Putting it all together, we find that to solve all the local pragmatics problems posed by sentence (4), we must derive the following expression:

$$\begin{aligned}&(\exists\, e,x,c,k_1,k_2,y,a,o) Past(e) \wedge disengage'(e,x,c) \wedge compressor(c)\\ &\quad \wedge after(k_1,k_2) \wedge event(k_1) \wedge rel(k_1,y) \wedge y \in \{c,e\}\\ &\quad \wedge event(k_2) \wedge rel(k_2,a) \wedge alarm(a) \wedge nn(o,a) \wedge lube\text{-}oil(o)\end{aligned}$$

But this is just the logical form of the sentence[7] together with the constraints that predicates impose on their arguments, allowing for coercions. That is, it is the first half of our characterization (1) of what it is to interpret a sentence.

When parts of this expression cannot be derived, assumptions must be made, and these assumptions are taken to be the new information. The likelihood that different conjuncts in this expression will be new information varies according to how the information is presented, linguistically. The main verb is more likely to convey new information than a definite noun phrase. Thus, we assign a cost to each of the conjuncts—the cost of assuming that conjunct. This cost is expressed in the same currency in which other factors involved in the "goodness" of an interpretation are expressed; among these factors are likely to be the length of the proofs used and the salience of the axioms they rely on. Since a definite noun phrase is generally used referentially, an interpretation that simply assumes the existence of the referent and thus fails to identify it should be an expensive one. It is therefore given a high assumability cost. For purposes of concreteness, let's just call this \$10. Indefinite noun phrases are not usually used referentially, so they are given a low cost, say, \$1. Bare noun phrases are given an intermediate cost, say, \$5. Propositions presented non-nominally are usually new information, so they are given a low cost, say, \$3. One does not usually use selectional constraints to convey new information, so they

---

[7] For justification for this kind of logical form for sentences with quantifiers and intensional operators, see Hobbs(1983b, 1985a).

are given the same cost as definite noun phrases. Coercion relations and the compound nominal relations are given a very high cost, say \$20, since to assume them is to fail to solve the interpretation problem. If we place the assumability costs as superscripts on their conjuncts in the above logical form, we get the following expression:

$$\begin{aligned}&(\exists\, e, x, c, k_1, k_2, y, a, o) Past(e)^{\$3} \wedge disengage'(e, x, c)^{\$3} \wedge compressor(c)^{\$5}\\ &\quad \wedge after(k_1, k_2)^{\$3} \wedge event(k_1)^{\$10} \wedge rel(k_1, y)^{\$20} \wedge y \in \{c, e\} \wedge event(k_2)^{\$10}\\ &\quad \wedge rel(k_2, a)^{\$20} \wedge alarm(a)^{\$5} \wedge nn(o, a)^{\$20} \wedge lube\text{-}oil(o)^{\$5}\end{aligned}$$

While this example gives a rough idea of the relative assumability costs, the real costs must mesh well with the inference processes and thus must be determined experimentally. The use of numbers here and throughout the next section constitutes one possible regime with the needed properties. This issue is addressed more fully in Section 8.3.

4 Weighted Abduction

In deduction, from $(\forall\, x) p(x) \supset q(x)$ and $p(A)$, one concludes $q(A)$. In induction, from $p(A)$ and $q(A)$, or more likely, from a number of instances of $p(A)$ and $q(A)$, one concludes $(\forall\, x) p(x) \supset q(x)$. Abduction is the third possibility. From $(\forall\, x) p(x) \supset q(x)$ and $q(A)$, one concludes $p(A)$. One can think of $q(A)$ as the observable evidence, of $(\forall\, x) p(x) \supset q(x)$ as a general principle that could explain $q(A)$'s occurrence, and of $p(A)$ as the inferred, underlying cause or explanation of $q(A)$. Of course, this mode of inference is not valid; there may be many possible such $p(A)$'s. Therefore, other criteria are needed to choose among the possibilities.

One obvious criterion is the consistency of $p(A)$ with the rest of what one knows. Two other criteria are what Thagard (1978) has called *simplicity* and *consilience*. Roughly, simplicity is that $p(A)$ should be as small as possible, and consilience is that $q(A)$ should be as big as possible. We want to get more bang for the buck, where $q(A)$ is bang, and $p(A)$ is buck.

There is a property of natural language discourse, noticed by a number of linguists (e.g., Joos, 1972; Wilks, 1972), that suggests a role for simplicity and consilience in interpretation—its high degree of redundancy. Consider

> Inspection of oil filter revealed metal particles.

An inspection is a looking at that *causes one to learn* a property relevant to the *function* of the inspected object. The *function* of a filter is to capture *particles* from a fluid. To reveal is to *cause one to learn*. If we assume the two causings to learn are identical, the two sets of particles are identical, and the two functions are identical, then we have explained the sentence in a minimal fashion. Because we have exploited this redundancy, a small number of inferences and assumptions (simplicity) have explained a large number of syntactically independent propositions in the sentence (consilience). As a by-product, we have moreover shown that the inspector is the one to whom the particles are revealed and that the particles are in the filter, facts which are not explicitly conveyed by the sentence.

Another issue that arises in abduction in choosing among potential explanations is what might be called the "informativeness-correctness tradeoff". Many previous uses of abduction in AI from a theorem-proving perspective have been in diagnostic reasoning (e.g., Pople, 1973; Cox and Pietrzykowski, 1986), and they have assumed "most-specific abduction". If we wish to explain chest pains, it is not sufficient to assume the cause is simply chest pains. We want something more specific, such as "pneumonia". We want the most specific possible explanation. In natural language processing, however, we often want the least specific assumption. If there is a mention of a fluid, we do not necessarily want to assume it is lube oil. Assuming simply the existence of a fluid may be the best we can do.[8] However, if there is corroborating evidence, we may want to make a more specific assumption. In

Alarm sounded. Flow obstructed.

we know the alarm is for the lube oil pressure, and this provides evidence that the flow is not merely of a fluid but of lube oil. The more specific our assumptions are, the more informative our interpretation is. The less specific they are, the more likely they are to be correct.

We therefore need a scheme of abductive inference with three features. First, it should be possible for goal expressions to be assumable, at varying costs. Second, there should be the possibility of making assumptions at various levels of specificity. Third, there should be a way of exploiting the natural redundancy of texts.

We have devised just such an abduction scheme.[9] First, every conjunct in the logical form of the sentence is given an assumability cost, as described at the end of Section 3. Second, this cost is passed back to the antecedents in Horn clauses by assigning weights to them. Axioms are stated in the form

(6) $P_1^{w_1} \wedge P_2^{w_2} \supset Q$

This says that P_1 and P_2 imply Q, but also that if the cost of assuming Q is c, then the cost of assuming P_1 is $w_1 c$, and the cost of assuming P_2 is $w_2 c$.[10] Third, factoring or synthesis is allowed. That is, goal expressions may be unified, in which case the resulting expression is given the smaller of the costs of the input expressions. Thus, if the goal expression is of the form

$$\ldots \wedge q(x) \wedge \ldots \wedge q(y) \wedge \ldots$$

where $q(x)$ costs \$20 and $q(y)$ costs \$10, then factoring assumes x and y to be identical and yields an expression of the form

$$\ldots \wedge q(x) \wedge \ldots$$

where $q(x)$ costs \$10. This feature leads to minimality through the exploitation of redundancy.

---

[8] Sometimes a cigar is just a cigar.

[9] The abduction scheme is due to Mark Stickel, and it, or a variant of it, is described at greater length in Stickel (1989).

[10] Stickel (1989) generalizes this to arbitrary functions of c.

Note that in (6), if $w_1 + w_2 < 1$, most-specific abduction is favored—why assume Q when it is cheaper to assume P_1 and P_2. If $w_1 + w_2 > 1$, least-specific abduction is favored—why assume P_1 and P_2 when it is cheaper to assume Q. But in

$$P_1^{.6} \wedge P_2^{.6} \supset Q$$

if P_1 has already been derived, it is cheaper to assume P_2 than Q. P_1 has provided evidence for Q, and assuming the "balance" P_2 of the necessary evidence for Q should be cheaper.

Factoring can also override least-specific abduction. Suppose we have the axioms

$$P_1^{.6} \wedge P_2^{.6} \supset Q_1$$
$$P_2^{.6} \wedge P_3^{.6} \supset Q_2$$

and we wish to derive $Q_1 \wedge Q_2$, where each conjunct has an assumability cost of \$10. Assuming $Q_1 \wedge Q_2$ will then cost \$20, whereas assuming $P_1 \wedge P_2 \wedge P_3$ will cost only \$18, since the two instances of P_2 can be unified. Thus, the abduction scheme allows us to adopt the careful policy of favoring least-specific abduction while also allowing us to exploit the redundancy of texts for more specific interpretations.

Finally, we should note that whenever an assumption is made, it first must be checked for consistency. Problems associated with this requirement are discussed in Section 8.1.

In the above examples we have used equal weights on the conjuncts in the antecedents. It is more reasonable, however, to assign the weights according to the "semantic contribution" each conjunct makes to the consequent. Consider, for example, the axiom

$$(\forall x) car(x)^{.8} \wedge no\text{-}top(x)^{.4} \supset convertible(x)$$

We have an intuitive sense that *car* contributes more to *convertible* than *no-top* does. We are more likely to assume something is a convertible if we know that it is a car than if we know it has no top.[11] The weights on the conjuncts in the antecedent are adjusted accordingly.

In the abductive approach to interpretation, we determine what implies the logical form of the sentence rather than determining what can be inferred from it. We backward-chain rather than forward-chain. Thus, one would think that we could not use superset information in processing the sentence. Since we are backward-chaining from the propositions in the logical form, the fact that, say, lube oil is a fluid, which would be expressed as

(7) $(\forall x) lube\text{-}oil(x) \supset fluid(x)$

could not play a role in the analysis of a sentence containing "lube oil". This is inconvenient. In the text

Flow obstructed. Metal particles in lube oil filter.

---

[11] To prime this intuition, imagine two doors. Behind one is a car. Behind the other is something with no top. You pick a door. If there's a convertible behind it, you get to keep it. Which door would you pick?

we know from the first sentence that there is a fluid. We would like to identify it with the lube oil mentioned in the second sentence. In interpreting the second sentence, we must prove the expression

$$(\exists\, x)\mathit{lube\text{-}oil}(x)$$

If we had as an axiom

$$(\forall\, x)\mathit{fluid}(x)\ \supset\ \mathit{lube\text{-}oil}(x)$$

then we could establish the identity. But of course we don't have such an axiom, for it isn't true. There are lots of other kinds of fluids. There would seem to be no way to use superset information in our scheme.

Fortunately, however, there is a way. We can make use of this information by converting the axiom to a biconditional. In general, axioms 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 as in the above example, we will not be able to prove the differentiae, and in many cases the differentiae cannot even be spelled out. But in our abductive scheme, this does not matter; they can simply be assumed. In fact, we need not state them explicitly. We can simply introduce a predicate which stands for all the remaining properties. It will never be provable, but it will be assumable. Thus, we can rewrite (7) as

$$(\forall\, x)\mathit{fluid}(x)^{.6} \wedge \mathit{etc}_1(x)^{.6} \equiv \mathit{lube\text{-}oil}(x)$$

Then the fact that something is fluid can be used as evidence for its being lube oil, since we can assume $etc_1(x)$. With the weights distributed according to semantic contribution, we can go to extremes and use an axiom like

$$(\forall\, x)\mathit{mammal}(x)^{.2} \wedge \mathit{etc}_2(x)^{.9} \supset \mathit{elephant}(x)$$

to allow us to use the fact that something is a mammal as (weak) evidence for its being an elephant.

The introduction of "et cetera" predications is a very powerful, and liberating, device. Before we hit upon this device, in our attempts at axiomatizing a domain in a way that would accommodate many texts, we were always "arrow hacking"—trying to figure out which way the implication had to go if we were to get the right interpretations, and lamenting when that made no semantic sense. With "et cetera" predications, that problem went away, and for principled reasons. Implicative relations could be used in either direction. Moreover, their use is liberating when constructing axioms for a knowledge base. It is well-known that almost no concept can be defined precisely. We are now able to come as close to a definition as we can and introduce an "et cetera" predication with an appropriate weight to indicate how far short we feel we have fallen. The "et cetera" predications play a role analogous to the abnormality predications of circumscriptive logic (McCarthy, 1987), a connection we explore a bit further in Section 8.3.

Exactly how the weights and costs should be assigned is a matter of continuing research. Our experience so far suggests that which interpretation is chosen is sensitive to whether the weights add up to more or less than one, but that otherwise the system's performance is fairly impervious to small changes in the values of the weights and costs. In Section 8.1, there some further discussion about the uses the numbers can be put to in making the abduction procedure more efficient, and in Section 8.3, there is a discussion of the semantics of the numbers.

5 Examples

5.1 Distinguishing the Given and the New

Let us examine four successively more difficult definite reference problems in which the given and the new information are intertwined and must be separated.[12] The first is

> Retained sample and filter element.

Here "sample" is new information. It was not known before this sentence in the message that a sample was taken. The "filter element", on the other hand, is given information. It is already known that the compressor's lube oil system has a filter, and that a filter has a filter element as one of its parts. These facts are represented in the knowledge base by the axioms

$$filter(F)$$

$$(\forall\, f) filter(f) \supset (\exists\, fe) filter\text{-}element(fe) \wedge part(fe, f)$$

Noun phrase conjunction is represented by the predicate $andn$. The expression $andn(x, s, fe)$ says that x is the typical element of the set consisting of the elements s and fe. Typical elements can be thought of as reified universally quantified variables. Roughly, their properties are inherited by the elements of the set. (See Hobbs, 1983b.) An axiom of pairs says that a set can be formed out of any two elements:

$$(\forall\, s, fe)(\exists\, x) andn(x, s, fe)$$

The logical form for the sentence is, roughly,

$$(\exists\, e, y, x, s, fe) retain'(e, y, x) \wedge andn(x, s, fe) \wedge sample(s) \wedge filter\text{-}element(fe)$$

That is, y retained x where x is the typical element of a set consisting of a sample s and a filter element fe. Let us suppose we have no metonymy problems here. Then interpretation is simply a matter of deriving this expression. We can prove the existence

---

[12] In all the examples of Section 5, we will ignore weights and costs, show the path to the correct interpretation, and assume the weights and costs are such that this interpretation will be chosen. A great deal of theoretical and empirical research will be required before this will happen in fact, especially in a system with a very large knowledge base.

of the filter element from the existence of the filter F. We cannot prove the existence of the sample s, so we assume it. It is thus new information. Given s and fe, the axiom of pairs gives us the existence of x and the truth of $andn(x, s, fe)$. We cannot prove the existence of the retaining e, so we assume it; it is likewise new information.

The next example is a bit trickier, because new and old information about the same entity are encoded in a single noun phrase.

> There was adequate lube oil.

We know about the lube oil already, and there is a corresponding axiom in the knowledge base.

$$lube\text{-}oil(O)$$

Its adequacy is new information, however. It is what the sentence is telling us.

The logical form of the sentence is, roughly,

$$(\exists\, o) lube\text{-}oil(o) \wedge adequate(o)$$

This is the expression that must be derived. The proof of the existence of the lube oil is immediate. It is thus old information. The adequacy cannot be proved and is hence assumed as new information.

The next example is from Clark (1975), and illustrates what happens when the given and new information are combined into a single lexical item:

> John walked into the room.
> The chandelier shone brightly.

What chandelier is being referred to?

Let us suppose we have in our knowledge base the fact that rooms have lights:

$$(\forall\, r) room(r) \supset (\exists\, l) light(l) \wedge in(l, r) \tag{8}$$

Suppose we also have the fact that lighting fixtures with several branches are chandeliers:

$$(\forall\, l) light(l) \wedge has\text{-}branches(l) \supset chandelier(l) \tag{9}$$

The first sentence has given us the existence of a room—$room(R)$. To solve the definite reference problem in the second sentence, we must prove the existence of a chandelier. Back-chaining on axiom (9), we see we need to prove the existence of a light with branches. Back-chaining from $light(l)$ in axiom (8), we see we need to prove the existence of a room. We have this in $room(R)$. To complete the derivation, we assume the light l has branches. The light is thus given by the room mentioned in the previous sentence, while the fact that it has several branches is new information.

This example may seem to have an unnatural, pseudo-literary quality. There are similar examples, however, which are completely natural. Consider

I saw my doctor last week.
He told me to get more exercise.

Who does "he" in the second sentence refer to?

Suppose in our knowledge base we have axioms encoding the fact that a doctor is a person,

$$(10) \quad (\forall\, d) doctor(d) \supset person(d)$$

and the fact that a male person is a "he",

$$(11) \quad (\forall\, d) person(d) \wedge male(d) \supset he(d)$$

To solve the reference problem, we must derive

$$(\exists\, d) he(d)$$

Back-chaining on axioms (11) and (10), matching with the doctor mentioned in the first sentence, and assuming the new information $male(d)$ gives us a derivation.[13]

5.2 Exploiting Redundancy

We next show the use of the abduction scheme in solving internal coreference problems. Two problems raised by the sentence

> The plain was reduced by erosion to its present level.

are determining what was eroding and determining what "it" refers to. Suppose our knowledge base consists of the following axioms:

$$(\forall p, l, s) decrease(p, l, s) \wedge vertical(s) \wedge etc_3(p, l, s) \equiv (\exists\, e) reduce'(e, p, l)^{14}$$

or e is a reduction of p to l if and only if p decreases to l on some (real or metaphorical) vertical scale s (plus some other conditions).

$$(\forall p) landform(p) \wedge flat(p) \wedge etc_4(p) \equiv plain(p)$$

or p is a plain if and only if p is a flat landform (plus some other conditions).

$$(\forall\, e, y, l, s) at'(e, y, l) \wedge on(l, s) \wedge vertical(s) \wedge flat(y) \wedge etc_5(e, y, l, s) \\ \equiv level'(e, l, y)$$

or e is the condition of l's being the level of y if and only if e is the condition of y's being at l on some vertical scale s and y is flat (plus some other conditions).

---

[13] Sexists will find this example more compelling if they substitute "she" for "he".

[14] This and the subsequent axioms are written as biconditionals, but they would be used as implications (from left to right), and the weighting scheme would operate accordingly.

$$(\forall\, x, l, s) decrease(x, l, s) \;\wedge\; landform(x) \;\wedge\; altitude(s) \;\wedge\; etc_6(y, l, s)$$
$$\equiv\; (\exists\, e) erode'(e, x)$$

or e is an eroding of x if and only if x is a landform that decreases to some point l on the altitude scale s (plus some other conditions).

$$(\forall\, s) vertical(s) \;\wedge\; etc_7(s) \;\equiv\; altitude(s)$$

or s is the altitude scale if and only if s is vertical (plus some other conditions).

Now the analysis. The logical form of the sentence is roughly

$$(\exists\, e_1, p, l, e_2, x, e_3, y) reduce'(e_1, p, l) \;\wedge\; plain(p) \;\wedge\; erode'(e_2, x) \;\wedge\; present(e_2)$$
$$\wedge\, level'(e_3, l, y)$$

Our characterization of interpretation says that we must derive this expression from the axioms or from assumptions. Back-chaining on $reduce'(e_1, p, l)$ yields

$$decrease(p, l, s_1) \;\wedge\; vertical(s_1) \;\wedge\; etc_3(p, l, s_1)$$

Back-chaining on $erode'(e_1, x)$ yields

$$decrease(x, l_2, s_2) \;\wedge\; landform(x) \;\wedge\; altitude(s_2) \;\wedge\; etc_6(x, l_2, s_2)$$

and back-chaining on $altitude(s_2)$ in turn yields

$$vertical(s_2) \;\wedge\; etc_7(s_2)$$

We unify the goals $decrease(p, l, s_1)$ and $decrease(x, l_2, s_2)$, and thereby identify the object x of the erosion with the plain p. The goals $vertical(s_1)$ and $vertical(s_2)$ also unify, telling us the reduction was on the altitude scale. Back-chaining on $plain(p)$ yields

$$landform(p) \;\wedge\; flat(p) \;\wedge\; etc_4(p)$$

and $landform(x)$ unifies with $landform(p)$, reinforcing our identification of the object of the erosion with the plain. Back-chaining on $level'(e_3, l, y)$ yields

$$at'(e_3, y, l) \;\wedge\; on(l, s_3) \;\wedge\; vertical(s_3) \;\wedge\; flat(y) \;\wedge\; etc_5(e_3, y, l, s_3)$$

and $vertical(s_3)$ and $vertical(s_2)$ unify, as do $flat(y)$ and $flat(p)$, thereby identifying "it", or y, as the plain p. We have not written out the axioms for this, but note also that "present" implies the existence of a change of level, or a change in the location of "it" on a vertical scale, and a decrease of a plain is a change of the plain's location on a vertical scale. Unifying these would provide reinforcement for our identification of "it" with the plain. Now assuming the most specific atomic formulas we have derived including all the "et cetera" conditions, we arrive at an interpretation that is minimal and that solves the internal coreference problems as a by-product.[15]

---

[15] This example was analyzed in a similar manner in Hobbs (1978) but not in such a clean fashion, since it was without benefit of the abduction scheme.

5.3 The Four Local Pragmatics Problems At Once

Let us now return to the example of Section 3.

> Disengaged compressor after lube-oil alarm.

Recall that we must resolve the reference of "compressor" and "alarm", discover the implicit relation between the lube oil and the alarm, attach "after alarm" to either the compressor or the disengaging, and expand "after alarm" into "after the sounding of the alarm".

The knowledge base includes the following axioms: There are a compressor C, an alarm A, lube oil O, and the pressure P of the lube oil O at A:

$$compressor(C),\ alarm(A),\ lube\text{-}oil(O),\ pressure(P, O, A)$$

The alarm is for the lube oil:

$$for(A, O)$$

The for relation is a possible nn relation:

$$(\forall\, a, o) for(a, o) \supset nn(o, a)$$

A disengaging e_1 by x of c is an event:

$$(\forall\, e_1, x, c) disengage'(e_1, x, c) \supset event(e_1)$$

If the pressure p of the lube oil o at the alarm a is not adequate, then there is a sounding e_2 of the alarm, and that sounding is the function of the alarm:

$$(\forall\, a, o, p) alarm(a) \wedge lube\text{-}oil(o) \wedge pressure(p, o, a) \wedge \neg adequate(p)$$
$$\supset (\exists\, e_2) sound'(e_2, a) \wedge function(e_2, a)$$

A sounding is an event:

$$(\forall\, e_2, a) sound'(e_2, a) \supset event(e_2)$$

An entity can be coerced into its function:

$$(\forall\, e_2, a) function(e_2, a) \supset rel(e_2, a)$$

Identity is a possible coercion:

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

Finally, we have axioms encoding set membership:

$$(\forall\, y, s) y \in \{y\} \cup s$$
$$(\forall\, y, x, s) y \in s \supset y \in \{x\} \cup s$$

Of the possible metonymy problems, let us confine ourselves to one posed by "after". Then the expression that needs to be derived for an interpretation is

$$\begin{aligned}&(\exists\, e_1, x, c, k_1, k_2, y, a, o)\mathit{disengage}'(e_1, x, c) \wedge \mathit{compressor}(c) \wedge \mathit{after}(k_1, k_2)\\ &\quad \wedge \mathit{event}(k_1) \wedge \mathit{rel}(k_1, y) \wedge y \in \{c, e_1\} \wedge \mathit{event}(k_2) \wedge \mathit{rel}(k_2, a)\\ &\quad \wedge \mathit{alarm}(a) \wedge \mathit{lube\text{-}oil}(o) \wedge \mathit{nn}(o, a)\end{aligned}$$

One way for $rel(k_1, y)$ to be true is for k_1 and y to be identical. We can back-chain from $event(k_1)$ to obtain $disengage'(k_1, x_1, c_1)$. This can be merged with $disengage'(e_1, x, c)$, yielding an interpretation in which the attachment y of the prepositional phrase is to "disengage". This identification of y with e_1 is consistent with the constraint $y \in \{c, e_1\}$. The conjunct $disengage'(e_1, x, c)$ cannot be proved and must be assumed as new information.

The conjuncts $compressor(c)$, $lube\text{-}oil(o)$, and $alarm(a)$ can be proved immediately, resolving c to C, o to O, and a to A. The compound nominal relation $nn(O, A)$ is true because $for(A, O)$ is true. One way for $event(k_2)$ to be true is for $sound'(k_2, a)$ to be true, and $function(k_2, A)$ is one way for $rel(k_2, A)$ to be true. Back-chaining on each of these and merging the results yields the goals $alarm(A)$, $lube\text{-}oil(o)$, $pressure(p, o, A)$, and $\neg adequate(p)$. The first three of these can be derived immediately, thus identifying o as O and p as P, and $\neg adequate(p)$ is assumed. We have thereby coerced the alarm into the sounding of the alarm, and as a by-product we have drawn the correct implicature, or assumed, that the lube oil pressure is inadequate.

5.4 Schema Recognition

One of the most common views of "understanding" in artificial intelligence has been that to understand a text is to match it with some pre-existing schema. In our view, this is far too limited a notion. But it is interesting to note that this sort of processing falls out of our abduction scheme, provided schemas are expressed as axioms in the right way.

Let us consider an example. RAINFORM messages are messages about sightings and pursuits of enemy submarines, generated during naval maneuvers. A typical message might read, in part,

> Visual sighting of periscope followed by attack with ASROC and torpedoes. Submarine went sinker.

An "ASROC" is an air-to-surface rocket, and to go sinker is to submerge. These messages generally follow a single, rather simple schema. An enemy sub is sighted by one of our ships. The sub either evades our ship or is attacked. If it is attacked, it is either damaged or destroyed, or it escapes.

A somewhat simplified version of this schema can be encoded in an axiom as follows:

$$\begin{aligned}&(\forall\, e_1, e_2, e_3, x, y, \ldots)\mathit{sub\text{-}sighting\text{-}schema}(e_1, e_2, e_3, x, y, \ldots)\\ &\quad \supset \mathit{sight}'(e_1, x, y) \wedge \mathit{friendly}(x) \wedge \mathit{ship}(x) \wedge \mathit{enemy}(y) \wedge \mathit{sub}(y)\\ &\qquad \wedge \mathit{then}(e_1, e_2) \wedge \mathit{attack}'(e_2, x, y) \wedge \mathit{outcome}(e_3, e_2, x, y)\end{aligned}$$

That is, if we are in a submarine-sighting situation, with all of its associated roles e_1, x, y, and so on, then a number of things are true. There is a sighting e_1 by a friendly ship x of an enemy sub y. Then there is an attack e_2 by x on y, with some outcome e_3.

Among the possible outcomes is y's escaping from x, which we can express as follows:

$$(\forall\, e_3, e_2, x, y) outcome(e_3, e_2, x, y) \;\wedge\; etc_1(e_3) \;\equiv\; escape'(e_3, y, x)$$

We express it this way because we will have to backward-chain from the escape to the outcome, and on to the schema.

The other facts that need to be encoded are as follows:

$$(\forall\, y) sub(y) \;\supset\; (\exists\, z) periscope(z) \;\wedge\; part(z, y)$$

That is, a sub has a periscope as one of its parts.

$$(\forall\, e_1, e_2) then(e_1, e_2) \;\supset\; follow(e_2, e_1)$$

That is, if e_1 and e_2 occur in temporal succession (*then*), then e_2 follows e_1.

$$(\forall\, e_3, y, x) escape'(e_3, y, x) \;\wedge\; etc_2(e_3, x, y) \;\equiv\; submerge'(e_3, y)$$

That is, submerging is one way of escaping.

$$(\forall\, e_3, y) submerge'(e_3, y) \;\equiv\; go\text{-}sinker'(e_3, y)$$

That is, going sinker and submerging are equivalent.

In order to interpret the first sentence of the example, we must prove its logical form, which is, roughly,

$$\begin{aligned}(\exists\, e_1, x, z, e_2, u, v, a, t) & sight'(e_1, x, z) \;\wedge\; visual(e_1) \;\wedge\; periscope(z)\\ & \wedge follow(e_2, e_1) \;\wedge\; attack'(e_2, u, v) \;\wedge\; with(e_2, a)\\ & \wedge ASROC(a) \;\wedge\; with(e_2, t) \;\wedge\; torpedo(t)\end{aligned}$$

and the logical form for the second sentence, roughly, is the following:

$$(\exists\, e_3, y_1) go\text{-}sinker'(e_3, y_1) \;\wedge\; sub(y_1)$$

When we backward-chain from the logical forms using the given axioms, we end up, most of the time, with different instances of the schema predication

$$sub\text{-}sighting\text{-}schema(e_1, e_2, e_3, x, y, \ldots)$$

as goal expressions. Since our abductive inference method merges unifiable goal expressions, all of these are unified, and this single instance is assumed. Since it is almost the only expression that had to be assumed, we have a very economical interpretation for the entire text.

To summarize, when a large chunk of organized knowledge comes to be known, it can be encoded in a single axiom whose antecedent is a "schema predicate" applied to all of the role fillers in the schema. When a text describes a situation containing many of the entities and properties that occur in the consequent of the schema axiom, then very often the most economical interpretation of the text will be achieved by assuming the schema predicate, appropriately instantiated. If we were to break up the schema axiom into a number of axioms, each expressing different stereotypical features of the situation and each having in its antecedent the conjunction of a schema predication and an et cetera predication, default values for role fillers could be inferred where and only where they were appropriate and consistent.

When we do schema recognition in this way, there is no problem, as there is in other approaches, with merging several schemas. It is just a matter of assuming more than one schema predication with the right instantiations of the variables.

6 A Thorough Integration of Syntax, Semantics, and Pragmatics

6.1 The Integration

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.[16]

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

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

$$(\forall\, i,j,k) np(i,j) \wedge verb(j,k) \supset s(i,k)$$
$$(\forall\, i,j,k,l) det(i,j) \wedge noun(j,k) \wedge noun(k,l) \supset np(i,l)$$

That is, suppose the indices i, j, k, and l stand for the "interword points", from 0 to the number of words in the sentence. If there is a noun phrase from point i to point j and a verb from point j to point k, then there is a sentence from point i to point k, and similarly for the second rule. To parse a sentence is to prove $s(0,N)$, where N is the number of words in the sentence.

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:

(12) $(\forall\, i,j,k,y,p,e,x) np(i,j,y) \wedge verb(j,k,p) \wedge p'(e,x) \wedge rel(x,y) \wedge Req(p,x)$
$\supset s(i,k,e)$

(13) $(\forall\, i,j,k,l,w_1,w_2,y,z) det(i,j,the) \wedge noun(j,k,w_1) \wedge noun(k,l,w_2)$
$\wedge w_1(z) \wedge w_2(y) \wedge nn(z,y) \supset np(i,l,y)$

The third arguments of the "lexical" predicates *noun*, *verb*, and *det* are the words themselves (or the predicates of the same name), such as *Boston*, *office* or *call*. The atomic formula $np(i,j,y)$ means that there is a noun phrase from point i to point j 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

---

[16] 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 (12) can then be paraphrased as follows: "If there is a noun phrase from point i to point j referring to y, and the verb p (denoting the predicate p) from point j to point k, 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 there is a sentence from point i to point k describing eventuality e." Axiom (13) can be paraphrased as follows: "If there is the determiner *the* from point i to point j, and the noun w_1 occurs from point j to point k, and the noun w_2 occurs from point k to point l, and the predicate w_1 is true of some entity z, and the predicate w_2 is true of some entity y, and there is some implicit relation nn between z and y, then there is a noun phrase from point i to point l referring to the entity y. Note that the conjuncts from line (3a) in the logical form have been incorporated into axiom (12) and the conjuncts from line (3b) into axiom (13).[17]

Before when we proved $s(0, N)$, we proved there was a sentence from point 0 to point N. Now, if we prove $(\exists e)s(0, N, e)$, we prove there is an *interpretable* sentence from point 0 to point N and that the eventuality e is its interpretation.

Each axiom in the "grammar" then has a "syntactic" part—the conjuncts like $np(i, j, y)$ and $verb(j, k, p)$—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 $(\exists e)s(0, N, e)$, 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 pragmatics part.

One final modification is necessary, since the elements of the pragmatics part have to be assumable. If we wish to get the same costs on the conjuncts in the logical form that we proposed at the end of Section 3, we need to augment our formalism to allow attaching assumability costs directly to some of the conjuncts in the antecedents of Horn clauses. Continuing to use the arbitrary costs we have used before, we would thus rewrite the axioms as follows:

$$(14) \quad (\forall i, j, k, y, p, e, x)np(i, j, y) \wedge verb(j, k, p) \wedge p'(e, x)^{\$3} \wedge rel(x, y)^{\$20} \\ \wedge Req(p, x)^{\$10} \supset s(i, k, e)$$

$$(15) \quad (\forall i, j, k, l, w_1, w_2, y, z)det(i, j, the) \wedge noun(j, k, w_1) \wedge noun(k, l, w_2) \\ \wedge w_1(z)^{\$5} \wedge w_2(y)^{\$10} \wedge nn(z, y)^{\$20} \supset np(i, l, y)$$

The first axiom now says what it did before, but in addition we can assume $p'(e, x)$ for a cost of \$3, $rel(x, y)$ for a cost of \$20, and $Req(p, x)$ for a cost of \$10.[18]

---

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

[18] The costs, rather than weights, on the conjuncts in the antecedents are already permitted if we allow, as Stickel (1989) does, arbitrary functions rather than multiplicative weights.

Implementations of different orders of interpretation, or different sorts of interaction among syntax, compositional semantics, and local pragmatics, can then be seen as different orders of search for a proof of $(\exists\, e)s(0, N, e)$. In a syntax-first order of interpretation, one would try first to prove all the "syntactic" atomic formulas, such as $np(i, j, y)$, before any of the "local pragmatics" atomic formulas, such as $p'(e, x)$. Verb-driven interpretation would first try to prove $verb(j, k, p)$ and would then use the information in the requirements associated with the verb to drive the search for the arguments of the verb, by deriving $Req(p', x)$ before back-chaining on $np(i, j, y)$. But more fluid orders of interpretation are obviously possible. This formulation allows one to prove those things first which are easiest to prove, and therefore allows one to exploit the fact that the strongest clues to the meaning of a sentence can come from a variety of sources—its syntax, the semantics of its main verb, the reference of its noun phrases, and so on. It is also easy to see how processing could occur in parallel, insofar as parallel Prolog is possible.

6.2 Syntactically Ill-Formed Utterances

It is straightforward to extend this approach to deal with ill-formed or unclear utterances, by first giving the expression to be proved $(\exists\, e)s(0, N, e)$ an assumability cost and then adding weights to the syntactic part of the axioms. Thus, axiom (14) can be revised as follows:

$$(\forall\, i, j, k, y, p, e, x) np(i, j, y)^{.6} \wedge verb(j, k, p) \wedge p'(e, x)^{\$3} \wedge rel(x, y)^{\$20} \wedge Req(p, x)^{\$10} \\ \supset\ s(i, k, e)$$

This says that if you find a verb, then for a small cost you can go ahead and assume there is a noun phrase, allowing us to interpret utterances without subjects, which are very common in certain kinds of informal discourse, including equipment failure reports and naval operation reports. In this case, the variable y will have no identifying properties other than what the verb phrase gives it.

More radically, we can revise the axiom to

$$(\forall\, i, j, k, y, p, e, x) np(i, j, y)^{.4} \wedge verb(j, k, p)^{.8} \wedge p'(e, x)^{\$3} \wedge rel(x, y)^{\$20} \wedge Req(p, x)^{\$10} \\ \supset\ s(i, k, e)$$

This allows us to assume there is a verb as well, although for a higher cost than for assuming a noun phrase (since presumably a verb phrase provides more evidence for the existence of a sentence than a noun phrase does). That is, either the noun phrase or the verb can constitute a sentence if the string of words is otherwise interpretable. In particular, this allows us to handle cases of ellipsis, where the subject is given but the verb is understood. In these cases we will not be able to prove $Req(p, x)$ unless we first identify p by proving $p'(e, x)$. The solution to this problem is likely to come from salience in context or from considerations of discourse coherence, such as recognizing a parallel with a previous segment of the discourse.

Similarly, axiom (15) can be rewritten to

$$(\forall\, i, j, k, l, w_1, w_2, y, z) det(i, j, the)^{.2} \wedge noun(j, k, w_1) \wedge noun(k, l, w_2) \wedge w_1(z)^{\$5} \\ \wedge w_2(y)^{\$10} \wedge nn(z, y)^{\$20} \supset\ np(i, l, y)$$

to allow omission of determiners, as is also very common in some kinds of informal discourse.

6.3 Recognizing the Coherence Structure of Discourse

In Hobbs (1985d) a theory of discourse structure is outlined in which coherence relations such as 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.

It looks as if it would be relatively straightforward to extend our method of interpretation as abduction to the recognition of some aspects of this coherence structure of the discourse. The hierarchical structure can be captured by the axiom

$$(\forall\, i,j,e) s(i,j,e) \supset \mathit{Segment}(i,j,e)$$

specifying that a sentence is a discourse segment, and axioms of the form

$$(\forall\, i,j,k,e_1,e_2,e) \mathit{Segment}(i,j,e_1) \wedge \mathit{Segment}(j,k,e_2) \wedge \mathit{CoherenceRel}(e_1,e_2,e) \\ \supset \mathit{Segment}(i,k,e)$$

saying that if there is a segment from i to j whose assertion or topic is e_1, and a segment from j to k asserting e_2, and *CoherenceRel* is one of the coherence relations where e is the assertion or topic of the composed segment as determined by the definition of the coherence relation, then there is a segment from i to k asserting e.

A first approximation of the definition for "explanation", for example, would be the following:

$$(\forall\, e_1,e_2) \mathit{cause}(e_2,e_1) \supset \mathit{Explanation}(e_1,e_2,e_1)$$

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, and the assertion of the composed segment is the assertion of the first segment.

The expansion relations, such as "elaboration", "parallel", and "contrast", are more difficult to capture in this way, since they require second-order formulations. For example, the parallel relation might be encoded in an axiom schema as follows:

$$(\forall\, e_1,e_2,x,y) p'(e_1,x) \wedge p'(e_2,y) \wedge q(x) \wedge q(y) \supset \mathit{Parallel}(e_1,e_2,e_1\&e_2)$$

That is, the two segments assert that two entities x and y, which are similar by virtue of both having property q, have some property p. The assertion of the composed segment is the conjunction of the assertions of the constituent segments.[19]

To interpret an N-word text, one must then prove the expression

$$(\exists\, e) \mathit{Segment}(0,N,e)$$

---

[19] See Hobbs (1985b) for explication of the notation $e_1\&e_2$.

The details of this approach remain to be worked out.

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.

We should point out a subtle shift of perspective we have just gone through. In Sections 3, 4, and 5 of this paper, the problem of interpretation was viewed as follows: One is given certain observable facts, namely, the logical form of the sentence, and one has to find a proof that demonstrates why they are true. In this section, we no longer set out to prove the observable facts. Rather we set out to prove that we are viewing a coherent situation, and it is built into the rules that specify what situations are coherent that an explanation must be found for the observable facts. We return to this point in the conclusion.

6.4 Below the Level of the Word

Interpretation can be viewed as abduction below the level of the word as well. Let us consider written text first. Prolog-style rules can decompose words into their constituent letters. The rule that says the word "it" occurs between point i and point k would be

$$(\forall i, j, k) I(i, j) \wedge T(j, k) \supset pro(i, k, it)$$

For most applications, this is not, of course, an efficient way to proceed. However, if we extend the approach to ill-formed or unclear input described above to the spellings of words, we have a way of recognizing and correcting spelling errors where the misspelling is itself an English word. Thus, in

> If is hard to recognize speech.

we are able to use constraints of syntax and pragmatics to see that we would have a good interpretation if "it" were the first word in the sentence. The letter "i" occurring as the first word's first letter provides supporting evidence that that is what we have. Thus, to get the best interpretation, we simply assume the second letter is "t" and not "f".

It is also likely that this approach could be extended to speech recognition by using Prolog-style rules to decompose morphemes into their phonemes, or into phonetic features, or into whatever else an acoustic processor can produce, and weighting these elements according to their acoustic prominence.

Suppose, for example, that the acoustic processor produces a word lattice, that is, a list of items saying that there is a certain probability that a certain word occurs between two points in the input stream. These can be expressed as atomic formulas of the form $word(i, j)$ with associated assumability costs corresponding to their probabilities. Thus, for the sentence

> It is hard to recognize speech.

we might have the atomic formulas

$$recognize(i_1, i_4), wreck(i_1, i_2), a(i_2, i_3), nice(i_3, i_5), speech(i_4, i_6), beach(i_5, i_6),$$

each with associated assumability costs.

If the accoustic processor produces trigrams indicating the probabilities that portions of the input stream convey certain phonemes flanked by certain other phonemes, the compositions of words can be similarly expressed by axioms.

$$(\forall\, i_1, i_2, i_3, i_4, i_5)\, {}_{\#}s^{p}(i_1, i_2) \wedge {}_{s}p^{i}(i_2, i_3) \wedge {}_{p}i^{\check{c}}(i_3, i_4) \wedge {}_{i}\check{c}^{\#}(i_4, i_5) \supset speech(i_1, i_5)$$

The acoustic component would then assert propositions such as ${}_{s}p^{i}(i_2, i_3)$, with an assumability cost corresponding to the goodness of fit of the input with the pre-stored pattern for that trigram.

Finally, if the acoustic processor recognized distinctive features of the phonemes, axioms could also express the composition of these features into phonemes:

$$(\forall\, i_1, i_2)[-Voiced](i_1, i_2) \wedge [+Stop](i_1, i_2) \wedge [+Bilabial](i_1, i_2) \supset P(i_1, i_2)$$

Again, assumability costs would be lower for the features that were detected with more reliability.

With any of these interfaces with accoustic processors, the approach described above for handling ill-formed and unclear input would allow us to assume our way past elements of the accoustic stream that were not sufficiently clear to resolve, in whatever way accords best with syntactic and pragmatic interpretation. Thus, in the last example, if we could not prove $[-Voiced](i_1, i_2)$ and if assuming it led to the best interpretation syntactically and pragmatically, then we could, at an appropriate cost, go ahead and assume it.

None of this should be viewed as a suggestion that the most efficient technique for recognizing speech is unconstrained abductive theorem-proving. It is rather a framework that allows us to see all of the processes, from phonology to discourse pragmatics, as examples of the same sort of processing. Abduction gives us a unified view of language understanding. Where efficient, special-purpose techniques exist for handling one aspect of the problem, these can be viewed as special-purpose procedures for proving certain of the propositions.

6.5 Generation as Abduction

A commonly cited appeal for declarative formalisms for grammars is that they can be used bidirectionally, for either parsing or generation. Having thoroughly integrated parsing and pragmatic interpretation in a declarative formalism, we can now use the formalism for generation as well as interpretation. In interpretation, we know that there is some sentence with N words, and our task is to discover the eventuality e that it is describing. That is, we must prove

$$(\exists\, e) s(0, N, e)$$

In generation, the problem is just the opposite. We know some eventuality E that we want to describe, and our task is to prove the existence of a sentence of some length n which expresses it. That is, we must prove

$$(\exists\, n) s(0, n, E)$$

In interpretation, what we have to assume is the new information. In generation, we have to assume the terminal categories of the grammar. That is, we have to assume the occurrence of the words in particular positions. We stipulate that when these assumptions are made, the words are spoken.[20]

Let us look again at the simple grammar of Section 6.1, this time from the point of view of generation. A little arithmetic is introduced to avoid axioms that say a word is one word long.

$$(12') \quad (\forall\, i,k,y,p,e,x) np(i,k-1,y) \wedge verb(k-1,k,p) \wedge p'(e,x) \wedge rel(x,y) \\ \wedge Req(p,x) \supset s(i,k,e)$$

$$(13') \quad (\forall\, i,w_1,w_2,y,z) det(i,i+1,the) \wedge noun(i+1,i+2,w_1) \\ \wedge noun(i+2,i+3,w_2) \wedge w_1(z) \wedge w_2(y) \wedge nn(z,y) \supset np(i,i+3,y)$$

We will also be referring to the world knowledge axioms of Section 1. Suppose we want to assert the existence of an eventuality E which is a calling event by John who works for the office in Boston. We need to prove there is a sentence that realizes it. A plausible story about how this could be done is as follows. The way to prove $s(0,n,E)$ is to prove each of the conjuncts in the antecedent of axiom (12′). Working from what we know, namely E, we try to instantiate $p'(E,x)$ and we find $call'(E,J_1)$. Now that we know *call* and J_1 we try to prove $Req(call,J_1)$, and do so by finding $person(J_1)$. We next try to prove $rel(J_1,y)$. At this point we could choose the coercion relation to be identity, in which case there would be no metonymy. Let us instead pick $work\text{-}for(J_1,O_1)$. Now that we have instantiated y as O_1, we use axiom (13′) to prove $np(0,k-1,O_1)$. Since $det(0,1,the)$ is a terminal category, we can assume it, which means that we utter the word "the". We next need to find a way of describing O_1 by proving the expression

$$w_1(z) \wedge w_2(O_1) \wedge nn(z,O_1)$$

We can do this by instantiating w_2 to *office*, by finding $in(O_1,B_1)$, and then by proving $w_1(B_1)$ by instantiating w_1 to the predicate $Boston$. We now have the terminal category $noun(1,2,Boston)$, which we assume, thus uttering "Boston". We also have the terminal category $noun(2,3,office)$, which we assume, thus uttering "office". Finally, we return to axiom (12′) where we complete the proof, and thus the sentence, by assuming $verb(3,4,call)$, thereby saying the word "call". As usual in pedagogical examples, we ignore tense.

The (admittedly naive) algorithm used here for searching for a proof, and thus for a sentence, is to try to prove next those goal atomic formulas that are partially instantiated and thus have the smallest branch factor for backward-chaining. Left-to-right generation is enforced by initially having only 0 as an instantiated interword point.

---

[20] This combines Shieber's idea of merging interpretation as abduction and parsing as deduction with another idea of Shieber's (Shieber, 1988) on the relation of parsing and generation in declarative representations of the grammar.

There are at least two important facets of generation that have been left out of this story. First of all, we choose a description of an entity in a way that will enable our hearer to identify it. That is, we need to find properties $w_2(O_1)$, and so on, that are mutually known and that describe the entity uniquely among all the entities in focus. A more complex story can be told that incorporates this facet. Second, utterances are actions in larger plans that the speaker is executing to achieve some set of goals. But planning itself can be viewed as a theorem-proving process, and thus the atomic formula $s(0, n, E)$ can be viewed as a subgoal in this plan. This view of generation as abduction fits nicely with the view of generation as planning.

Some will find this unified view of interpretation and generation psychologically implausible. It is a universal experience that we are able to interpret more utterances than we typically, or ever, generate. Does this not mean that the grammars we use for interpretation and generation are different? We think it is not necessary to tell the story like this, for several reasons. The search order for interpretation and generation will necessarily be very different, and it could be that paths that are never taken in generation are nevertheless available for interpretation. We can imagine a philosopher, for example, who is deathly afraid of category errors and never uses metonymy. In proving $rel(z, x)$ in axiom $(12')$ during generation, he always uses identity. But he may still have other ways of proving it during interpretation, that he uses when he finds it necessary to talk to non-philosophers. Furthermore, there is enough redundancy in natural language discourse that in interpretation, even where one lacks the necessary axioms, one is usually able, by making appropriate assumptions, to make sense out of an utterance one would not have generated.

It is worth pointing out that translation from one language to another can be viewed elegantly in this framework. Let s in our grammar above be renamed to s_E for English, and suppose we have a grammar for Japanese similarly incorporating semantics and local pragmatics, whose "root predicate" is s_J. Then the problem of translating from English to Japanese can be viewed as the problem of proving for a sentence of length N the expression

$$(\exists\, e, n) s_E(0, N, e) \ \wedge \ s_J(0, n, e)$$

That is, there is some eventuality e described by the given English sentence of N words and which can be expressed in Japanese by a sentence of some length n. In the simplest cases, lexical transfer would occur by means of axioms such as

$$(\forall\, x) mountain(x) \ \equiv \ yama(x)$$

Because of the expressive power of first-order logic, much more complicated examples of lexical transfer could be stated axiomatically as well. Some of the details of an abductive approach to translation are explored by Hobbs and Kameyama (1990).

6.6 The Role of Assumptions

We have used assumptions for many purposes: to accept new information from the speaker, to accommodate the speaker when he seems to assume something is mutually known when it is not, to glide over uncertainties and imperfections in the speech stream, and to utter words, or more generally, to take actions. Is there anything that all of these uses have in common? We think there is. In all the cases, there is a proposition that is not mutually known, and we somehow have to treat it as if it were mutually known. In interpreting an utterance and accepting it as true, we do this by entering the assumption into our mutual knowledge. In parsing the speech stream, we accommodate the speaker by assuming, or pretending if necessary, that the most appropriate token did occur in copresence with the speaker and is thus mutual knowledge. In generation, we *make* the assumption true in copresence with the hearer, and thus make it mutually known, by uttering the word or by taking the action.

6.7 Integration versus Modularity

For the past several decades, there has been quite a bit of discussion in linguistics, psycholinguistics, and related fields about the various modules involved in language processing and their interactions. A number of researchers have, in particular, been concerned to show that there was a syntactic module that operated in some sense independently of processes that accessed general world knowledge. Fodor (1983) has been perhaps the most vocal advocate of this position. He argues that human syntactic processing takes place in a special "informationally encapsulated" input module, immune from top-down influences from "central processes" involving background knowledge. This position has been contentious in psycholinguistics. Marslen-Wilson and Tyler (1987), for example, present evidence that if there is any information encapsulation, it is not in a module that has logical form as its output, but rather one that has a mental model or some other form of discourse representation as its output. Such output requires background knowledge in its construction. At the very least, if linguistic processing is modular, it is not immune from top-down context dependence.

Finally, however, Marslen-Wilson and Tyler argue that the principal question about modularity—"What interaction occurs between modules?"—is ill-posed. They suggest that there may be no neat division of the linguistic labor into modules, and that it therefore does not make sense to talk about interaction between modules. This view is very much in accord with the integrated approach we have presented here. Knowledge of syntax is just one kind of knowledge of the world. All is given a uniform representation. Any rule used in discourse interpretation can in principle, and often in fact will, involve predications about syntactic phenomena, background knowledge, the discourse situation, or anything else. In such an approach, issues of modularity simply go away.

In one extended defense of modularity, Fodor (n.d.) begins by admitting that the arguments against modularity are powerful. "If you're a modularity theorist, the fundamental problem in psycholinguistics is to talk your way out of the massive effects of context on language comprehension" (p. 15). He proceeds with a valiant attempt to do just that. He begins with an assumption: "Since a structural description is really the union of representations of an utterance in a variety of different theoretical vocabularies, it's natural to assume that the internal structure of the parsers is correspondingly functionally differentiated" (p. 10). But in our framework, this assumption is incorrect. Facts about syntax and pragmatics are expressed in different theoretical vocabularies only in the sense that facts about doors and airplanes are expressed in different theoretical vocabularies—different predicates are used. But the "internal structure of the parsers" is the same. It is all abduction.

In discussing certain sentences in which readers are "garden-pathed" by applying the syntactic strategy of "minimal attachment", Fodor proposes two alternatives, the first interactionist and the second modular: "Does context bias by penetrating the parser and *suspending* the (putative) preference for minimal attachment? Or does it bias by correcting the *output* of the parser when minimal attachment yields implausible analyses?" (p. 37) In our view, neither of these is true. The problem is to find the interpretation of the utterance that best satisfies a set of syntactic, semantic, and pragmatic constraints. Thus, all the constraints are applied simultaneously and the best interpretation satisfying them all is selected.

Moreover, often the utterance is elliptical, obscure, ill-formed, or unclear in parts. In these cases, various interpretive moves are available to the hearer, among them the local pragmatics moves of assuming metonymy or metaphor, the lexical move of assuming a very low-salience sense of a word, and the syntactic move of inserting a word to repair the syntax. The last of these is required in a sentence in a rough draft that was circulated of Fodor's paper:

> By contrast, on the Interactive model, it's assumed that the same processes have access to linguistic information can also access cognitive background. (p. 57–8)

The best way to interpret this sentence is to assume that a "that" should occur between "processes" and "have". There is no way of knowing *a priori* what interpretive moves will yield the best interpretation for a given utterance. This fact would dictate that syntactic analysis be completed even where purely pragmatic processes could repair the utterance to interpretability.

In Bever's classic example (Bever, 1970),

> The horse raced past the barn fell.

there are at least two possible interpretive moves: insert an "and" between "barn" and "fell", or assume the rather low-frequency, causative sense of "race". People generally make the first of these moves. However, Fodor himself gives examples, such as

> The performer sent the flowers was very pleased.

in which no such low-frequency sense needs to be accessed and the sentence is more easily interpreted as grammatical.

Our approach to this problem is in the spirit of Crain and Steedman (1985), who argue that interpretation is a matter of minimizing the number of presuppositions it is necessary to assume are in effect. Such assumptions add to the cost of the interpretation.

There remains, of course, the question of the optimal order of search for a proof for any particular input text. As pointed out in Section 6.1, the various proposals of modularizations can be viewed as suggestions for order of search. But in our framework, there is no particular reason to assume a rigid order of search. It allows what seems to us the most plausible account—that sometimes syntax drives interpretation and sometimes pragmatics does.

It should be pointed out that if Fodor were to adopt our position, it would only be with the utmost pessimism. According to him, we would have taken a peripheral, modular process that is, for just that reason, perhaps amenable to investigation, and turned it into one of the central processes, the understanding of which, on his view, would be completely intractable. However, it seems to us that nothing can be lost in this move. Insofar as syntax is tractable and the syntactic processing can be traced out, this information can be treated as information about efficient search orders in the central processes.

Finally, the reader may object to this integration because syntax and the other so-called modules constitute coherent domains of inquiry, and breaking down the barriers between them can only result in conceptual confusion. This is not a necessary consequence, however. One can still distinguish, if one wants, between linguistic axioms such as (12) and background knowledge axioms such as (8). It is just that they will both be expressed in the same formal language and used in the same fashion. What the integration has done is to remove such distinctions from the code and put them into the comments.

7 Relation to Other Work

7.1 Previous and Current Research on Abduction

Prior to the late seventeenth century science was viewed as deductive, at least in the ideal. It was felt that, on the model of Euclidean geometry, one should begin with propositions that were self-evident and deduce whatever consequences one could from them. The modern view of scientific theories, probably best expressed by Lakatos (1970), is quite different. One tries to construct abstract theories from which observable events can be deduced or predicted. There is no need for the abstract theories to be self-evident, and they usually are not. It is only necessary for them to predict as broad a range as possible of the observable data and for them to be "elegant", whatever that means. Thus, the modern view is that science is fundamentally abductive. We seek hidden principles or causes from which we can deduce the observable evidence.

This view of science, and hence the notion of abduction, can be seen first in some passages in Newton's *Principia* (1934 [1686]). It is understandable why Newton might have been driven to the modern view of scientific theories, as the fundamental principles of his system were in no way self-evident. In his "Preface to the First Edition" (p. xvii) he says, "The whole burden of philosophy seems to consist in this—from the phenomena of motions to investigate the forces of nature, and from these forces to demonstrate the other phenomena." The phenomena of motions and other phenomena correspond to the Q of our schema and the forces of nature correspond to our P and $P \supset Q$. At the beginning of Book III, before presenting the Universal Law of Gravitation, he argues for a parsimony of causes in his first "rule of reasoning in philosophy" (p. 308): "We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances." This seems to presuppose a view of scientific theorizing as abduction;

where he says "admit", we would say "assume"; his causes are our P and $P \supset Q$, and his appearances are our Q. At the end of *Principia* (p. 547), in a justification for not seeking the cause of gravity, he says, "And to us it is enough that gravity does really exist, and act according to the laws which we have explained, and abundantly serves to account for all the motions of the celestial bodies, and of our sea." The justification for gravity and its laws is not in its self-evidential nature but in what it accounts for.

The term "abduction" was first used by C. S. Pierce (e.g., 1955), who also called the process "retroduction". His definition of it is as follows:

> The surprising fact, C, is observed;
> But if A were true, C would be a matter of course,
> Hence, there is reason to suspect that A is true. (p. 151)

Pierce's C is what we have been calling $q(A)$ and A is what we have been calling $p(A)$. To say "if A were true, C would be a matter of course" is to say that for all x, $p(x)$ implies $q(x)$, that is, $(\forall x)p(x) \supset q(x)$. He goes on to describe what he refers to as "abductory induction". In our terms, this is when, after abductively hypothesizing $p(A)$, one checks a number of, or a random selection of, properties q_i such that $(\forall x)p(x) \supset q_i(x)$, to see whether $q_i(A)$ holds. This, in a way, corresponds to our check for consistency. Then Pierce says that "in pure abduction, it can never be justifiable to accept the hypothesis otherwise than as an interrogation", and that "the whole question of what one out of a number of possible hypotheses ought to be entertained becomes purely a question of economy." This corresponds to our evaluation scheme.

The first use of abduction in artificial intelligence was by Pople (1973), in the context of medical diagnosis. He gave the formulation of abduction that we have used and showed how it can be implemented in a theorem-proving framework. Literals that are "abandoned by deduction in the sense that they fail to have successor nodes" (p. 150) are taken as the candidate hypotheses. Those hypotheses are best that account for the most data, and in service of this principle, he introduced factoring or synthesis, which, just as in our scheme, attempts to unify goal literals. Hypotheses where this is used are favored. No further scoring criteria are given, however.

Work on abduction in artificial intelligence was revived in the early 1980s at several sites. Reggia and his colleagues (e.g., Reggia et al., 1983; Reggia, 1985) formulated abductive inference in terms of parsimonious covering theory. One is given a set of disorders (our $p(A)$'s) and a set of manifestations (our $q(A)$'s) and a set of causal relations between disorders and manifestations (our rules of the form $(\forall x)p(x) \supset q(x)$). An explanation for any set of manifestations is a set of disorders which together can cause all of the manifestations. The minimal explanation is the best one, where minimality can be defined in terms of cardinality or irredundancy. More recently, Peng and Reggia (1987a, 1987b) have begun to incorporate probabilistic considerations into their notion of minimality. For Reggia, the sets of disorders and manifestations are distinct, as is appropriate for medical diagnosis, and there is no backward-chaining to deeper causes; our abduction method is more general than his in that we can assume any proposition—one of the manifestations or an underlying cause of arbitrary depth.

In their textbook, Charniak and McDermott (1985) presented the basic pattern of abduction and then discuss many of the issues involved in trying to decide among alternative hypotheses on probabilistic grounds. Reasoning in uncertainty and its application to expert systems are presented as examples of abduction.

Cox and Pietrzykowski (1986) present a formulation in a theorem-proving framework that is very similar to Pople's, though apparently independent. It is especially valuable in that it considers abduction abstractly, as a mechanism with a variety of possible applications, and not just as a handmaiden to diagnosis. The test used to select a suitable hypothesis is that it should be what they call a "dead end"; that is, it should not be possible to find a stronger consistent assumption by backward-chaining from the hypothesis using the axioms in the knowledge base. However, this method is subject to a criticism theoretically. By insisting on the logically strongest hypothesis available, the dead-end test forces the abductive reasoning system to overcommit—to produce overly specific hypotheses. Often it does not seem reasonable, intuitively, to accept *any* of a set of very specific assumptions as the explanation of the fact that generated them by backward-chaining in the knowledge base. Moreover, the location of these dead ends is often a rather superficial and incidental feature of the knowledge base that has been constructed. Backward-chaining is a reasonable way to establish that the abductive hypothesis, in conjunction with the knowledge base, will logically imply the fact to be explained. But this is equally true whether or not a dead end has been reached. More backward-chaining is not necessarily better. Other tests must be sought to distinguish among the hypotheses reached by backward-chaining. It is in part to overcome such objections that we devised our *weighted* abduction scheme.

In recent years there has been an explosion of interest in abduction in artificial intelligence. A good overview of this research can be obtained from O'Rorke (1990).

In most of the applications of abduction to diagnosis, it is assumed that the relations expressed by the rules are all causal, and in fact Josephson (1990) has argued that that is necessarily the case in explanation. It seems to us that when one is diagnosing physical devices, of course explanations must be in terms of physical causality. But when we are working within an informational system, such as language or mathematics, then the relations are implicational and not necessarily causal.

7.2 Inference in Natural Language Understanding

The problem of using world knowledge in the interpretation of discourse, and in particular of drawing the appropriate inferences, has been investigated by a number of researchers for the last two decades. Among the earliest work was that of Rieger (Rieger, 1974; Schank, 1975). He and his colleagues implemented a system in which a sentence was mapped into an underlying representation on the basis of semantic information, and then all of the possible inferences that could be drawn were drawn. Where an ambiguity was present, those interpretations were best that yielded the most inferences. Rieger's work was seminal in that of those who appreciated the importance of world knowledge in text interpretation, his implementation was probably the most general and on the largest scale. But because he imposed no constraints on what inferences should be drawn, his method was inherently combinatorially explosive.

Recent work by Sperber and Wilson (1986) takes an approach very similar to Rieger's. They present a noncomputational attempt to characterize the relevance of utterances in discourse. They first define a contextual implication of some new information, say, that provided by a new utterance, to be a conclusion that can be drawn from the new information plus currently highlighted background knowledge but that cannot be drawn from either alone. An utterance is then relevant to the extent, essentially, that it has a large number of easily derived contextual implications. To extend this to the problem of interpretation, we could say that the best interpretation of an ambiguous utterance is the one that gives it the greatest relevance in the context.

In the late 1970s and early 1980s, Roger Schank and his students scaled back from the ambitious program of Rieger. They adopted a method for handling extended text that combined keywords and scripts. The text was scanned for particular keywords which were used to select the pre-stored script that was most likely to be relevant. The script was then used to guide the rest of the processing. This technique was used in the FRUMP program (DeJong, 1977; Schank et al., 1980) for summarizing stories on the Associated Press news wire that dealt with terrorist incidents and with disasters. Unconstrained inference was thereby avoided, but at a cost. The technique was necessarily limited to very narrow domains in which the texts to be processed described stereotyped scenarios and in which the information was conveyed in stereotyped ways. The more one examines even the seemingly simplest examples of spoken or written discourse, the more one realizes that very few cases satisfy these criteria.

In what can be viewed as an alternative response to Rieger's project, Hobbs (1980) proposed a set of constraints on the inferences that should be drawn in knowledge-based text processing: those inferences should be drawn that are required for the most economical solution to the discourse problems posed by the text. These problems include interpreting vague predicates, resolving definite references, discovering the congruence of predicates and their arguments, discovering the coherence relations among adjacent segments of text, and detecting the relation of the utterances to the speaker's or writer's overall plan. For each problem a discourse operation was defined, characterizing the forward and backward inferences that had to be drawn for that problem to be solved.

The difference in approaches can be characterized briefly as follows: The Rieger and the Sperber and Wilson models assume the unrestricted drawing of forward inferences, and the best interpretation of a text is the one that maximizes this set of inferences. The selective inferencing model posits certain external constraints on what counts as an interpretation, namely, that certain discourse problems must be solved, and the best interpretation is the the set of inferences, some backward and some forward, that satisfies these constraints most economically. In the abductive model, there is only one constraint, namely, that the text must be explained, and the best interpretation is the set of backward inferences that does this most economically. Whereas Rieger and Sperber and Wilson were forward-chaining from the text and trying to maximize implications, we are backward-chaining from the text and trying to minimize assumptions.

7.3 Abduction in Natural Language Understanding

Grice (1975) introduced the notion of "conversational implicature" to handle examples like the following:

A: How is John doing on his new job at the bank?
B: Quite well. He likes his colleagues and he hasn't embezzled any money yet.

Grice argues that in order to see this as coherent, we must assume, or draw as a conversational implicature, that both A and B know that John is dishonest. An implicature can be viewed as an abductive move for the sake of achieving the best interpretation.

Lewis (1979) introduces the notion of "accommodation" in conversation to explain the phenomenon that occurs when you "say something that requires a missing presupposition, and straightaway that presupposition springs into existence, making what you said acceptable after all." The hearer accommodates the speaker.

Thomason (1985) argued that Grice's conversational implicatures are based on Lewis's rule of accommodation. We might say that implicature is a procedural characterization of something that, at the functional or interactional level, appears as accommodation. When we do accommodation, implicature is what our brain does.

Hobbs (1979) recognized that many cases of pronoun reference resolution were in fact conversational implicatures, drawn in the service of achieving the most coherent interpretation of a text. Hobbs (1983a) gave an account of the interpretation of a spatial metaphor as a process of backward-chaining from the content of the utterance to a more specific underlying proposition, although the details are vague. Hobbs (1982b) showed how the notion of implicature can solve many problematic cases of definite reference. However, in none of this work was there a recognition of the all-pervading role of abductive explanation in discourse interpretation.

A more thorough-going early use of abduction in natural language understanding was in the work of Norvig (1983, 1987), Wilensky (1983; Wilensky et al., 1988), and their associates. They propose an operation of "concretion", one of many that take place in the processing of a text. It is a "kind of inference in which a more specific interpretation of an utterance is made than can be sustained on a strictly logical basis" (Wilensky et al., 1988, p. 50). Thus, "to use a pencil" generally means to write with a pencil, even though one could use a pencil for many other purposes. The operation of concretion works as follows: "A concept represented as an instance of a category is passed to the concretion mechanism. Its eligibility for membership in a more specific subcategory is determined by its ability to meet the constraints imposed on the subcategory by its associated relations and aspectual constraints. If all applicable conditions are met, the concept becomes an instance of the subcategory" (*ibid.*). In the terminology of our schema,

From $q(A)$ and $(\forall x)p(x) \supset q(x)$, conclude $p(A)$,

A is the concept, q is the higher category, and p is the more specific subcategory. Whereas Wilensky et al. view concretion as a special and somewhat questionable inference from $q(A)$, in the abductive approach it is a matter of determining the best explanation for $q(A)$.

The "associated relations and aspectual constraints" are other consequences of $p(A)$. In part, checking these is checking for the consistency of $p(A)$. In part, it is being able to explain the most with the least.

Norvig (1987), in particular, describes this process in terms of marker passing in a semantic net framework, deriving originally from Quillian (1968). Markers are passed from node to node, losing energy with each pass, until they run out of energy. When two markers collide, the paths they followed are inspected, and if they are of the right shape, they constitute the inferences that are drawn. Semantic nets express implicative relations, and their links can as easily be expressed as axioms. Hierarchical relations correspond to axioms of the form

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

and slots correspond to axioms of the form

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

Marker passing therefore is equivalent to forward- and backward-chaining in a set of axioms. Although we do no forward-chaining, the use of "et cetera" propositions described in Section 4 accomplishes the same thing. Norvig's "marker energy" corresponds to our costs; when the weights on antecedents sum to greater than one, that means cost is increasing and hence marker energy is decreasing. Norvig's marker collision corresponds to our factoring. We believe ours is a more compelling account of interpretation. There is really no justification for the operation of marker passing beyond the pretheoretic psychological notion that there are associations between concepts and one concept reminds us of another. And there is no justification at all for why marker collision is what should determine the inferences that are drawn and hence the interpretation of the text. In our formulation, by contrast, the interpretation of a text is the best explanation of why it would be true, "marker passing" is the search through the axioms in the knowledge base for a proof, and "marker collision" is the discovery of redundancies that yield more economic explanations.

Charniak and his associates have also been working out the details of an abductive approach to interpretation for a number of years. Charniak (1986) expresses the fundamental insight: "A standard platitude is that understanding something is relating it to what one already knows. ... One extreme example would be to prove that what one is told must be true on the basis of what one already knows. ... We want to prove what one is told *given certain assumptions*."

To compare Charniak's approach with ours, it is useful to examine in detail one of his operations, that for resolving definite references. In Charniak and Goldman (1988) the rule is given as follows:

```
(inst ?x ?frame) ⇒
      (OR (PExists (y :  ?frame)(== ?x ?y))^.9
           (→OR (role-inst ?x ?superfrm ?slot)
                (Exists (?s :  ?superfrm)
                     (== (?slot ?s) ?x))))^.1)
```

For the sake of concreteness, we will look at the example

John bought a new car. The engine is already acting up.

where the problem is to resolve "the engine". For the sake of comparing Charniak and Goldman's with our approach, let us suppose we have the axiom

(16) $(\forall y) car(y) \supset (\exists x) engine\text{-}of(x,y) \wedge engine(x)$

That is, if y is a car, then there is an engine x which is the engine of y. The relevant portion of the logical form of the second sentence is

$$(\exists \ldots, x, \ldots) \ldots \wedge engine(x) \wedge \ldots$$

and after the first sentence has been processed, $car(C)$ is in the knowledge base.

Now, Charniak and Goldman's expression `(inst ?x ?frame)` says that an entity `?x`, say, the engine, is an instance of a frame `?frame`, such as the frame `engine`. In our terminology, this is simply $engine(x)$. The first disjunct in the conclusion of the rule says that a `y` instantiating the same frame previously exists (`PExists`) in the text and is equal to (or the best name for) the mentioned engine. For us, that corresponds to the case where we already know $engine(E)$ for some E. In the second disjunct, the expression `(role-inst ?x ?superfrm ?slot)` says that `?x` is a possible filler for the `?slot` slot in the frame `?superfrm`, as the engine x is the engine x is a possible filler for the `engine-of` slot in the `car` frame. In our formulation, that corresponds to backward-chaining using axiom (16) and finding the predicate car. The expression

```
(Exists (?s :  ?superfrm)(== (?slot ?s) ?x))
```

says that some entity `?s` instantiating the frame `?superfrm` must exist, and its `?slot` slot is equal to (or the best name for) the definite entity `?x`. So in our example, we need to find a car whose existence is known or can be inferred. The operator →`OR` tells us to infer its first argument in all possible ways and then to prove its second argument with one of the resulting bindings. The superscripts on the disjuncts are probabilities that result in favoring the first over the second, thereby favoring shorter proofs.

The two disjuncts of Charniak and Goldman's rule therefore correspond to the two cases of not having to use axiom (16) in the proof of the engine's existence and having to use it. There are two ways of viewing the difference between Charniak and Goldman's formulation and ours. The first is that whereas they must explicitly state complex rules for definite reference, lexical disambiguation, case disambiguation, plan recognition, and other discourse operations in a complex metalanguage, we simply do backward-chaining on a set of axioms expressing our knowledge of the world. Their rules can be viewed as descriptions of this backward-chaining process: If you find $r(x)$ in the text, then look for an $r(A)$ in the preceding text, or, if that fails, look for an axiom of the form

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

and a $p(B)$ in the preceding text or the knowledge base, and make the appropriate identifications.

Alternatively, we can view Charniak and Goldman's rule as an axiom schema, one of whose instances is

$$
\begin{aligned}
(\forall\, x) engine(x) \supset\ & [(\exists\, y) engine(y) \wedge y = x] \\
& \vee [(\exists\, y) car(y) \wedge \textit{engine-of}(x, y)] \\
& \vee [(\exists\, y) truck(y) \wedge \textit{engine-of}(x, y)] \\
& \vee [(\exists\, y) plane(y) \wedge \textit{engine-of}(x, y)] \\
& \vee \ldots
\end{aligned}
$$

Konolige (1990) points out that abduction can be viewed as nonmonotonic reasoning with closure axioms and minimization over causes. That is, where there are a number of potential causes expressed as axioms of the form $P_i \supset Q$, we can write the closure axiom $Q \supset P_1 \vee P_2 \vee \ldots$, saying that if Q holds, then one of the P_i's must be its explanation. Then instead of backward-chaining through axioms of the first sort, we forward chain through axioms of the second sort. Minimization over the P_i's, or assuming as many of them as possible to be false, then selects the most economic conjunctions of P_i's for explaining Q. Our approach is of the first sort, Charniak and Goldman's of the second.

In more recent work, Goldman and Charniak (1990; Charniak and Goldman, 1989) have begun to implement their interpretation procedure in the form of an incrementally built belief network (Pearl, 1988), where the links between the nodes, representing influences between events, are determined from the axioms, stated as described above. They feel that one can make not unreasonable estimates of the required probabilities, giving a principled semantics to the numbers. The networks are then evaluated and ambiguities are resolved by looking for the highest resultant probabilities.

It is clear that minimality in the number of assumptions is not adequate for choosing among interpretations; this is why we have added weights. Ng and Mooney (1990) have proposed another criterion, which they call "explanatory coherence". They define a "coherence metric" that gives special weight to observations explained by other observations. One ought to be able to achieve this by factoring, but they give examples where factoring does not work. Their motivating examples, however, are generally short, two-sentence texts, where they fail to take into account that one of the facts to be explained is the adjacency of the sentences in a single, coherent text. When one does, one sees that their supposedly simple but low-coherence explanations are bad just because they explain so little. We believe it remains to be established that the coherence metric achieves anything that a minimality metric does not.

There has been other recent work on using abduction in the solution of various natural language problems, including the problems of lexical ambiguity (Dasigi, 1988, 1990), structural ambiguity (Nagao, 1989), and lexical selection (Zadrozny and Kokar, 1990).

8 Future Directions

8.1 Making Abduction More Efficient

Deduction is explosive, and since the abduction scheme augments deduction with two more options at each node—assumption and factoring—it is even more explosive. We are currently engaged in an empirical investigation of the behavior of this abductive scheme on a knowledge base of nearly 400 axioms, performing relatively sophisticated linguistic processing. So far, we have begun to experiment, with good results, with three different techniques for controlling abduction—a type hierarchy, unwinding or avoiding transitivity axioms, and various heuristics for reducing the branch factor of the search.

We expect our investigation to continue to yield techniques for controlling the abduction process.

The Type Hierarchy: The first example on which we tested the abductive scheme was the sentence

> There was adequate lube oil.

The system got the correct interpretation, that the lube oil was the lube oil in the lube oil system of the air compressor, and it assumed that that lube oil was adequate. But it also got another interpretation. There is a mention in the knowledge base of the adequacy of the lube oil pressure, so the system identified that adequacy with the adequacy mentioned in the sentence. It then assumed that the pressure was lube oil.

It is clear what went wrong here. Pressure is a magnitude whereas lube oil is a material, and magnitudes can't be materials. In principle, abduction requires a check for the consistency of what is assumed, and our knowledge base should have contained axioms from which it could be inferred that a magnitude is not a material. In practice, unconstrained consistency checking is undecidable and, at best, may take a long time. Nevertheless, one can, through the use of a type hierarchy, eliminate a very large number of possible assumptions that are likely to result in an inconsistency. We have consequently implemented a module that specifies the types that various predicate-argument positions can take on, and the likely disjointness relations among types. This is a way of exploiting the specificity of the English lexicon for computational purposes. This addition led to a speed-up of two orders of magnitude.

A further use of the type hierarchy speeds up processing by a factor of 2 to 4. The types provide prefiltering of relevant axioms for compound nominal, coercion, and other very general relations. Suppose, for example, that we wish to prove $rel(a, b)$, and we have the two axioms

$$p_1(x,y) \supset rel(x,y)$$
$$p_2(x,y) \supset rel(x,y)$$

Without a type hierarchy we would have to backward-chain on both of these axioms. If, however, the first of the axioms is valid only when x and y are of types t_1 and t_2, respectively, and the second is valid only when x and y are of types t_3 and t_4, respectively, and a and b have already been determined to be of types t_1 and t_2, respectively, then we need to backward-chain on only the first of the axioms.

There is a problem with the type hierarchy, however. In an ontologically promiscuous notation, there is no commitment in a primed proposition to truth or existence in the real world. Thus, $\textit{lube-oil}'(e, o)$ does not say that o is lube oil or even that it exists; rather it says that e is the eventuality of o's being lube oil. This eventuality may or may not exist in the real world. If it does, then we would express this as $Rexists(e)$, and from that we could derive from axioms the existence of o and the fact that it is lube oil. But

e's existential status could be something different. For example, e could be nonexistent, expressed as $not(e)$ in the notation, and in English as "The eventuality e of o's being lube oil does not exist," or simply as "o is not lube oil." Or e may exist only in someone's beliefs or in some other possible world. While the axiom

$$(\forall\, x) pressure(x) \supset \neg lube\text{-}oil(x)$$

is certainly true, the axiom

$$(\forall\, e_1, x) pressure'(e_1, x) \supset \neg(\exists\, e_2) lube\text{-}oil'(e_2, x)$$

would not be true. The fact that a variable occupies the second argument position of the predicate $lube\text{-}oil'$ does not mean it is lube oil. We cannot properly restrict that argument position to be lube oil, or fluid, or even a material, for that would rule out perfectly true sentences like "Truth is not lube oil."

Generally, when one uses a type hierarchy, one assumes the types to be disjoint sets with cleanly defined boundaries, and one assumes that predicates take arguments of only certain types. There are a lot of problems with this idea. In any case, in our work, we are not buying into this notion that the universe is typed. Rather, we are using the type hierarchy strictly as a heuristic, as a set of guesses not about what could or could not *be* but about what it would or would not occur to someone to *say*. When two types are declared to be disjoint, we are saying that they are certainly disjoint in the real world, and that they are very probably disjoint everywhere except in certain bizarre modal contexts. This means, however, that we risk failing on certain rare examples. We could not, for example, deal with the sentence, "It then assumed that the pressure was lube oil."

Unwinding or Avoiding Transitivity Axioms: At one point, in order to conclude from the sentence

> Bombs exploded at the offices of French-owned firms in Catalonia.

that the country in which the terrorist incident occurred was Spain, we wrote the following axiom:

$$(\forall\, x, y, z) in(x, y) \wedge partof(y, z) \supset in(x, z)$$

That is, if x is in y and y is a part of z, then x is also in z. The interpretation of this sentence was taking an extraordinarily long time. When we examined the search space, we discovered that it was dominated by this one axiom. We replaced the axiom with several axioms that limited the depth of recursion to three, and the problem disappeared.

In general, one must exercise a certain discipline in the axioms one writes. Which kinds of axioms cause trouble and how to replace them with adequate but less dangerous axioms is a matter of continuing investigation.

Reducing the Branch Factor of the Search: It is always useful to reduce the branch factor of the search for a proof wherever possible. We have devised several heuristics so far for accomplishing this.

The first heuristic is to prove the easiest, most specific conjuncts first, and then to propagate the instantiations. For example, in the domain of naval operations reports, words like "Lafayette" are treated as referring to classes of ships rather than to individual ships. Thus, in the sentence

Lafayette sighted.

"Lafayette" must be coerced into a physical object that can be sighted. We must prove the expression

$$(\exists\, x, y) sight(z, y) \,\wedge\, rel(y, x) \wedge Lafayette(x)$$

The predicate *Lafayette* is true only of the entity *LAFAYETTE-CLASS*. Thus, rather than trying to prove $rel(y, x)$ first, leading to a very explosive search, we try first to prove $Lafayette(x)$. We succeed immediately, and propagate the value *LAFAYETTE-CLASS* for x. We thus have to prove $rel(y,$*LAFAYETTE-CLASS*$)$. Because of the type of *LAFAYETTE-CLASS*, only one axiom applies, namely, the one allowing coercions from types to tokens that says that y must be an instance of *LAFAYETTE-CLASS*.

Similar heuristics involve solving reference problems before coercion problems and proving conjuncts whose source is the head noun of a noun phrase before proving conjuncts derived from adjectives.

Another heuristic is to eliminate assumptions wherever possible. We are better off if at any node, rather than having either to prove an atomic formula or to assume it, we only have to prove it. Some predicates are therefore marked as nonassumable. One category of such predicates is the "closed-world predicates", those predicates such that we know all entities of which the predicate is true. Predicates representing proper names, such as *Enterprise*, and classes, such as *Lafayette*, are examples. We don't assume these predicates because we know that if they are true of some entity, we will be able to prove it.

Another category of such predicates is the "schema-related" predicates. In the naval operations domain, the task is to characterize the participants in incidents described in the message. This is done as described in Section 5.4. A schema is encoded by means of a schema predication, with an argument for each role in the schema. Lexical realizations and other consequences of schemas are encoded by means of schema axioms. Thus, in the jargon of naval operations reports, a plane can splash another plane. The underlying schema is called *Init-Act*. There is thus an axiom

$$(\forall\, x, y, \ldots) Init\text{-}Act(x, y, attack, \ldots) \,\supset\, splash(x, y)$$

Schema-related predicates like *splash* occurring in the logical form of a sentence are given very large assumption costs, effectively preventing their being assumed. The weight associated with the antecedent of the schema axioms is very very small, so that the schema predication can be assumed very cheaply. This forces backward-chaining into the schema.

In addition, in the naval operations application, coercion relations are never assumed, since constraints on the arguments of predicates are what drives the use of the type hierarchy.

Factoring also multiplies the size of the search tree wherever it can occur. As explained above, it is a very powerful method for coreference resolution. It is based on the principle that where it can be inferred that two entities have the same property, there is a good possibility that the two entities are identical. However, this is true only for fairly specific properties. We don't want to factor predicates true of many things. For example, to resolve the noun phrase ships and planes

we need to prove the expression

$$(\exists\, x, s_1, y, s_2) Plural(x, s_1) \wedge ship(x) \wedge Plural(y, s_2) \wedge plane(y)$$

where $Plural$ is taken to be a relation between the typical element of a set and the set itself. If we applied factoring indiscriminately, then we would factor the conjuncts $Plural(x, s_1)$ and $Plural(y, s_2)$, identifying x with y and s_1 with s_2. If we were lucky, this interpretation would be rejected because of a type violation—planes aren't ships. But this would waste time. It is more reasonable to say that very general predicates such as $Plural$ provide no evidence for identity.

The type hierarchy, the discipline imposed in writing axioms, and the heuristics for limiting search all make the system less powerful than it would otherwise be, but we implement these techniques for the sake of efficiency. We are trying to locate the system on a scale whose extremes are efficiency and power. Where on that scale we achieve optimal performance is a matter of ongoing investigation.

8.2 Other Pragmatics Problems

In this paper we have described our approach to the problems of reference resolution, compound nominal interpretation, syntactic ambiguity, metonymy resolution, and schema recognition. These approaches have been worked out, implemented, and tested on a fairly large scale. We intend similarly to work out the details of an abductive treatment of other problems in discourse interpretation. These include the local pragmatics problems of lexical ambiguity, metaphor interpretation, and the resolution of quantifier scope ambiguities. Other problems of interest are the recognition of discourse structure (what Agar and Hobbs (1982) call local coherence) the recognition of the relation between the utterance and the speaker's plan (global coherence), and the drawing of quantity and similar implicatures. We will indicate very briefly for each of these problems what an abductive approach might look like.

Lexical Ambiguity: It appears that the treatment of lexical ambiguity is reasonably straightforward in our framework, adopting an approach advocated by Hobbs (1982a) and similar to the "polaroid word" method of Hirst (1987). An ambiguous word, like "bank", has a corresponding predicate $bank$ which is true of both financial institutions and the banks of rivers. There are two other predicates, $bank_1$ true of financial institutions and $bank_2$ true of banks of rivers. The three predicates are related by the two axioms

$$(\forall\, x) bank_1(x) \supset bank(x)$$
$$(\forall\, x) bank_2(x) \supset bank(x)$$

All world knowledge is then expressed in terms of either $bank_1$ or $bank_2$, not in terms of $bank$. In interpreting the text, we use one or the other of the axioms to reach into the knowledge base, and whichever one we use determines the intended sense of the word. Where these axioms are not used, it is apparently because the best interpretation of the text did not require the resolution of the lexical ambiguity.

This approach is essentially the same as the first-order approach to the compound nominal and metonymy problems.

Metaphor Interpretation: Hobbs (1983a) gave an account of metaphor interpretation within an inferential framework. There it was argued that metaphor interpretation is a matter of selecting the right inferences from what is said and rejecting the wrong ones. Thus, from

> John is an elephant.

we may infer that John is large or clumsy or has a good memory, but we won't infer that we should kill him for ivory. It was also shown how large-scale metaphor schemas could be handled in the same way. (See also Lakoff and Johnson, 1980, and Indurkhya, 1987.) This account was developed in a framework that ran the arrows in the opposite direction from the way they are in an abductive account. It was asked what one could infer from the text rather than what the text could be inferred from. But as described in Section 4, in the abductive approach implications can be converted into biconditionals, so it may be that this account of metaphor interpretation can be converted relatively easily into an abductive approach. The details remain to be worked out, however.

Resolving Quantifier Scope Ambiguities: Hobbs (1983b) proposed a flat representation for sentences with multiple quantifiers, consisting of a conjunction of atomic formulas, by admitting variables denoting sets and typical elements of sets, where the typical elements behave essentially like reified universally quantified variables, similar to McCarthy's (1977) "inner variables". Webber (1978), Van Lehn (1978), Mellish (1985), and Fahlman (1979) have all urged similar approaches in some form or other, although the technical details of such an approach are by no means easy to work out. (See Shapiro, 1980.) In such an approach, the initial logical form of a sentence, representing all that can be determined from syntactic analysis alone without recourse to world knowledge, is neutral with respect to the various possible scopings. As various constraints on the quantifier structure are discovered during pragmatics processing, the information is represented in the form of predications expressing "functional dependence" relations among sets and their typical elements. For example, in

> Three women in our group had a baby last year.

syntactic analysis of the sentence tells us that there is an entity w that is the typical example of a set of women, the cardinality of which is three, and there is an entity b that in some sense is a baby. What needs to be inferred is that b is functionally dependent on w.

In an abductive framework, what needs to be worked out is what mechanism will be used to infer the functional dependency. Is it, for example, something that must be assumed in order to avoid contradiction when the main predication of the sentence is assumed? Or is it something that we somehow infer directly from the propositional content of the sentence. Again, the problem remains to be worked out.

It may also be that if the quantifier scoping possibilities were built into the grammar rules in the integrated approach of Section 6, much as Montague (1974) did, the whole problem of determining the scopes of quantifiers will simply disappear into the larger problem of searching for the best interpretation, just as the problem of syntactic ambiguity did.

Discourse Structure: Hobbs (1985d) presented an account of discourse coherence in terms of a small number of "coherence relations" that can obtain between adjacent segments of text, recognizable by the content of the assertions of the segments. There are two possible approaches to this sort of discourse structure that we expect to explore. The first is the approach outlined in Section 6.3 above.

There is a second approach we may also explore, however. In 1979, Hobbs published a paper entitled "Coherence and Coreference", in which it was argued that coreference problems are often solved as a by-product of recognizing coherence. It may be appropriate, however, to turn this observation on its head and to see the coherence structure of the text as a kind of higher-order coreference. (This is similar to the approach of Lockman and Klapholz (1980) and Lockman (1978).) Where we see two sentences as being in an elaboration relation, for example, it is because we have inferred the same eventuality from the assertions of the two sentences. Thus, from both of the sentences

> John can open Bill's safe.
> He knows the combination.

we infer that there is some action that John/he can do that will cause the safe to be open. Rather than taking this to be the definition of a coherence relation of elaboration, we may instead want to view the second sentence as inferrable from the first, as long as certain other assumptions of a default nature are made. From this point of view, recognizing elaborations looks very much like ordinary reference resolution, as described in Section 3.

Causal relations can be treated similarly. Axioms would tell us in a general way what kinds of things cause and are caused by what. In

> John slipped on a banana peel,
> and broke his back.

we cannot infer the entire content of the second clause from the first, but we know in a general way that slipping tends to cause falls, and falls tend to cause injuries. If we take the second clause to contain an implicit definite reference to an injury, we can recover the causal relation between the two events, and the remainder of the specific information about the injury is new information and can be assumed.

Recognizing parallelism is somewhat more complex, but perhaps it can be seen as a kind of definite reference to types.

A disadvantage of this approach to discourse coherence is that it does not yield the large-scale coherence structure of the discourse in the same way as in the approach based on coherence relations. This is important because the coherence structure structures the context against which subsequent sentences are interpreted.

Recognizing the Speaker's Plan: It is a very common view that to interpret an utterance is to discover its relation to the speaker's presumed plan, and on any account, this relation is an important component of an interpretation. The most fundamental of the objections that Norvig and Wilensky (1990) raise to current abductive approaches to discourse interpretation is that they take as their starting point that the hearer must explain why the utterance is true rather than what the speaker was trying to accomplish with it. We agree with this criticism. Let us look at things from the broadest possible context. An intelligent agent is embedded in the world. Just as a hearer must explain why a sequence of words is a sentence or a coherent text, our agent must, at each instant, explain why the complete set of observables it is encountering constitutes a coherent situation. Other agents in the environment are viewed as intentional, that is, as planning mechanisms, and that means their observable actions are sequences of steps in a coherent plan. Thus, making sense of the environment entails making sense of other agents' actions in terms of what they are intended to achieve. When those actions are utterances, the utterances must be related to the goals those agents are trying to achieve. That is, the speaker's plan must be recognized.

Recognizing the speaker's plan is 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, it is a matter of deduction to prove that a sequence or more complex arrangement of actions will achieve an agent's goals, given the agent's beliefs. Unfortunately, we rarely have complete knowledge. We will almost always have to make assumptions. That is, abduction will be called for. To handle this aspect of interpretation in our framework, therefore, we can take it as one of our tasks, in addition to proving the logical form, to 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 begun research on how weighted abduction can be used for the plan ascription problem.

There is a point, however, at which the "intentional" view of interpretation becomes trivial. It tells us that the proper interpretation of a compound nominal like "coin copier" means what the speaker intended it to mean. This is true enough, but it offers us virtually no assistance in determining what it really *does* mean. It is at this point where the "informational" view of interpretation comes into play. We are working for the most part in the domain of common knowledge, so in fact what the speaker intended a sentence to mean is just what can be proved to be true from that base of common knowledge. That is, the best interpretation of the sentence is the best explanation for why it would be true, given the speaker and hearer's common knowledge. So while we agree that the intentional view of interpretation is correct, we believe that the informational view is a necessary component of that, a component that moreover, in analyzing long written texts and monologues, completely overshadows all other components.

Quantity Implicatures: When someone says,

(17) I have two children.

we conclude, in most circumstances, in a kind of implicature, that he does not have three children. If he had three children, he would have said so. This class of implicature has been studied by Levinson (1983), among others.

The general problem is that often the inferences we draw from an utterance are determined by what else the speaker could have said but didn't. Thus, in Grice's (1975) example,

> Miss X produced a series of sounds that corresponded closely with the score of "Home sweet home".

we conclude from the fact that the speaker could have said, "Miss X sang 'Home sweet home'", that in fact opening the mouth and making noises did not constitute singing, even though we might normally assume it would.

The logical structure of this phenomenon is the following: The speaker utters U_1. The best interpretation for U_1 is I_1. But the hearer uses his own generation processes to determine that if one wanted to convey meaning I_1, the most reasonable utterance would be U_2. There must be some reason the speaker chose to say U_1 instead. The hearer thus determines the content of U_2 that is not strictly entailed by U_1, and concludes that that difference does not hold. From sentence (17), the most reasonable interpretation I_1 is that $\mid Children \mid \geq 2$. If the speaker had three children, the most natural utterance U_2 would be "I have three children." Thus, we draw as an implicature the negation of the difference between U_2 and U_1, namely, $\neg(\mid Children \mid > 2)$.

This is a rather formidable phenomenon to proceduralize, because it seems to involve the hearer in the whole process of generation, and not just of one sentence, but rather of all the different ways the same information could have been conveyed.

We do not have a clear idea of how we would handle this phenomenon in our framework. But we are encouraged by the fact that interpretation and generation can be captured in exactly the same framework, as described in Section 6.6. It is consequently quite possible that this framework will give us a mechanism for examining not just the interpretation of an utterance but also adjacent possible realizations of that interpretation.

8.3 What the Numbers Mean

The problem of how to combine symbolic and numeric schemes in the most effective way, exploiting the expressive power of the first and the evaluative power of the second, is one of the most significant problems that faces researchers in artificial intelligence today. The abduction scheme we have presented attempts just this. However, our numeric component is highly *ad hoc* at the present time. We need a more principled account of what the numbers mean. Here we point out several possible lines of investigation.

First let us examine the roles of weights. It seems that a principled approach is most likely to be one that relies on probability. But what is the space of events over which the probabilities are to be calculated? Suppose we are given our corpus of interest. Imagine that a TACITUS-system-in-the-sky runs on this entire corpus, interpreting all the texts and instantiating all the abductive inferences it has to draw. This gives us a set of propositions Q occurring in the texts and some propositions P drawn from the knowledge base. It is possible that the weights w_i should be functions of probabilities and conditional probabilities involving instances of the concepts P and instances of concepts Q.

Given this space of events, the first question is how the weights should be distributed across the conjuncts in the antecedents of Horn clauses. In formula (6), repeated here for convenience,

$$(6) \quad P_1^{w_1} \wedge P_2^{w_2} \supset Q$$

one has the feeling that the weights should correspond somehow to the semantic contribution that each of P_1 and P_2 make to Q. The semantic contribution of P_i to Q may best be understood in terms of the conditional probability that an instance of concept Q is an instance of concept P_i in the space of events, $Pr(Q \mid P_i)$. If we distribute the total weight w of the antecedent of (6) according to these conditional probabilities, then

$$w_i = \frac{w\, Pr(Q|P_i)}{Pr(Q|P_1)+Pr(Q|P_2)}$$

The next question is what the total weight on the antecedent should be. To address this question, let us suppose that all the axioms have just one conjunct in the antecedent. Then we consider the set of axioms that have Q as the conclusion:

$$\begin{array}{l} P_1^{w_1} \supset Q \\ P_2^{w_2} \supset Q \\ \ldots \\ P_k^{w_k} \supset Q \end{array}$$

Intuitively, the price we will have to pay for the use of each axiom should be inversely related to the likelihood that Q is true by virtue of that axiom. That is, we want to look at the conditional probability that P_i is true given Q, $Pr(P_i \mid Q)$. The weights w_i should be ordered in the reverse order of these conditional probabilities. We need to include in this ordering the likelihood of Q occurring in the space of events without any of the P_i's occurring, $Pr(\neg(P_1 \wedge \ldots \wedge P_k) \mid Q)$, to take care of those cases where the best assumption for Q was simply Q itself. In assigning weights, this should be anchored at 1, and the weights w_i should be assigned accordingly.

All of this is only the coarsest pointer to a serious treatment of the weights in terms of probabilities.

A not entirely dissimilar approach to the question is in terms of model preference relations for nonmonotonic logics (Shoham, 1987). This is suggested by the apparent resemblance between our abduction scheme and various forms of nonmonotonic logic. For example, in circumscriptive theories (McCarthy, 1987) it is usual to write axioms like

$$(\forall x) bird(x) \wedge \neg Ab_1(x) \supset flies(x)$$

This certainly looks like the axiom

$$(\forall x) bird(x) \wedge etc_1(x)^{w_1} \supset flies(x)$$

The literal $\neg Ab_1(x)$ says that x is not abnormal in some particular respect. The literal $etc_1(x)$ says that x possesses certain unspecified properties, for example, that x is not abnormal in that same respect. In circumscription, one minimizes over the abnormality predicates, assuming they are false wherever possible, perhaps with a partial ordering on abnormality predicates to determine which assumptions to select (e.g., Poole, 1989). Our abduction scheme generalizes this a bit: The literal $etc_1(x)$ may be assumed if no contradiction results and if the resulting proof is the most economical one available. Moreover, the "et cetera" predicates can be used for any kind of differentiae distinguishing a species from the rest of a genus, and not just for those related to normality.

This observation suggests that a semantics can be specified for the abduction scheme along the lines developed for nonmonotonic logic. Appelt (1990) is exploring an approach to the semantics of the weights, based not on probabilities but on preference relations among models. Briefly, when we have two axioms of the form

$$\begin{array}{l} P_1^{w_1} \supset Q \\ P_2^{w_2} \supset Q \end{array}$$

where w_1 is less than w_2, we take this to mean that if then every model in which P_1, Q, and $\neg P_2$ are true is preferred over some model in which P_2, Q, and $\neg P_1$ are true. Appelt's approach exposes problems of unintended side-effects. Elsewhere among the axioms, P_2 may entail a highly preferred proposition, even though w_2 is larger than w_1. To get around this problem, Appelt must place very tight global constraints on the assignment of weights. This difficulty may be fundamental, resulting from the fact that the abduction scheme attempts to make global judgments on the basis of strictly local information.

So far we have only talked about the semantics of the weights, and not the costs. Hasida (personal communication) has suggested that the costs and weights be viewed along the lines of an economic model of supply and demand. The requirement to interpret texts creates a demand for propositions to be proved. The costs reflect that demand. Those most likely to anchor the text referentially are the ones that are in the greatest demand; therefore, they cost the most to assume. The supply, on the other hand, corresponds to the probability that the propositions are true. The more probable the proposition, the less it should cost to assume, hence the smaller the weight.

Charniak and Shimony (1990) have proposed a probabilistic semantics for weighted abduction schemes. They make the simplifying assumption that a proposition always has the same cost, wherever it occurs in the inference process, although rules themselves may also have an associated cost. They consider only the propositional case, so, for example, no factoring or equality assumptions are needed. They further assume that the axioms are acyclic. Finally, they concern themselves only with the probability that the propositions are true, and do not try to incorporate utilities into their cost functions as we do. They show that a set of axioms satisfying these restrictions can be converted into a Bayesian network where the negative logarithms of the prior probabilities of the nodes are the assumability costs of the propositions. They then show that the assignment of truth values to the nodes in the Bayesian network with maximum probability given the evidence is equivalent to the assignment of truth values to the propositions that minimizes cost. We view this as a promising start toward a semantics for the less restricted abduction scheme we have used.

A further requirement for the scoring scheme is that it incorporate not only the costs of assumptions, but also the costs of inference steps, where highly salient inferences cost less than inferences of low salience. The obvious way to do this is to associate costs with the use of each axiom, where the costs are based on the axiom's salience, and to levy that cost as a charge for each proof step involving the axiom. If we do this, we need a way of correlating the cost of inference steps with the cost of assumptions; there must be a common coin of the realm. Can we develop a semantics for the numbers that relates assumption costs and inference costs? Two moves are called for: interpreting the cost of inference as uncertainty and interpreting salience as truth in a local theory.

The first move is to recognize that virtually all of our knowledge is uncertain to some degree. Then we can view the cost of using an axiom to be a result of the greater uncertainty that is introduced by assuming that axiom is true. This can be done with "et cetera" propositions, either at the level of the axiom as a whole or at the level of its instantiations. To associate the cost with the general axiom, we can write our axioms as follows:

$$(\forall x)[p(x) \wedge etc_1^{\$c_1} \supset q(x)]$$

That is, there is no dependence on x. Then we can use any number of instances of the axiom once we pay the price c_1. To associate the cost with each instantiation of the axiom, we can write our axioms as follows:

$$(\forall x)[p(x) \wedge etc_1(x)^{\$c_1} \supset q(x)]$$

Here we must pay the price of c_1 for every instance of the axiom we use. The latter style seems more reasonable.

Furthermore, it seems reasonable not to charge for multiple uses of particular instantiations of axioms; we need to pay for $etc_1(A)$ only once for any given A. This intuition supports the uncertainty interpretation of inference costs.

It is easy to see how a salience measure can be implemented in this scheme. Less salient axioms have higher associated costs c_1. These costs can be changed from situation to situation if we take the cost c_1 to be not a constant but a function that is sensitive somehow to the contextual factors affecting the salience of different clusters of knowledge. Alternatively, if axioms are grouped into clusters and tagged with the cluster they belong to, as in

$$(\forall x)p(x) \wedge cluster^{\$c_1} \supset q(x)$$

then whole clusters can be moved from low salience to high salience by paying the cost $\$c_1$ of the "proposition" *cluster* exactly once.

But can this use of the costs also be interpreted as a measure of uncertainty? We suspect it can, based on ideas discussed in Hobbs (1985c). There it is argued that whenever intelligent agents are interpreting and acting in specific environments, they are doing so not on the basis of everything they know, their entire knowledge base, but rather on the basis of local theories that are already in place for reasoning about this type of situation or are constructed somehow for the occasion. At its simplest, a local theory is a relatively small subset of the entire knowledge base; more complex versions are also imaginable, in which axioms are modified in some way for the local theory. In this view, a local theory creates a binary distinction between the axioms that are true in the local theory and the axioms in the global theory that are not necessarily true. However, in the abductive framework, the local theory can be given a graded edge by assigning values to the costs c_1 in the right way. Thus, highly salient axioms will be in the core of the local theory and will have relatively low costs. Low-salience axioms will be ones for which there is a great deal of uncertainty as to whether they are relevant to the given situation and thus whether they should actually be true in the local theory; they will have relatively high costs. Salience can thus be seen as a measure of the certainty that an axiom is true in the local theory.

Josephson et al. (1987) have argued that an evaluation scheme must consider the following criteria when choosing a hypothesis H to explain some data D:

1. How decisively does H surpass its alternatives?
2. How good is H by itself, independent of the alternatives?
3. How thorough was the search for alternatives?
4. What are the risks of being wrong and the benefits of being right?
5. How strong is the need to come to a conclusion at all?

Of these, our abduction scheme uses the weights and costs to formalize criterion 2, and the costs at least in part address criteria 4 and 5. But criteria 1 and 3 are not accommodated at all. The fact that our abduction scheme does not take into account the competing possible interpretations is a clear shortcoming that needs to be corrected.

A theoretical account, such as the one we have sketched, can inform our intuitions, but in practice we can only assign weights and costs by a rough, intuitive sense of semantic contribution, importance, and so on, and refine them by successive approximation on a representative sample of the corpus. But the theoretical account would at least give us a clear view of what the approximations are approximating.

9 Conclusion

Interpretation in general may be viewed as abduction. When we look out the window and see a tree waving back and forth, we normally assume the wind is blowing. There may be other reasons for the tree's motion; for example, someone below window level might be shaking it. But most of the time the most economical explanation coherent with the rest of what we know will be that the wind is blowing. This is an abductive explanation. Moreover, in much the same way as we try to exploit the redundancy in natural language discourse, we try to minimize our explanations for the situations we encounter by identifying disparately presented entities with each other wherever possible. If we see a branch of a tree occluded in the middle by a telephone pole, we assume that there is indeed just one branch and not two branches twisting bizarrely behind the telephone pole. If we hear a loud noise and the lights go out, we assume one event happened and not two.

These observations make the abductive approach to discourse interpretation more appealing. Discourse interpretation is seen, as it ought to be seen, as just a special case of interpretation. From the viewpoint of Section 6.3, to interpret a text is to prove abductively that it is coherent, where part of what coherence is is an explanation for why the text would be true. Similarly, one could argue that faced with any scene or other situation, we must prove abductively that it is a coherent situation, where part of what coherence means is explaining why the situation exists.[21]

Moreover, the particular abduction scheme we use, or rather the ultimate abduction scheme of which our scheme is an initial version, has a number of other attractive properties. It gives us the expressive power of predicate logic. It allows the defeasible reasoning of nonmonotonic logics. Its numeric evaluation method begins to give reasoning the "soft corners" of neural nets. It provides a framework in which a number of traditionally difficult problems in pragmatics can be formulated elegantly in a uniform manner. Finally, it gives us a framework in which many types of linguistic processing can be formalized in a thoroughly integrated fashion.

Acknowledgments

The authors have profited from discussions with Douglas Edwards, Eugene Charniak, Todd Davies, Koiti Hasida, John Lowrance, Fernando Pereira, Stuart Shieber, Mabry Tyson, and Sheryl Young about this work. The research was funded by the Defense Advanced Research Projects Agency under Office of Naval Research contract N00014–85–C–0013.

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] Appelt, Douglas, 1990. "A Theory of Abduction Based on Model Preference", in P. O'Rorke, ed., *Working Notes*, AAAI Spring Symposium on Automated Abduction, Stanford, California, March 1990, pp. 67–71.

[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] Bear, John, and Jerry R. Hobbs, 1988. "Localizing the Expression of Ambiguity", *Proceedings*, Second Conference on Applied Natural Language Processing, Austin, Texas, February, 1988.

[5] Bever, Thomas, 1970. "The Cognitive Basis for Linguistic Structures", in J. Hayes, ed., *Cognition and the Development of Language*, pp. 279–352, John Wiley & Sons, New York.

---

[21] When this viewpoint is combined with that of Section 6.6 of action as abduction, one begins to suspect the brain is primarily a large and complex abduction machine.

[6] Charniak, Eugene, 1986. "A Neat Theory of Marker Passing", *Proceedings*, AAAI-86, Fifth National Conference on Artificial Intelligence, Philadelphia, Pennsylvania, pp. 584–588.

[7] Charniak, Eugene, and Robert Goldman, 1988. "A Logic for Semantic Interpretation", *Proceedings*, 26th Annual Meeting of the Association for Computational Linguistics, pp. 87–94, Buffalo, New York, June 1988.

[8] Charniak, Eugene, and Robert Goldman, 1989. "A Semantics for Probabilistic Quantifier-Free First-Order Languages, with Particular Application to Story Understanding", *Proceedings*, Eleventh International Joint Conference on Artificial Intelligence, pp. 1074–1079. Detroit, Michigan. August 1989.

[9] Charniak, Eugene, and Drew McDermott, 1985. *Introduction to Artificial Intelligence*, Addison-Wesley Publishing Co., Reading, Massachusetts.

[10] Charniak, Eugene, and Solomon E. Shimony, 1990. "Probabilistic Semantics for Cost Based Abduction", Technical Report CS-90-02, Department of Computer Science, Brown University, February 1990.

[11] Clark,Herbert, 1975. "Bridging", in R. Schank and B. Nash-Webber, eds., *Theoretical Issues in Natural Language Processing*, pp. 169–174, Cambridge, Massachusetts.

[12] Cox, P. T., and T. Pietrzykowski, 1986. "Causes for Events: Their Computation and Applications", in J. Siekmann, ed., *Proceedings*, 8th International Conference on Automated Deduction (CADE-8), Springer-Verlag, Berlin.

[13] Crain, S., and Mark Steedman, 1985. "On Not Being Led Up the Garden Path: The Use of Context by the Psychological Parser", in D. Dowty, L. Karttunen, and A. Zwicky, eds., *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*, Cambridge University Press, Cambridge, England.

[14] Dasigi, Venu R., 1988. *Word Sense Disambiguation in Descriptive Text Interpretation: A Dual-Route Parsimonious Covering Model* (doctoral dissertation), Technical Report TR-2151, Department of Computer Science, University of Maryland, College Park, December, 1988. Also published as Technical Report WSU-CS-90-03, Department of Computer Science and Engineering, Wright State University, Dayton, Ohio.

[15] Dasigi, Venu R., 1990. "A Dual-Route Parsimonious Covering Model of Descriptive Text Interpretation", in F. Gardin et al., eds., *Computational Intelligence II*, North-Holland, New York.

[16] DeJong, Gerald F., 1977. "Skimming Newspaper Stories by Computer", Research Report 104, Department of Computer Science, Yale University.

[17] Downing, Pamela, 1977. "On the Creation and Use of English Compound Nouns", *Language*, Vol. 53, No. 4, pp. 810–842.

[18] Fahlman, Scott E., 1979. *NETL: A System for Representing and Using Real-World Knowledge*, MIT Press, Cambridge, Massachusetts.

[19] Fodor, Jerry A., 1983. *The Modularity of Mind: An Essay on Faculty Psychology*, Bradford Books, MIT Press, Cambridge, Massachusetts.

[20] Fodor, Jerry A., n.d. "On the Modularity of Parsing: A Review", manuscript.

[21] Goldman, Robert P., and Eugene Charniak, 1990. "Incremental Construction of Probabilistic Models for Language Abduction: Work in Progress", in P. O'Rorke, ed., *Working Notes: AAAI Spring Symposium on Automated Abduction*, Stanford University, Stanford, California, March 1990, pp. 1–4.

[22] Grice, H. P., 1975. "Logic and Conversation", in P. Cole and J. Morgan, eds., *Syntax and Semantics*, Vol. 3, pp. 41–58, Academic Press, New York.

[23] Hirst, Graeme, 1987. *Semantic Interpretation and the Resolution of Ambiguity*. Cambridge University Press, Cambridge, England.

[24] Hobbs, Jerry R., 1978, "Resolving Pronoun References", *Lingua*, Vol. 44, pp. 311–338. Also in B. Grosz, K. Sparck-Jones, and B. Webber, eds., *Readings in Natural Language Processing*, pp. 339–352, Morgan Kaufmann Publishers, Los Altos, California.

[25] Hobbs, Jerry, 1979, "Coherence and Coreference", *Cognitive Science*, Vol. 3, No. 1, pp. 67–90.

[26] Hobbs, Jerry R., 1980. "Selective Inferencing", *Proceedings*, Third National Conference of the Canadian Society for Computational Studies of Intelligence, pp. 101–114, Victoria, British Columbia, May 1980.

[27] Hobbs, Jerry R., 1982a. "Representing Ambiguity", *Proceedings*, First West Coast Conference on Formal Linguistics, Stanford, California, January 1982, pp. 15–28.

[28] Hobbs, Jerry R., 1982b. "Implicature and Definite Reference", talk delivered at the Workshop on Modelling Real-time Language Processes, Port Camargue, France, June 1982. Published as Report No. CSLI-88-99, Center for the Study of Language and Information, Stanford University, Stanford, California, May 1987.

[29] Hobbs, Jerry R., 1983a. "Metaphor Interpretation as Selective Inferencing: Cognitive Processes in Understanding Metaphor", *Empirical Studies in the Arts*, Vol. 1, No. 1, pp. 17–34, and Vol. 1, No. 2, pp. 125–142.

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

[31] Hobbs, Jerry R. 1985a. "Ontological promiscuity." *Proceedings*, 23rd Annual Meeting of the Association for Computational Linguistics, pp. 61–69.

[32] Hobbs, Jerry R., 1985b, "The Logical Notation: Ontological Promiscuity", unpublished manuscript.

[33] Hobbs, Jerry R., 1985c. "Granularity", *Proceedings*, Ninth International Joint Conference on Artificial Intelligence, pp. 432–435. Los Angeles, California. August 1985. Also in D. Weld and J. de Kleer, eds., *Readings in Qualitative Reasoning about Physical Systems*, pp. 542–545, Morgan Kaufmann Publishers, San Mateo, California, 1989.

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

[35] Hobbs, Jerry R., 1986. "Overview of the TACITUS Project", *Computational Linguistics*, Vol. 12, No. 3.

[36] Hobbs, Jerry R., and John Bear, 1990. "Two Principles of Parse Preference", in H. Karlgren, ed., *Proceedings*, Thirteenth International Conference on Computational Linguistics, Helsinki, Finland, Vol. 3, pp. 162–167, August, 1990.

[37] Hobbs, Jerry R., William Croft, Todd Davies, Douglas Edwards, and Kenneth Laws, 1987. "Commonsense Metaphysics and Lexical Semantics", *Computational Linguistics*, Vol. 13, nos. 3–4, July–December 1987, pp. 241–250.

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

[39] Hobbs, Jerry R., and Paul Martin 1987. "Local Pragmatics". *Proceedings*, International Joint Conference on Artificial Intelligence, pp. 520–523. Milano, Italy, August 1987.

[40] Indurkhya, Bipin, 1987. "Approximate Semantic Transference: A Computational Theory of Metaphors and Analogies", *Cognitive Science*, Vol. 11, No. 4, pp. 445–480, October–December 1987.

[41] Joos, Martin, 1972. "Semantic Axiom Number One", *Language*, Vol. 48, pp. 257–265.

[42] Josephson, John R., 1990. "On the 'Logical Form' of Abduction", in P. O'Rorke, ed., *Working Notes*, AAAI Spring Symposium on Automated Abduction, Stanford, California, March 1990, pp. 140–144.

[43] Josephson, John R., B. Chandrasekaran, J. W. Smith, and M. C. Tanner, 1987. "A Mechanism for Forming Composite Explanatory Hypotheses", *IEEE Transactions on Systems, Man and Cyberbetics*, Vol. 17, pp. 445–54.

[44] Konolige, Kurt, 1990. "A General Theory of Abduction", in P. O'Rorke, ed., *Working Notes: AAAI Spring Symposium on Automated Abduction*, Stanford University, Stanford, California, March 1990, pp. 62–66.

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

[46] Lakatos, Imre, 1970. "Falsification and the Methodology of Scientific Research Programmes", in I. Lakatos and A. Musgrave, eds., *Criticism and the Growth of Knowledge*, Cambridge University Press, Cambridge, England.

[47] Lakoff, George, and Mark Johnson, 1980. *Metaphors We Live By*, University of Chicago Press, Chicago.

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

[49] Levinson, Stephen C., 1983. *Pragmatics*, Cambridge University Press, Cambridge, England.

[50] Lewis, David, 1979. "Scorekeeping in a Language Game," *Journal of Philosophical Logic*, Vol. 6, pp. 339–59.

[51] Lockman, Abraham, 1978. "Contextual Reference Resolution in Natural Language Processing", Ph.D. thesis, Department of Computer Science, Columbia University, May 1978.

[52] Lockman, Abraham, and David Klapholz, 1980. "Toward a Procedural Model of Contextual Reference Resolution", *Discourse Processes*, Vol. 3, pp. 25–71.

[53] Marslen-Wilson, William, and Lorraine Tyler, 1987. "Against Modularity", in J. L. Garfield, ed., *Modularity in Knowledge Representation and Natural Language Processing*, MIT Press, Cambridge, Massachusetts.

[54] McCarthy, John, 1977. "Epistemological Problems of Artificial Intelligence", *Proceedings*, International Joint Conference on Artificial Intelligence, pp. 1038–1044, Cambridge, Massachusetts, August 1977.

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

[56] Mellish, Chris, 1985. *Computer Interpretation of Natural Language Descriptions*, Ellis Horwood / John Wiley, Chichester, England.

[57] Montague, Richard, 1974. "The Proper Treatment of Quantification in Ordinary English", in R. H. Thomason, ed., *Formal Philosophy: Selected Papers of Richard Montague*, pp. 247–270, Yale University Press, New Haven, Connecticut.

[58] Nagao, Katashi, 1989. "Semantic Interpretation Based on the Multi-World Model", in *Proceedings*, Eleventh International Conference on Artificial Intelligence. Detroit, Michigan.

[59] Newton, Isaac, 1934 [1686]. *Mathematical Principles of Natural Philosophy*, Vol. 1: *The Motion of Bodies*, and Vol. 2: *The System of the World*, translated by Andrew Motte and Florian Cajori, University of California Press, Berkeley, California.

[60] Ng, Hwee Tou, and Raymond J. Mooney, 1990. "The Role of Coherence in Constructing and Evaluating Abductive Explanations", in P. O'Rorke, ed., *Working Notes*, AAAI Spring Symposium on Automated Abduction, Stanford, California, March 1990.

[61] Norvig, Peter, 1983. "Frame Activated Inferences in a Story Understanding Program", *Proceedings*, 8th International Joint Conference on Artificial Intelligence, Karlsruhe, West Germany, pp. 624–626.

[62] Norvig, Peter, 1987. "Inference in Text Understanding", *Proceedings*, AAAI-87, Sixth National Conference on Artificial Intelligence, Seattle, Washington, July 1987.

[63] 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.

[64] Nunberg, Geoffery, 1978. "The Pragmatics of Reference", Ph. D. thesis, City University of New York, New York.

[65] O'Rorke, Paul (editor), 1990. *Working Notes: AAAI Spring Symposium on Automated Abduction*, Stanford University, Stanford, California, March 1990.

[66] Pearl, Judea, 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, San Mateo, California.

[67] Peng, Yun, and James A. Reggia, 1987a. "A Probabilistic Causal Model for Diagnostic Problem Solving, Part One: Integrating Symbolic Causal Inference with Numeric Probabilistic Inference", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-17, No. 2, pp. 146–162, March/April 1987.

[68] Peng, Yun, and James A. Reggia, 1987b. "A Probabilistic Causal Model for Diagnostic Problem Solving—Part II: Diagnostic Strategy", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-17, No. 3, pp. 395–406, May/June 1987.

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

[70] Pierce, Charles Sanders, 1955. "Abduction and Induction", in J. Buchler, ed., *Philosophical Writings of Pierce*, pp. 150–156, Dover Books, New York.

[71] Poole, David, 1989. "Explanation and Prediction: An Architecture for Default and Abductive Reasoning", *Computational Intelligence*, Vol. 5, No. 2, pp. 97–110.

[72] Pople, Harry E., Jr., 1973, "On the Mechanization of Abductive Logic", *Proceedings*, Third International Joint Conference on Artificial Intelligence, pp. 147–152, Stanford, California, August 1973.

[73] Quillian, M. Ross, 1968. "Semantic Memory", in M. Minsky, ed., *Semantic Information Processing*, pp. 227–270, MIT Press, Cambridge, Massachusetts.

[74] Reggia, James A., 1985. "Abductive Inference", in K. N. Karna, ed., *Proceedings*, Expert Systems in Government Symposium, pp. 484–489, IEEE Computer Society Press, New York.

[75] Reggia, James A., Dana S. Nau, and Pearl Y. Wang, 1983. "Diagnostic Expert Systems Based on a Set Covering Model", *International Journal of Man-Machine Studies*, Vol. 19, pp. 437–460.

[76] Rieger, Charles J., III., 1974. "Conceptual Memory: A Theory and Computer Program for Processing the Meaning Content of Natural Language Utterances", Memo AIM-233, Stanford Artificial Intelligence Laboratory, Stanford University.

[77] Robinson, Jane, 1982. "DIAGRAM: A Grammar for Dialogues", *Communications of the ACM*, Vol. 25, No. 1, pp. 27–47, January 1982.

[78] Sager, Naomi, 1981. *Natural Language Information Processing: A Computer Grammar of English and Its Applications*, Addison-Wesley, Reading, Massachusetts.

[79] Schank, Roger. 1975. *Conceptual Information Processing.* Elsevier, New York.

[80] Schank, Roger C., Michael Lebowitz, and Lawrence Birnbaum, 1980. "An Integrated Understander", *American Journal of Computational Linguistics*, Vol. 6, No. 1, January–March 1980.

[81] Shapiro, Stuart C., 1980. "Review of *NETL: A System for Representing and Using Real-World Knowledge*, by Scott E. Fahlman", *American Journal of Computational Linguistics*, Vol. 6, Nos. 3–4, pp. 183–186, July–December 1980.

[82] Shieber, Stuart M., 1988. "A Uniform Architecture for Parsing and Generation", *Proceedings*, 12th International Conference on Computational Linguistics, pp. 614–619, Budapest, Hungary.

[83] Shoham, Yoav, 1987. "Nonmonotonic Logics: Meaning and Utility", *Proceedings*, International Joint Conference on Artificial Intelligence, pp. 388–393. Milano, Italy, August 1987.

[84] Sperber, Dan, and Deirdre Wilson, 1986. *Relevance: Communication and Cognition*, Harvard University Press, Cambridge, Massachusetts.

[85] Stickel, Mark E., 1989. "Rationale and Methods for Abductive Reasoning in Natural-Language Interpretation", in R. Studer, ed., *Proceedings*, Natural Language and Logic, International Scientific Symposium, Hamburg, Germany, May 1989, *Lecture Notes in Artificial Intelligence* #459, pp. 233–252, Springer-Verlag, Berlin.

[86] Thagard, Paul R., 1978. "The Best Explanation: Criteria for Theory Choice", *The Journal of Philosophy*, pp. 76–92.

[87] Thomason, Richmond H., 1985. "Accommodation, Conversational Planning, and Implicature", *Proceedings*, Workshop on Theoretical Approaches to Natural Language Understanding, Halifax, Nova Scotia, May 1985.

[88] Tyson, Mabry, and Jerry R. Hobbs, 1990. "Domain-Independent Task Specification in the TACITUS Natural Language System", Technical Note 488, Artificial Intelligence Center, SRI International, May 1990.

[89] Van Lehn, Kurt, 1978. "Determining the Scope of English Quantifiers", Massachusetts Institute of Technology Artificial Intelligence Laboratory Technical Report AI-TR-483, Cambridge, Massachusetts, June 1978.

[90] Webber, Bonnie L., 1978. "A Formal Approach to Discourse Anaphora", BBN Report No. 3761, Bolt, Beranek, and Newman Inc. Cambridge, Mass. May 1978.

[91] Wilensky, Robert, 1983. *Planning and Understanding: A Computational Approach to Human Reasoning*, Addison-Wesley, Reading, Massachusetts.

[92] Wilensky, Robert, David N. Chin, Marc Luria, James Martin, James Mayfield, and Dekai Wu, 1988. "The Berkeley UNIX Consultant Project", *Computational Linguistics*, Vol. 14, No. 4, December 1988, pp. 35–84.

[93] Wilks, Yorick, 1972. *Grammar, Meaning, and the Machine Analysis of Language*, Routledge and Kegan Paul, London.

[94] Zadrozny, Wlodek, and Mieczyslaw M. Kokar, 1990. "A Logical Model of Machine Learning: A Study of Vague Predicates", in P. Benjamin, ed., *Change of Representation and Inductive Bias*, pp. 247–266, Kluwer, Amsterdam.