Appears in *Proceedings of the Eighth National Conference on Artificial Intelligence (AAAI-90)*, pp.337-342, Boston, MA, July, 1990

On the Role of Coherence in Abductive Explanation*

Hwee Tou Ng
Raymond J. Mooney
Department of Computer Sciences
University of Texas at Austin
Austin, Texas 78712
htng@cs.utexas.edu, mooney@cs.utexas.edu

Abstract

Abduction is an important inference process underlying much of human intelligent activities, including text understanding, plan recognition, disease diagnosis, and physical device diagnosis. In this paper, we describe some problems encountered using abduction to understand text, and present some solutions to overcome these problems. The solutions we propose center around the use of a different criterion, called *explanatory coherence*, as the primary measure to evaluate the quality of an explanation. In addition, explanatory coherence plays an important role in the construction of explanations, both in determining the appropriate level of specificity of a preferred explanation, and in guiding the heuristic search to efficiently compute explanations of sufficiently high quality.

1 Introduction

Finding explanations for properties and events is an important aspect of text understanding and of intelligent behavior in general. The philosopher C.S. Peirce defined abduction as the process of finding the best explanation for a set of observations; i.e. inferring cause from effect. The standard formalization of abductive reasoning in artificial intelligence defines an explanation as a set of assumptions which, together with background knowledge, logically entails a set of observations [Charniak and McDermott, 1985].

Natural language understanding has recently been studied in terms of abduction [Charniak, 1986, Hobbs *et al.*, 1988]. Specifically, abduction has been used to solve problems ranging from resolving anaphoric references and syntactic ambiguity [Hobbs *et al.*, 1988] to recognizing characters' plans in a narrative [Charniak, 1986].

*This research is supported by the NASA Ames Research Center under grant NCC-2-429. The first author was also partially supported by a University of Texas MCD fellowship.

We have built an understanding system called ACCEL (Abductive Construction of Causal Explanations for Language) that is capable of constructing deep, causal explanations for natural language text (both narrative and expository text) through the use of abduction. In this paper, we focus on several problems that arise when applying abduction to text understanding. These problems include: (1) the inadequacy of the frequently used simplicity criterion, i.e. "Occam's Razor", in selecting the preferred abductive explanation; (2) the determination of an abductive proof of the appropriate level of specificity; and (3) the computational intractability of abduction. The solutions we propose to these problems center around the use of a different criterion, called *explanatory coherence*, to construct and evaluate abductive explanations.

2 The Basic Abductive Mechanism

A generic abductive inference procedure operates as follows. The background knowledge is encoded in Horn clause axioms. Given a conjunction of positive literals which encodes the input sentences, the abductive inference procedure computes all possible abductive proofs by backward-chaining on the input literals using the Horn clause axioms in the knowledge base, much like Prolog. However, when there is no fact or consequent of a rule in the knowledge base that unifies with a subgoal in the current partial proof, instead of the proof failing, the abductive procedure has the choice of making that subgoal an assumption, if it is consistent to do so. Inconsistency is detected by calling a theorem prover to try to deduce a contradiction. Since consistency checking is undecidable in general, we impose a depth bound in the theorem prover such that it only tries to search for a proof of inconsistency within the depth bound.

Hence, an abductive proof is a proof of the conjunction of the input literals using Horn clause rules and facts in the knowledge base, together with a set of assumptions. An abductive proof is considered an explanation, or an interpretation of the input sentences.

3 Problems and Solutions

The basic abductive mechanism has several problems. These problems and their solutions are described below.

3.1 Occam's Razor Isn't Sharp Enough

Most previous research on abduction, whether applied to plan recognition, language understanding, disease diagnosis, or physical device diagnosis, only uses "Occam's Razor", i.e. the simplicity criterion, as the basis for selecting the best explanation. For instance, in [Charniak, 1986], the best interpretation is one that maximizes $E - A$, where $E =$ the number of explained observations, and $A =$ the number of assumptions made. The work of [Hobbs *et al.*, 1988] also adopts a similar simplicity criterion. In [Reiter, 1987], he proposed a simplicity criterion based on subset-containment minimality. Other related work, though not explicitly utilizing abduction, also relies on some kind of simplicity criterion to select the best explanation. For example, [Kautz and Allen, 1986] explicitly incorporates the assumption of minimizing the number of top level events in deducing the plan that an agent is pursuing.

Though an important factor, the simplicity criterion is not sufficient by itself to select the best explanation. We believe that some notion of explanatory coherence is more important in deciding which explanation is the best. This is especially true in the area of language understanding and plan recognition. Consider the sentences: "John was happy. The exam was easy." The sentences translate into the conjunction of the following literals:[1] `(name j john)`, `(happy j)`, `(exam e)`, `(easy e)`. A knowledge base of axioms relevant to these input literals are:

```
(happy ?x) <- (optimist ?x)
(happy ?x) <- (succeed ?x ?y)
(succeed ?x ?y) <- (exam ?y) (easy ?y)
                   (study ?x ?y) (take ?x ?y)
```

Based on these axioms, there are two possible interpretations of these sentences, as shown in Figure 1a and 1b. Relying on the simplicity metric results in selecting the interpretation that John was happy because he is an optimist, someone who always feels good about life in general (Figure 1b). This is in contrast with our preferred interpretation of the sentence — John was happy because he did well on the easy exam (Figure 1a).[2]

Intuitively, it seems that the first interpretation (Figure 1a) is better because the input observations are connected more "coherently" than in the second interpretation (Figure 1b). We manage to connect "John was happy" with the "easy exam" in the first interpretation, whereas in the second interpretation, they are totally unrelated. This is the intuitive notion of what we mean by *explanatory coherence*. It is clear that "Occam's Razor", i.e. making the minimum number of assumptions, is not the dominant deciding factor here at all. Rather, we select an explanation based on its coherence, i.e. how well the various observations are "tied together" in the explanation.[3]

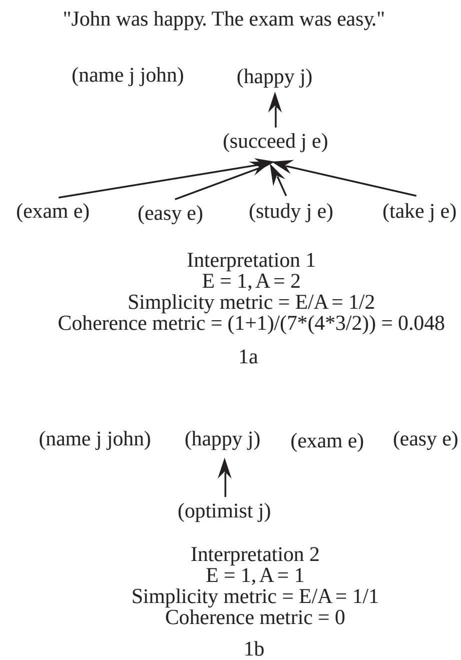

Figure 1: The importance of explanatory coherence

That sentences in a natural language text are connected in a coherent way is reflected in the well known "Grice's conversational maxims" [Grice, 1975], which are principles governing the production of natural language utterances, such as "be relevant", "be informative", etc. Although the notion that natural language text is coherently structured has long been recognized by researchers in natural language processing (see for example [Allen, 1987]), previous work on abduction applying to the tasks of text understanding and plan recognition has not included this criterion in its evaluation of explanations. The use of explanatory coherence here attempts to remedy this problem.

[1] Since we do not focus on the parsing aspect of language understanding, we assume the existence of some appropriate parser that translates the given set of input sentences into a logical representation consisting of a set of literals.

[2] Note that the simplicity criterion of [Reiter, 1987] based on subset minimality also does not work well for this example — it is indifferent towards both interpretations, instead of choosing the preferred one.

[3] [Thagard, 1989] has independently proposed a computational theory of explanatory coherence that applies to the evaluation of scientific theories. However, his theory of explanatory coherence consists of seven principles — symmetry, explanation, analogy, data priority, contradiction, acceptability, and system coherence. Independent criteria like simplicity and connectedness have been collapsed into one measure which he termed "explanatory coherence".

We would like to formulate our coherence metric so as to possess several desirable properties. In particular, explanations with more connections between any pair of observations, as well as those with fewer disjoint partitions are more coherent. Also, a coherence metric with values lying in a unit range 0–1 will facilitate the comparison of explanations.

We have developed a formal characterization of what we mean by explanatory coherence in the form of a coherence metric that satisfies the above mentioned properties. The metric is defined as follows :

$$\mathcal{C} = \frac{\sum_{1 \leq i < j \leq l} N_{i,j}}{N \binom{l}{2}}$$

where
l = the total number of observations;
N = the total number of nodes in the proof graph;
$\binom{l}{2} = l(l-1)/2$;
$N_{i,j}$ = the number of distinct nodes n_k in the proof graph such that there is a (possibly empty) sequence of directed edges from n_k to n_i and a (possibly empty) sequence of directed edges from n_k to n_j, where n_i and n_j are observations.

The numerator of this metric is a measure of the total number of nodes in the explanation connecting pairs of observations. This measure is constructed so that it increases with the number of nodes in the explanation which simultaneously lend support to a given connection. The denominator of the metric simply scales the result according to the size of the explanation so that the final value falls between 0 and 1.

To illustrate the computation of the coherence metric, consider the explanation in Figure 1a. Let n_1 = (name j john), n_2 = (happy j), n_3 = (exam e), n_4 = (easy e), n_5 = (succeed j e), n_6 = (study j e), and n_7 = (take j e). In this explanation, the total number of nodes $N = 7$, and the total number of observations $l = 4$. $N_{2,3} = 1$, since there is exactly one node (namely n_3) such that there is a directed path from n_3 to n_2 and also a directed path from n_3 to n_3 (the trivial empty path). Similarly, $N_{2,4} = 1$. All other $N_{i,j} = 0$. This results in the coherence metric $\mathcal{C} = 0.048$, as shown in Figure 1a.

The coherence metric as defined above can be efficiently computed. Using a standard depth-first search graph algorithm[4], it can be readily shown that $\mathcal{C}$ can be computed in time $O(l \cdot N + e)$, where l = the total number of observations, N = the total number of nodes in the proof graph, and e = the total number of directed edges in the proof graph [Ng and Mooney, 1989]. Based on the coherence metric, ACCEL has successfully selected the best interpretation for a half dozen examples of expository and narrative text that we have tested. (See [Ng and Mooney, 1989] for the list of examples successfully processed by ACCEL.)

[4] We assume here that the proof graph contains no cycles, since circular justification is not considered a good trait of an explanation.

We note here some additional advantages of our coherence metric. One observation is that coherent explanations also tend to be simple explanations. This is because in a coherent explanation, propositions tend to be more tightly connected. This increases the likelihood of assumptions being unified, and leads to a reduction in the number of assumptions made and thus a simpler explanation.

In addition, compared to the simplicity metric, the coherence metric is less vulnerable to changes in the underlying representation of the knowledge base. It is relatively easy to encode the axioms in a knowledge base in a slightly different way so as to change the number of assumptions made in an explanation. However, connections between propositions are less dependent (relatively speaking) on such changes. For example, suppose we change the axioms in the given example slightly so that as long as one takes an easy exam, one will succeed in the exam without having to study for it. Also, suppose one has to be wealthy as well as an optimist to be happy. Given this modified set of axioms, the first interpretation now only requires one assumption, while the second interpretation requires two. So all of a sudden, the first interpretation becomes the simpler explanation of the two. However, the coherence metric of the first interpretation (= (1+1)/(6*(4*3/2)) = 0.056) is still higher than that of the second (which remains at zero).

3.2 Deciding on the Appropriate Level of Specificity of Explanations

Another problem in constructing a good explanation is determining the appropriate level of specificity of an abductive proof. Previous approaches fall into one of three categories : most specific abduction, least specific abduction, and weighted abduction.[5]

In most specific abduction, the assumptions made must be *basic*, i.e. they cannot be "intermediate" assumptions that are themselves provable by assuming some other (more basic) assumptions. This is the approach used in the diagnosis work of [Cox and Pietrzykowski, 1987]. In least specific abduction, the only allowable assumptions are literals in the input observations. [Stickel, 1988] claims that least specific abduction is best suited for natural language interpretation. He argues that what one learns from reading a piece of text is often close to its surface form, and that assuming deeper causes is unwarranted. In weighted abduction [Hobbs *et al.*, 1988], weights (or costs) are assigned to the antecedents of

[5] [Stickel, 1988] describes yet another form of abduction known as predicate specific abduction, which has been used primarily in planning and design-synthesis tasks. In predicate specific abduction, the predicate of any assumption made must be one of a pre-specified set of predicates.

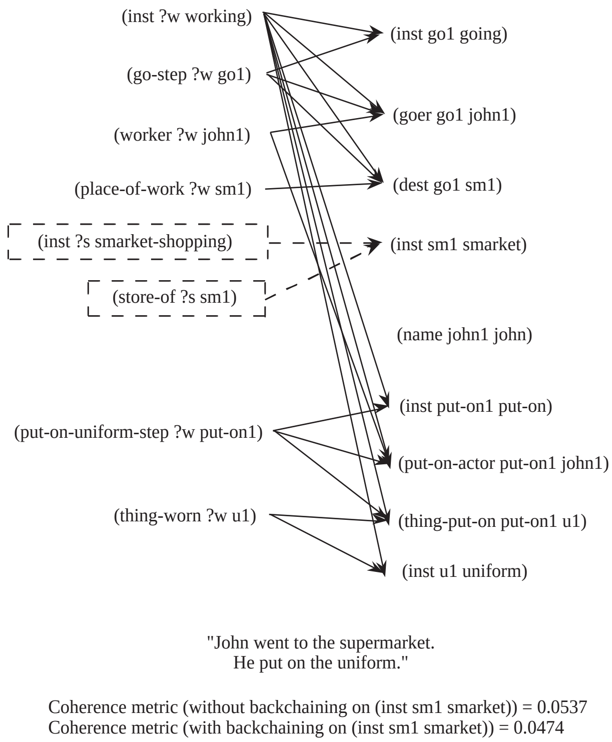

Figure 2: The level of specificity of explanation

backward-chaining rules in order to influence the decision on whether to backchain on a rule. In this case, the best interpretation is the one with assumptions that have the lowest combined total cost.

However, none of these approaches is completely satisfactory. Consider the sentences "John went to the supermarket. He put on the uniform." Both least specific and most specific abduction fail to generate the preferred interpretation in this case, which is that John is working at the supermarket. Figure 2 shows the proof graph of the preferred interpretation of this example (excluding the dashed lines and boxes). (See [Ng and Mooney, 1989] for the details of the relevant axiomatization.)

Note that nowhere in the input sentences is the concept of "working" mentioned at all. It has to be *inferred* by the reader. Since this preferred interpretation includes making the assumptions that there is a working event, that John is the worker of this working event, etc, it is evident that least specific abduction, in which the only allowable assumptions are input literals, is incapable of arriving at this explanation.

On the other hand, most specific abduction will not do the job either. Recall that most specific abduction always prefers backchaining on rules to prove a subgoal if possible rather than making that subgoal an assumption. Thus, applying most specific abduction to this example results in backchaining on the input literal (inst sm1 smarket) to the assumptions (inst ?s smarket-shopping) and (store-of ?s sm1), since in the present knowledge base, this is the *only* backchaining rule with a consequent that unifies with (inst sm1 smarket).[6] That is, we explain the going action, its agent and its destination by assuming that John is working there, and we are also forced to assume, by the requirement of most specific abduction, that there is some supermarket shopping event to explain the supermarket instance! This is because most specific abduction requires that we have an explanation for why John went to the supermarket as opposed to some other workplace. This is clearly undesirable.

However, determining the level of specificity of an explanation based on coherence produces the desired interpretation. That is, we backchain on rules to prove the subgoals in an explanation only if doing so increases its overall coherence, and thus we make assumptions just specific enough to connect the observations. In the current example, backchaining on (inst sm1 smarket) results in a decrease in the coherence metric value, since the total number of nodes in the proof graph increases by two but there is no increase in the number of connections among the input observations. Intuitively, explaining the supermarket instance by assuming a supermarket shopping event is completely unrelated to the rest of the explanation that John is working there. The coherence metric has been successfully used in ACCEL to determine the appropriate level of specificity of explanations, where the desired specificity is one which maximizes coherence.

The weighted abduction of [Hobbs *et al.*, 1988] would presumably arrive at the correct interpretation given the "appropriate" set of weights. However, it is unclear how to characterize the "semantic contribution" of each antecedent in a rule in order to assign the appropriate weights. In contrast, our method does not rely on tweaking such weights, and it produces the preferred interpretation with the desired level of specificity in all of our examples. We believe that allowing arbitrary weights on rules is too much of a burden on the knowledge engineer. It also provides too many degrees of freedom, which can lead to the knowledge engineer "hacking up" arbitrary weights in order to get the system to produce the desired explanation.

3.3 Taming the Intractability Problem

Finding a simplest abductive explanation has been shown to be NP-hard [Reggia *et al.*, 1985, Bylander *et al.*, 1989]. However, an optimal explanation in our system is one with the highest coherence, as opposed to the simplest explanation, and so none of the proofs in the above research applies directly. Nonetheless, we have a proof that finding a maximally coherent explanation that satisfies simple contradiction restrictions in a two-level, propositional abduc-

[6] (inst sm1 smarket) denotes "sm1 is an instance of a supermarket"; (inst ?s smarket-shopping) denotes "?s is an instance of a supermarket shopping event; and (store-of ?s sm1) denotes "the store of the shopping event ?s is sm1".

tion model is NP-hard. As such, the use of heuristic search to explore the vast space of possible solutions seems to be a good strategy to adopt. In fact, we have implemented a form of beam search that has successfully computed the preferred interpretation of all of our examples very efficiently.

We use a beam search algorithm with two beam widths, called *inter-observation beam width* (β_{inter}) and *intra-observation beam width* (β_{intra}), in order to reduce the explored search space. A queue of best explanations is kept by the beam search procedure, forming the "beam" of the beam search. At all times, explanations in the queue are sorted by coherence, where the best explanation is the one with the highest coherence.[7] Only at most β_{inter} number of the best explanations are kept in the queue after completing the processing of each input observation. Within the processing of an input observation, at most β_{intra} number of best explanations are kept in the queue. We have adopted two beam widths instead of one as in a typical beam search algorithm since we have found out empirically that optimal solutions can be computed most efficiently using two beam widths of different sizes.

Figure 3 shows how the quality of the best explanation varies with run time for the supermarket working example with different values of β_{inter} and β_{intra}. We use the ratio of the coherence metric value of an explanation over that of the optimal explanation to represent the quality of an explanation. All the run times reported in this paper are the actual execution times on a Texas Instruments Explorer II Lisp machine.

Each data point in the Figure represents a quality-time pair obtained by using some specific values of β_{inter} and β_{intra}. Each curve connects all the data points with the same β_{inter} but different β_{intra}. Without using any heuristic search (i.e., if a complete search is made), it takes more than 3 hours to compute the optimal solution, while setting $\beta_{inter} = 3$ and $\beta_{intra} = 8$ yields the optimal solution in 0.89 min, which represents a speed up of over 200 times! Also, fixing $\beta_{inter} = 4$ and $\beta_{intra} = 13$, the optimal explanations are computed in about one minute on average for the half dozen examples of expository and narrative text that we have tested.

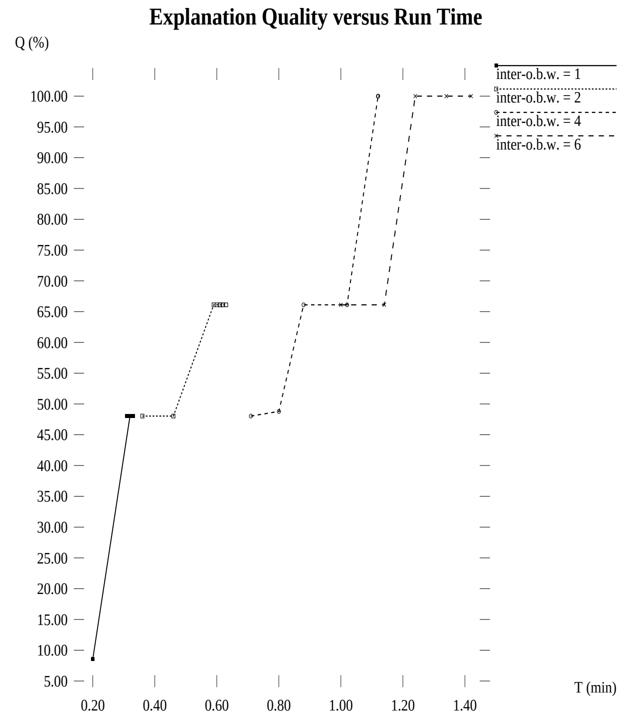

Figure 3: Explanation Quality versus Run Time

4 Related Work

Several research efforts have focused on abduction as an important inference process in plan recognition [Charniak, 1986], natural language understanding [Hobbs *et al.*, 1988], disease diagnosis [Pople, 1973], and physical device diagnosis [Cox and Pietrzykowski, 1987]. In the area of natural language understanding, [Hobbs *et al.*, 1988] describes the use of abduction in solving the four local pragmatics problems of text understanding. This work differs from ours in that unlike their emphasis on mostly linguistics issues like reference resolution and syntactic ambiguity resolution, ACCEL is concerned with constructing deep, causal explanations of the input text. The work of [Charniak, 1986] and [Charniak and Goldman, 1989] are most similar to ours. However, they are primarily concerned with recognizing characters' plans and goals in narrative stories, whereas ACCEL is also capable of constructing causal explanations for expository text. For example, a complete understanding of an encyclopedia text describing animals requires understanding the purpose or reason for the various features of the animals. (See [Ng and Mooney, 1989] for a list of expository text sentences that can be processed by ACCEL.) Also, explanations are evaluated based on their explanatory coherence in our work, as opposed to the simplicity criterion of [Charniak, 1986] and the probability criterion of [Charniak and Goldman, 1989]. Furthermore, the work of [Charniak, 1986] used marker passing to restrict the search for explanations, whereas we used a form of beam search for the efficient construction of explanations.

The Bayesian probabilistic approach to plan recognition and text understanding has been proposed by [Charniak and Goldman, 1989]. Besides the problem of engineering the numerous prior and posterior probabilities of the nodes in a Bayesian network, this approach does not take into account the importance of text coherence. For instance, in the sentences "John got a gun. He entered the grocery store.", one can set

[7] Ties are broken based on the simplicity metric of E/A, where E is the number of observations explained and A is the number of assumptions made.

up reasonable probability estimates such that the conditional probability that John was both hunting and shopping is higher than that of John robbing the store (given the propositions stated in the text). However, selecting an interpretation based solely on the probability of propositions about the situation being described is ignoring the fact that these propositions are adjacent sentences in a natural language text, not just random facts observed in the world. As illustrated by this example, text coherence dominates and results in the reader selecting the more coherent interpretation that John was robbing the store.

5 Future Research

We plan to investigate the efficiency gain which may be brought about by incorporating an ATMS (Assumption-based Truth Maintenance System) into the abductive inference procedure, so as to efficiently keep track of the dependency among the assumptions and propositions of various competing explanations.

Since uncertainty and likelihood information is needed in order to achieve a complete understanding of natural language text, and because a straightforward application of Bayesian probability theory does not give a completely satisfactory solution (as we have illustrated here), an approach that will integrate both the importance of text coherence and likelihood information is an important issue in future research.

6 Conclusion

In summary, we have described some problems encountered using abduction to understand text, and have presented some solutions to overcome these problems. The solutions we propose center around the use of explanatory coherence to evaluate the quality of explanations, to determine the appropriate level of specificity of explanations, and to guide the heuristic search to efficiently compute explanations of sufficiently high quality. These solutions have proven to be very effective on a range of examples in text understanding.

References

[Allen, 1987] James F. Allen. *Natural Language Understanding.* Benjamin/Cummings, Menlo Park, CA, 1987.

[Bylander *et al.*, 1989] Tom Bylander, Dean Allemang, Michael C. Tanner, and John R. Josephson. Some results concerning the computational complexity of abduction. In *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, pages 44–54, Toronto, Ontario, Canada, 1989.

[Charniak and Goldman, 1989] Eugene Charniak and Robert P. Goldman. A semantics for probabilistic quantifier-free first-order languages, with particular application to story understanding. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, 1989.

[Charniak and McDermott, 1985] Eugene Charniak and Drew McDermott. *Introduction to Artificial Intelligence.* Addison Wesley, Reading, MA, 1985.

[Charniak, 1986] Eugene Charniak. A neat theory of marker passing. In *Proceedings of the National Conference on Artificial Intelligence*, pages 584–588, Philadelphia, PA, 1986.

[Cox and Pietrzykowski, 1987] P. T. Cox and T. Pietrzykowski. General diagnosis by abductive inference. In *Proceedings of the 1987 Symposium on Logic Programming*, pages 183–189, 1987.

[Grice, 1975] H. P. Grice. Logic and conversation. In P. Cole and J. Morgan, editors, *Syntax and Semantics 3 : Speech Acts*, pages 41–58. Academic Press, New York, 1975.

[Hobbs *et al.*, 1988] Jerry R. Hobbs, Mark E. Stickel, Paul Martin, and Douglas Edwards. Interpretation as abduction. In *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, pages 95–103, Buffalo, New York, 1988.

[Kautz and Allen, 1986] Henry A. Kautz and James F. Allen. Generalized plan recognition. In *Proceedings of the National Conference on Artificial Intelligence*, pages 32–37, Philadelphia, PA, 1986.

[Ng and Mooney, 1989] Hwee Tou Ng and Raymond J. Mooney. Abductive explanation in text understanding : Some problems and solutions. Technical Report AI89-116, Artificial Intelligence Laboratory, Department of Computer Sciences, The University of Texas at Austin, October 1989.

[Pople, 1973] Harry E. Pople, Jr. On the mechanization of abductive logic. In *Proceedings of the Third International Joint Conference on Artificial Intelligence*, pages 147–152, 1973.

[Reggia *et al.*, 1985] James A. Reggia, Dana S. Nau, and Pearl Y. Wang. A formal model of diagnostic inference. I. problem formulation and decomposition. *Information Sciences*, 37:227–256, 1985.

[Reiter, 1987] Raymond Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–95, 1987.

[Stickel, 1988] Mark E. Stickel. A prolog-like inference system for computing minimum-cost abductive explanations in natural-language interpretation. Technical Note 451, SRI International, September 1988.

[Thagard, 1989] Paul Thagard. Explanatory coherence. *Behavioral and Brain Sciences*, 12(3):435–467, September 1989.