A Critical Evaluation of Commensurable Abduction Models for Semantic Interpretation *

Peter Norvig and **Robert Wilensky**

University of California, Berkeley
Computer Science Division, Evans Hall
Berkeley, CA 94720, USA

Introduction

Language interpretation involves mapping from a string of words to a representation of an interpretation of those words. The problem is to be able to combine evidence from the lexicon, syntax, semantics, and pragmatics to arrive at the best of the many possible interpretations. Given the well-worn sentence "The box is in the pen," syntax may say that "pen" is a noun, while lexical knowledge may say that "pen" most often means writing implement, less often means a fenced enclosure, and very rarely means a female swan. Semantics may say that the object of "in" is often an enclosure, and pragmatics may say that the topic is hiding small boxes of illegal drugs inside aquatic birds. Thus there is evidence for multiple interpretations, and one needs some way to decide between them.

In the past few years, some general approaches to interpretation have been advanced within an abduction framework. Charniak (1986) and Norvig (1987, 1989) are two examples. *Abduction* is a term coined by Pierce (1955) to describe the (unsound) inference rule that concludes A from the observation B and the rule $A \Rightarrow B$, along with the fact that there is no "better" rule explaining B.

In this paper we critically evaluate three recent abductive interpretation models, those of Charniak and Goldman (1989); Hobbs, Stickel, Martin and Edwards (1988); and Ng and Mooney (1990). These three models add the important property of commensurability: all types of evidence are represented in a common currency that can be compared and combined. While commensurability is a desirable property, and there is a clear need for a way to compare alternate explanations, it appears that a single scalar measure is not enough to account for all types of processing. We present other problems for the abductive approach, and some tentative solutions.

Cost Based Commensurability

Hobbs et al. (1988) view interpreting sentences as "providing the best explanation of why the sentences would be true." In this view a given sentence (or an entire text) is translated by an ambiguity-preserving parser into a logical form, L. Each conjunct in the logical form is annotated by a number indicating the cost, $\$C$, of assuming the conjunct to be true. Conjuncts corresponding to "new" information have a low cost of assumability, while those corresponding to "given" information have a higher cost, since to assume them is to fail to find the proper connection to mutual knowledge. Each conjunct must be either assumed or proved, using a rule or series of rules from the knowledge base. Each rule also has cost factors associated with it, and the proper interpretation, I, is the set of propositions with minimal cost that entails L.

As an example, consider again the sentence "The box is in the pen." The cost-annotated logical form (in a simplified notation omitting quantifiers) is:

$$L = box(x)^{\$10} \wedge pen(y)^{\$10} \wedge in(x,y)^{\$3}$$

where $P^{\$x}$ means the final interpretation must either assume P for $\$x$, or prove P, presumably for less. Consider the proof rules:

$$writing\,pen(x)^{.9} \supset pen(x)$$
$$enclosure(x)^{.3} \wedge fenced(x)^{.3} \wedge etc_1(x)^{.3} \supset pen(x)$$
$$female(x)^{.3} \wedge swan(x)^{.6} \supset pen(x)$$
$$enclosure(y)^{.3} \wedge inside(x,y)^{.6} \supset in(x,y)$$

The first rule says that anything that is a writing-pen is also a member of the class 'pen'—things that can be described with the word "pen". The superscripted numbers are preference information: the first rule says that $pen(x)^{\$10}$ can be derived by assuming $writing\,pen(x)^{\$9}$. Predicates of the form $etc_i(x)$, as in the second rule, denote conditions that are stated elsewhere, or, for some natural kind terms, can not be fully enumerated, but can only be assumed. They seem to be related to the abnormal predicates, $ab(x)$ used in circumscription theory (McCarthy 1986).

Below are two interpretations of L. The first just assumes the entire logical form for \$23, while the second applies the rules and shares the $enclosure(y)$ predicate common to one of the definitions of $pen(y)$ and the definition of $in(x,y)$ to arrive at a \$20.80 solution.

$$box(x)^{\$10} \wedge pen(y)^{\$10} \wedge in(x,y)^{\$3}.$$
$$box(x)^{\$10} \wedge enclosure(y)^{\$3} \wedge fenced(y)^{\$3} \wedge etc_1(y)^{\$3} \wedge enclosure(y)^{\$0} \wedge inside(x,y)^{\$1.8}$$

*Sponsored by the Defense Advanced Research Projects Agency (DoD), Arpa Order No. 4871, monitored by Space and Naval Warfare Systems Command under Contract N00039-88-C-0292. This paper benefitted from discussions with Michael Braverman, Dan Jurafsky, Nigel Ward, Dekai Wu, and other members of the BAIR seminar.

The second $enclosure(y)$ gets a cost of \$0 because it has already been assumed. Let me stress that the details here are ours, and the authors may have a different treatment of this example. For example, they do not discuss lexical ambiguity, although we believe we have been faithful to the sense of their proposal.

This approach has several problems, as we see it:

(1) A single number is being used for two separate measures: the cost of the assumptions and the quality of the explanation. Hobbs et al. hint at this when they discuss the "informativeness-correctness tradeoff." Consider their example "lube-oil alarm," which gets translated as:

$$lubeoil(o)^{\$5} \wedge alarm(a)^{\$5} \wedge nn(o,a)^{\$20}$$

where nn means noun-noun compound. It is given a high cost, \$20, because failing to find the relation means failing to fully understand the referent. Intuitively this motivation is valid. However, the nn should have a very *low* cost of assumption, because there is very strong evidence for it—the juxtaposition of two nouns in the input—so there is little doubt that nn holds. Thus we see nn should have two numbers associated with it: a low cost of assumption, and a low quality of explanation. It should not be surprising to see that two numbers are needed to search for an explanation: even in A^* search one needs both a cost function, g, and a heuristic function h'.

The low quality of explanation is often the sign of a need to search for a better explanation, but the need depends on the task at hand. To diagnose a failure in the compressor, it is useful to know that a "lube-oil alarm" is an alarm that sounds when the lube-oil pressure is low, and not, say, and alarm made out of lube-oil. However, if the input was "Get me a box of lube-oil alarms from the warehouse," then it may not be necessary to further explain the nn relation.[1] Mayfield (1989) characterizes a good explanation as being applicable to the needs of the explanation's user, grounded in what is already known, and completely accounting for the input.

To put it another way, consider the situation where a magician pulls a rabbit out of his hat. One possible explanation is that the rabbit magically appeared in the hat. This explanation is of very high quality—it perfectly explains the situation—but it has a prohibitive assumption cost. An alternate explanation is that the magician somehow used slight of hand to insert the rabbit in the hat when the audience was distracted. This is of fairly low quality—it fails to completely specify the situation—but it has a much lower assumption cost. Whether this is a sufficient explanation depends on the task. For a casual observer it may will do, but for a rival magician trying to steal the trick, a better explanation is needed.

(2) Translating, say, "the pen" as $pen(y)^{\$10}$ conflates two issues: the final interpretation must find a referent, y, and it must also disambiguate "pen". It is true that definite noun phrases are often used to introduce new information, and thus must be assumed, but an interpretation that does not disambiguate "pen" is not just making an assumption—rather it is failing altogether. One could accommodate this problem by writing disambiguation rules where the sum of the left-hand-side components is less than 1. Thus, the system will always prefer to find some interpretation for "pen", rather than leaving it ambiguous. In the case of vagueness rather than ambiguity, one would probably want the left-hand-side to total greater than 1. For example, in "He saw her duck", the word "duck" is ambiguous between a water fowl and a downward movement, and any candidate solution should be forced to decide between the two meanings. In contrast, "he" is vague between a boy and a man, but it is not necessary for a valid interpretation to make this choice. We could model this with the rules:

$$\begin{array}{l} duck_{fowl}(x)^{.9} \supset duck(x) \\ duck_{move}(x)^{.9} \supset duck(x) \\ male(x)^{.9} \wedge adult(x)^{.2} \supset he(x) \\ male(x)^{.9} \wedge child(x)^{.2} \supset he(x) \end{array}$$

However, this alone is not enough. Consider the sentence "The pen is in the box." By the rules above (and assuming a box is defined as an enclosure) we could derive three interpretations, where either a writing implement, a swan, or a fenced enclosure is inside a box. All three would get a cost of \$20.8. To choose among these three, we would have to add knowledge about the likelihood of these three things being in boxes, or add knowledge about the relative frequencies of the three senses of "pen". For example, we could change the numbers as follows:

$$\begin{array}{l} writing\,pen(x)^{.9} \supset pen(x) \\ enclosure(x)^{.31} \wedge fenced(x)^{.31} \wedge etc_1(x)^{.31} \\ \quad \supset pen(x) \\ female(x)^{.4} \wedge swan(x)^{.8} \supset pen(x) \end{array}$$

This has the effect of making the writing implement sense slightly more likely than the fenced enclosure sense, and much more likely than the female swan sense. These rules maintain the desirable property of commensurability, but the numbers are now even more overloaded. Hobbs et al. already are giving the numbers responsibility for both "probabilities" and "semantic relatedness", and now we have shown they must account for word frequency information, and both the cost of assumptions and the quality of the explanation, the two measures needed to control search. As our previous criticisms have shown, a single number cannot represent even the cost and quality of an explanation, much less these additional factors.

Also note that to constrain search, it is important to consider bottom-up clues, as in (Charniak 1986) and (Norvig 1987). It would be a mistake to use the rules given here in a strictly top-down manner, just because they are reminiscent of Prolog rules.

(3) There is no notion of a "good" or "bad" interpretation, except as an epiphenomenon of the interpretation rules. In the "pen" example, the difference between failing completely to understand "pen" and properly disambiguating it to fenced-enclosure is less than 10% of the total cost. The numbers in the rules could be changed to increase this difference, but it would still be a quantitative rather than qualitative difference. The problem is

[1] Translating "lube-oil alarm" as $(\exists o) lubeoil(o)$ is suspect; in the case of an alarm still in the box, there is not yet any particular oil for which it is the alarm.

that there are at least three reasons why we might want to maintain ambiguity: because we are unsure of the cause of an event, because it is so mundane as to not need an explanation, and because it is so unbelievable that there is no explanation. This theory does not distinguish these cases. The theory has no provision for saying "I don't understand–the only interpretation I can find is a faulty one," and then looking harder for a better interpretation.

(4) There is no way to enforce a penalty worse than the cost of an assumption. Consider the sentence "Mary said she had killed herself." The logical form is something like:

$$say(Mary, x)^{\$3} \wedge x = kill(Mary, Mary)^{\$3}.$$

Thus, for $6 we can just assume the logical form, without noticing the inherent contradiction. Now let's consider some rules. We've collapsed most of the interesting parts of these rules into etc predicates, leaving just the parts relevant to the contradiction:

$$alive(p)^{.1} \wedge etc_2(p, x)^{.9} \supset say(p, x)$$
$$\neg alive(p)^{.5} \wedge etc_3(m, p)^{.5} \supset kill(m, p)$$

We've ignored time here, but the intent is that the alive predicate is concerned with the time interval or situation after the killing, including the time of the saying. Now, an alternative interpretation of L is:

$$alive(Mary)^{\$.3} \wedge \neg alive(Mary)^{\$1.5}$$
$$\wedge\, etc_2(Mary, x)^{\$2.7} \wedge etc_3(Mary, Mary)^{\$1.5}$$

Presumably there should be some penalty (finite or infinite) for deriving a contradiction, so this interpretation will total more than $6. The problem is there is no way to propagate this contradiction back up to the first interpretation, where we just assumed both clauses. We would like to penalize that interpretation, too, so that it costs more than $6, but there is no way to do so.

A solution to this problem is to legislate that rather than finding a solution to the logical form of a sentence, L, the hearer must find a solution to the larger set of propositions, L', where L' is derived from L by some process of direct, "obvious" inference. We do not want the full deductive closure from L, of course, but we want to allow for some amount of automatic forward chaining from the input.

(5) We would like to be able to go on and find alternative explanations, perhaps one where Mary is speaking from the afterworld, or she is lying, or the speaker is lying. One could imagine rules for truthful and untruthful saying, and such rules could be applied to Mary's speech act. However, since the goal of the interpretation process is "providing the best explanation of why the sentences would be true," it does not seem that we could use the rules to consider the possibility of the speaker being untruthful. The truth of the text is assumed by the model, and the speaker is not modeled.

Probability Based Commensurability

Charniak and Goldman (1988) started out with a model very similar to Hobbs et al., but became concerned with the lack of theoretical grounding for the numbers in rules, much as we were. Charniak and Goldman (1989a, 1989b) switched to a system based strictly on probabilities in the world, combined by Bayesian probability theory. Although this solves some problems, other problems remain, and some new ones are introduced. For example:

(1) The approach in (1989a) is based on "events and objects in the real world". As the authors point out, it cannot deal with texts involving modal verbs, nor can it deal with speech acts by characters, or texts where the speaker is uncooperative. So problem (4) above remains.

(2) Because the probabilities are based on events in the real world, the basic system often failed to find stories as coherent as they should be. For example, the text:

Jack got a rope. He killed himself.

suggests suicide by hanging when interpreted as a text, but when interpreted as a partial report of events in the world, that interpretation is less compelling. (After all, the killing may have taken place years after the getting.) It is only when the two events are taken as a part of a coherent text that we assume they are related, temporally and causally. In Charniak and Goldman (1989a), the coherence of stories is explained by a (probabilistic) assumption of spatio-temporal locality between events mentioned in adjacent sentences in the text. Thus the story would be treated roughly as if it were:

Jack got a rope. Soon after, nearby, a male was found to have killed himself.

The Bayesian networks compute a probability of hanging of .3; this seems about right for the later story, but too low for the original version.

Perhaps anticipating some of these problems, Charniak and Goldman (1989b) introduce an alternate approach involving a parameter, E, which denotes the probability that two arbitrary things are the same. They claim that in stories this parameter should be set higher than in real life, and that this will lead, for example, to a high probability for the interpretation where the rope that Jack got is the one he used for hanging. But E does a poor job of capturing the notion of coherence. Consider:

John picked an integer from one to ten. Mary did so too.

Here the probability that they picked the same number should be .1, regardless of whether we are observing real life or reading a story, and regardless of the value of E.

Charniak and Goldman (1989b) go on to propose a theory of "mention" rather than a theory of coincidence, but they do not develop this alternative.

(3) It seems that for many inferences, frequency in the world does not play an important role at all. Consider the text:

Jack wanted to tie a mattress on top of his car. He also felt like killing himself. He got some rope.

Now, the probability of getting a rope to hang oneself given suicidal feelings must be quite low, maybe .001, while the probability of getting a rope for tying given a desire to secure a mattress is much higher, maybe .5. Thus the Charniak-Goldman model would strongly prefer the latter interpretation. With the "mention" theory, it would like both interpretations. Yet a sample of informants mostly found the text confusing–they reported finding both interpretations, and were unable to choose between them. It would be useful to find a better characterization of when frequencies in the world are useful, and when they appear to be ignored in favor of some more discrete notion of "reasonable connection."

Problems With Both Models

Neither model is completely explicit on how the final explanation is constructed, or on what to do with the final explanation. In a sense, Hobbs et al.'s system is like a justification-based truth-maintenance system that searches for a single consistent state, possibly exploring other higher-cost states along the way. Charniak and Goldman's system is like an assumption-based truth-maintenance system (ATMS) that keeps track of all possible worlds in one grand model, but needs a separate interpretation process to extract consistent solutions. Thus, the system does not really do interpretation to the level that could lead to decisions. Rather, it provides evidence upon which decisions can be based.

Both approaches are problematic. Imagine the situation where a hearer is driving a car, and is about to enter an intersection when a traffic officer says "don't - stop". The hearer derives two possible interpretations, one corresponding to "Don't stop." and the other corresponding to "Don't. Stop." Hobbs et al.'s system would assign costs and chose the one with the lower cost, no matter how slight the difference. A more prudent course of action might be to recognize the ambiguity, and seek more information to decide what was intended. Charniak and Goldman's system would assign probabilities to each proposition, but would offer no assistance as to what to do. However, if the model were extended from Bayesian networks to influence diagrams, then a decision could be made, and it would also be possible to direct search to the important parts of the network.

Deliberate ambiguity is also a problematic area. In a pun, for example, the speaker intends that the hearer recover two distinct interpretations. Such subtlety would be lost on the models discussed here. This issue is discussed in more depth in Norvig (1988).

A number of arguments show that strict maximization of probability (or minimization of cost) is a bad idea. First, as we have seen, we must sometimes admit that an input is truly ambiguous (intentionally or unintentionally).

Second, there is the problem of computational complexity. Algorithms that guarantee a maximal solution take exponential time for the models discussed here. Thus, a large-scale system will be forced to make some sort of approximation, using a less costly algorithm. This is particularly true because we desire an on-line system—one that computes a partial solution after each word is read, and updates the solution in a bounded period of time.

Third, communication by language has the property that "the speaker is always right". In chess, if I play optimally and my opponent plays sub-optimally, I win. But in language understanding, if I abduce the "optimal" interpretation when the speaker had something else in mind, then we have failed to communicate, and I in effect lose. Put another way, there is a clear "evolutionary" advantage for optimal chess strategies, but once language has evolved to the point where communication is possible, there is no point for a hearer to try to change his interpretation strategy to derive what an optimal speaker would have uttered to an optimal hearer–because there are no such optimal speakers. Indeed, there is an advantage for communication strategies that can be computed quickly, allowing the participants to spend time on other tasks. By the second point above, such a strategy must be sub-optimal.

Earlier we said that Charniak and Goldman (1989b) introduced the parameter E to account for the coherence of stories. But they also provide a brief sketch of another account, one where, in addition to deriving probabilities of events in the world, we also consider the probability that the speaker would mention a particular entity at all. Such a theory, if worked out, could account for the difficulty in processing speech acts that we have shown both models suffer from.

However, a theory of "mention" alone is not enough. We also need theories of representing, intending, believing, directly implying, predicting, and acting. The chain of reasoning and acting includes at least the following:

H attends to utterance U by speaker S
H infers "S said U to H"
H infers "L represents U"
H infers "L directly implies L'"
H infers "S intended H to believe S believes L'"
H infers "S intended H to believe L'"
H believes a portion of L' compatible with H's beliefs
H forms predictions about S's future speech acts
H acts accordingly

This still only covers the case of successful, cooperative communication, and it leaves out some steps. A successful model should be able to deal with all these rules, when necessary. However, the successful model should also be able to quickly bypass the rules in the default case. We believe that the coherence of stories stems primarily from the speaker presenting evidence to the hearer in a fashion that will lead the hearer to focus his attention on the evidence, and thereby derive the inferences intended by the speaker. Communication is possible because it consists primarily of building a single shared explanation. It is only in unusual cases where there are multiple possibilities that must be weighed against each other and carried forth.

Both models seem to have difficulty distinguishing ambiguity from multiple explanations. This makes a difference in cases like the following:

John was wondering about lunch when it started to rain. He ran into a restaurant.

Here there are two reasons why John would enter the restaurant–to satisfy hunger and to avoid the rain. In other words there are two explanations, say, $A \supset R$ and $B \supset R$, and we would like to combine them to yield $A \wedge B \supset R$. As we understand it, Hobbs et al. appear to use "exclusive or" in all cases, so they would not find this explanation. Charniak and Goldman allow competing explanations to be joined by an "or" node, but require competing lexical senses to be joined by "exclusive or" nodes. So they would find $A \vee B \supset R$. In other words, they would find both explanations probable, which is not quite the same thing as finding the conjunction probable. Now consider:

He's a real sweetheart.

This has a straight and an ironic reading: $sweetheart(x)$ and $\neg sweetheart(x)$. The disjunction is a tautology and the conjunction is a contradiction, so in this case the Hobbs approach of keeping the alternatives separate seems better than allowing their disjunction. Finally, consider:

Mary was herding water fowl while dodging hostile gunfire. John saw her duck.

Here we do not want to combine the two interpretations into a single interpretation. If we amend a model to allow multiple explanations, we must be careful that we don't go too far.

Coherence Based Commensurability

Much of the criticism above stems from the lack of a model of textual coherence. Intuitively, an explanation that makes the text cohere (by finding propositions that relate pieces of the text to other pieces) will be preferred to an equi-probable explanation that is not as coherent. Ng and Mooney (1990) attempt to formally define a measure of coherence. In their model the logical form of the input text is taken as a set of propositions that they call *observations*. The interpretation is the conjunction of these observations with another set of propositions called *assumptions,* where each assumption is introduced as a node in the proof graph of the observations. The most coherent interpretation is the one that maximizes the number of nodes that support (directly or indirectly) two observations. Nodes are counted multiple times if they support multiple pairs of observations. The coherence metric is normalized to the range 0-1 by dividing the count by the total number of possible connections.

Ng and Mooney give as an example the text "John was happy. The exam was easy." They propose two interpretations. The first relies on the (assumed) rule that optimists are happy. Thus, by making the single assumption that John is an optimist, one can explain the fact that he is happy. This takes one assumption to make one explanation, but it gets a coherence score of 0, because no pairs of observations are tied together. The other interpretation makes use of the (assumed) rules that succeeding on something makes one happy, and that studying for and taking an easy exam leads to succeeding on the exam. This makes two assumptions (studying and taking the exam) and again explains only one input (that John was happy), so it fares worse on the ration of explanations to assumptions. However, it has a higher coherence score, because it ties together the exam and the exam being easy with John being happy. Therefore, they conclude that coherence is more important than other metrics considered here.

We tend to agree with this conclusion. While the other models may be able to duplicate this particular example, consider "John was happy. The winning lottery number was 174625." Here, we assert, the best interpretation is that John has number 174625 and has won the lottery. However, a probability-based model would have to put a very low probability on John having that particular number, and would prefer some other explanation for his happiness.

However, we do not feel that coherence should be used alone without some notion of relative costs or probabilities, nor do we feel that Ng and Mooney have accurately captured the notion of coherence. There are several problems with their metric.

First, recall that if an assumption *A* supports two observations, it adds to the coherence metric, and that any further assumption that supports *A* also adds to the metric. Thus, Ng and Mooney prefer explanations with arbitrarily long backward chaining, no matter how improbable the connection. They have no way to cut off the explanation at an appropriate level of detail.

Second, they do not attempt to choose between alternatives. For example, in the "John was wondering about lunch when it started to rain. He ran into a restaurant." example, they would accept both explanations. Since two explanations are always more coherent than one, they can never reject a coherent explanation in favor of a better one.

Third, they have no way to guide the search for coherence to the right places. Suppose the input consists of three observations, and that the first two have been tied together. It would be prudent to try to tie the third observations in with the first two, but the coherence metric gives just as many points for finding another explanation for the first two as for connecting the third to one of the others. To their credit, Ng and Mooney discuss a heuristic search mechanism that is guided by coherence. We feel this is the right idea, but not with their exact coherence metric.

Finally, while coherence is important, it is not the only criteria by which texts are constructed. Consider "John was 22 years old. He lives in California." This is perfectly acceptable as the setting for a story to come. We would not want to try to explain this passage by assuming, for example, that John thought that California was a good place for 22-year-olds to live, and thus he moved there.

Conclusions

Abduction is a good model for language interpretation, and commensurability is a vital component of an abduction system. But the models discussed here have serious limitations, due to technical problems, and due to a failure to embrace language as a complex activity, involving actions, goals, beliefs, inferences, predictions, and the like. We don't believe that knowledge of probability in the world, plus a few general principles (such as E) can lead to a viable theory of language use. This "complicated" side of language has been studied in depth for over a decade (a list very similar to our chain of reasoning and acting appears in Morgan (1978)), so our task is clear: to marry these pre-theoretic "complicated" notions with the formal apparatus of commensurable abductive interpretation schemes.

References

Charniak, E. A neat theory of marker passing, *AAAI-86*.

Charniak, E. and Goldman, R. (1988) A logic for semantic interpretation, *Proc. of the 26th Meeting of the ACL.*

Charniak, E. and Goldman, R. (1989a) A semantics for probabilistic quantifier-free first-order languages, with particular application to story understanding, *IJCAI-89.*

Charniak, E. and Goldman, R. (1989b) Plan recognition in stories and in life, *Uncertainty Workshop, IJCAI-89.*

Hobbs, J. R., Stickel, M., Martin, P. and Edwards, D. (1988) Interpretation as abduction, *Proc. of the 26th Meeting of the ACL.*

Mayfield, J. M. (1989) Goal analysis: Plan recognition in dialogue systems, *Univ. of Cal. Berkeley EECS Dept.* Report No. UCB/CSD 89/521.

McCarthy, J. (1986) Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence,* 26(3).

Morgan, J. L. (1978) Toward a rational model of discourse comprehension. *Theoretical Issues in Natural Language Processing*.

Ng, H. T. and Mooney, R. J. (1990) The role of coherence in constructing and evaluating abductive explanations. Proceedings of the AAAI Spring Symposium on Automated Abduction.

Norvig, P. (1987) A Unified Theory of Inference for Text Understanding. *Univ. of Cal. Berkeley EECS Dept.* Report No. UCB/CSD 87/339.

Norvig, P. (1988) Multiple simultaneous interpretations of ambiguous sentences. *Proc. of the 10th Annual Conference of the Cognitive Science Society.*

Norvig, P. (1989) Marker passing as a Weak Method for Text Inferencing. *Cognitive Science,* **13**, 4, 569-620.

Pierce, C. S. (1955) *Abduction and Induction.* Dover, NY.