# Discourse versus Probability in the Theory of Natural Language Interpretation

**Robert Wilensky***
Division of Computer Science
571 Evans Hall
University of California, Berkeley
Berkeley, CA 94720
wilensky@cs.berkeley.edu

## Abstract

Recent probabilistic models of natural language interpretation pay insufficient attention to what a speaker might say, mistakenly focusing instead on what is probable in the world. As a result, these models produce incorrect results, require data that is not realistically obtainable, and entail the solution to problems that intractable but irrelevant to the task of natural language interpretation. In contrast, a proper understanding of the nature of discourse allows us to so simplify any probabilistic elements so as to trivialize their role. The challenger for natural language interpretation, then, is not to perfect a probabilistic account, but to develop a plausible account of discourse. Such an account needs to focus on the nature of the contents speakers convey, and on cooperative principles of communication.

## 1 Introduction

Recently, there have been a number of efforts to cast natural language understanding in a probabilistic framework. The argument that there is a probabilistic element in natural language interpretation is essentially the following: Most inferences or interpretation decisions cannot be made based on logical criteria alone. Rather, evidence of varying strengths needs to be combined, and probability theory offers a principled basis for doing so. For example, consider the following story[1]:

(1) John got his suitcase. He went to the airport.

The interpretation that John is a terrorist intent on blowing up an airplane, say, is not favored by human readers, and is probably not even considered by them, although structurally it is identical to the preferred interpretation in which John is an airline passenger. A probabilistic approach, however, would offer us some hope of coming to the correct conclusion.

Hobbs et al. (1988) and Goldman and Charniak (1988) offer models in which evidence can be accorded numerical values and hence combined so as to determine the best explanation. The models address different problems: plan recognition in the case of Goldman and Charniak; "local pragmatics" (specifically, metonymy, reference and syntactic ambiguity) in the case of Hobbs et al. Both models can be thought of as probabilistic, with Goldman and Charniak explicitly taking this stance, and Shimony and Charniak (1990) arguing that Hobbs et al. can be interpreted this way.

However, as formulated, these probabilistic approaches have significant, perhaps fatal shortcomings (some of which are broadly alluded to in Norvig and Wilensky, 1990). In a nutshell, the problem is as follows:

> The models pay insufficient attention to what a speaker might say; they focus instead on what is probable in the world. As a result, they produce incorrect results, require data that is not realistically obtainable, and require the solution to intractable problems whose solution is unnecessary for the task of natural language interpretation.

In this paper, we attempt to state this criticism more precisely, and then offer an alternative approach. This approach still has a role for probability to play, but it is not a very elaborate one – a proper understanding of the nature of discourse allows us to so simplify any probabilistic elements so as to trivialize them. In effect, our moral is that most of the action in text understanding is in understanding the nature of discourse; discourse, in turn, is primarily about the kinds of contents people tend to communicate.

*This work was supported by Nation Science Foundation grant IRI-9123336.

[1] from Charniak, personal communication

## 2 General Considerations

We begin by a general motivation of the importance of discourse. Then we look at precisely where particular probabilistic theories err.

Hobbs et al. characterize interpretation as

> "the process of providing the best explanation of why the sentences would be true."

Charniak and Goldman refer to this definition, but restate it as

> "reasoning from the text to the intentions of the language user and thence to the thing described."

Formulating the problem as addressing the question of "how an utterance can be true" requires a theory of what a speaker is likely to say. While both Goldman and Charniak and Hobbs et al. allude to this fact, the theories offered are basically concerned with determining the least costly or most probable assumptions one can make *so that a sentence follows from a knowledge base.* This is problematic in a number of ways. Consider the following utterances:

(2) The butcher has kidneys today.

(3) The man entering the room had one green eye.

(4) John doesn't put his money in a bank.

If we are to explain how the first sentence can be *true*, given our beliefs about the world, the interpretation in which the kidneys are inalienably part of the butcher is preferable to the one in which these are his merchandise: The first interpretation is a near certainty; the other has some rather lower empirical value. Thus, maximizing the probability of a proposition being true yields the wrong interpretation of the utterance.

Similarly, if the second sentence above were an observation, say, because the man was turned so that we could see only one of his eyes, we would conclude that the man's eyes were both green. However, if we were told this sentence, we would assume just the opposite. Both interpretations make use of the same probabilistic fact, namely, that people's eyes are usually the same color.

An analogous argument pertains to negation and disambiguation. In the third sentence above, it is presumably more likely that John doesn't put his money in the side of a river than that he doesn't put it in a financial institution. Again, maximizing the probability of an interpretation being true in the world yields a likely event – John not putting money in the side of a river – but a very unlikely interpretation.

Additionally, a "probability in the world" metric will always favor an ambiguous or vague interpretation over a more specific one, since the ambiguous or vague interpretation is always more probable. As Hobbs et al. states "The less specific they [the assumptions] are, the more likely they are to be correct." For example, consider the following sentences:

(5) **a.** Jan cut the grass.

**b.** Jan cut the cake.

**c.** Jan cut the ribbon.

Presumably, an utterance of the first sentence would normally mean about the same thing as "Jan mowed the law"; the second describes a slicing event, and the third, yet another kind of cutting. Moreover, determining the kind of cutting being described is a condition of understanding each utterance. However, the probability of any kind of cutting event, given a cutting event, is always greater than that of any specific kind of cutting event. So the interpretation of these sentences that is most likely to be true is the one specifying an undifferentiated cutting event; but this is not the interpretation natural language users assign these utterances.

These are familiar Gricean objections (Grice, 1975). The models ask the question: "Given that an utterance represents facts about the world, what else is likely to be true?" But this question discounts the nature of natural language communication: In the case of the butcher and his kidneys, the more-probable-in-the-world answer is uninformative; but speakers are normally informative. Similarly, we would be at a loss to explain why a speaker informed us that a man had one green eye if he knew that the man had two. In the case of ambiguous or vague terms, it is a fact about the nature of natural language that these terms can be used to refer to a subtype, not to a disjunction of subtypes or a very abstract category. In the negation examples, the communicative function is not simply to assert a non-event, but to deny a scenario that the hearer is assumed to consider.

## 3 What is Reasonable to Say Versus What is True in the World

In the remainder of this paper, we focus on Charniak and Goldman (1989) and Goldman (1990) ("Charniak and Goldman" henceforth), since we are concerned primarily about the same problem as they

are, i.e., text interpretation in the large. However, some of the criticisms also apply in different ways to Hobbs et al. as well. (We will self-servingly refrain, though, from a general critical of probabilistic interpretation in the small as espoused, for example, by Wu (1990, 1992).)

Charniak and Goldman start with a Bayesian network model of plan recognition. Elements of a text are treated as evidence for events, etc. They construct a Bayesian network for the text, using marker passing to delimit the size of the network constructed. For example, the events described in the text

(6) Jack got a rope. He killed himself.

are evidentially related to possible events like Jack's hanging himself. Simply evaluating the resulting Bayesian network should produce a compelling probability that a hanging transpired.

However, and Charniak and Goldman point out, in the naive version, the probability that Jack hanged himself is essentially unaffected by the evidence from the text. The problem is that the network constructed contains nodes representing, for example, the possibility that the rope mentioned in the story was same object as the one used in a hypothetical hanging event. These nodes are conditioned by the probability that any two objects in the universe of the task are the same object. This is a very small number indeed, so small as to relegate the observation of getting a rope to be irrelevant.

The nature of the problem is best stated this way: The network treats the text as describing *completely unrelated events.* That is, it is just as if we learned, somehow, that, at some point in his life, Jack once got a rope. Then, possibly many years later, Jack killed himself. Under these circumstances, there is indeed little reason to believe that Jack hung himself.

Charniak and Goldman address this problem by making an assumption of spatial and temporal locality ("STL"). The STL assumption raises the probability that any two entities are the same because there are assumed to be many fewer entities in the problem domain. With a restrictive enough assumption, the priors are much higher, and the existing of getting a rope has a non-negligible

While this assumption is necessary for the probabilistic model of plan recognition to process feasible results, as Charniak and Goldman point out, one must still account for the difference between natural language understanding and abduction "in the world". For example, they consider the following texts:

(7) **a.** There was a rope in the closet. Jack killed himself.

**b.** Jack got a rope. He killed himself.

Charniak and Goldman assume that in (b), hanging is a viable inference both in the story and in real life, i.e., just from knowing the facts, whereas in (a), hanging is a viable inference just in the story, but would not be warranted in the real world.

To account for this difference with a single model of plan recognition, Charniak and Goldman introduce a parameter reflecting the a priori bias against any two things being the same thing. This bias is to be large for the world, but smaller for texts. Thus, in the world, the chance that the object used as a rope in any putative hanging and the rope mentioned in the story are the same is relatively small; hence the likelihood of hanging is increased only a little by knowledge of a rope. In a text, it is more likely that the two are the same, so the probability of hanging is significantly increased by knowledge of the rope.

Unfortunately, every one of these assumptions appears to be erroneous. Consider first basing a theory of discourse on the likelihood of the identity of objects. Consider the following example:

(8) Jan randomly picked a number between 1 and 10. So did Lynn.

The assumption that it is more likely that two entities in a story are the same than are two entities in the world would force us to infer that the chance that Jan and Lynn picked the same number is more likely than 1 out of 100. But this inference is not warranted.

Second, the assumption of a different likelihood of the identity of objects between stories and the real world does *not* explain the difference between the (a) and (b) stories. The assumption of STL presumably makes hanging more likely in both the (a) and (b) "real world" cases, and the assumption of even greater likelihood of object identity in stories makes hanging even more likely in both textual cases. To put it another way, the degree that hanging in (a) is considered less likely than (b) in the world, it will be considered less likely as a story. But hanging is presumably just as correct an inference in both the (a) and (b) stories. (Charniak and Goldman do not give computed probabilities for all these cases, so it is difficult to determine what actually happens in their model.)

STL is a *terrible* assumption anyway. It predicts, erroneously, for example, that, as the number of ropes available to Jack increase, it is *less* likely that he hung himself with the rope he got. This approach would predict that, in

(9) Jack worked in a rope factory. He got a rope. He killed himself.

we infer that it is relatively unlikely that Jack hung himself with the rope he acquired.

Also, consider the following stories:

(10) Dick wanted to become president. He decided to run for governor first. He was elected. He started planning his presidential campaign.

(11) Asiatic people lived across the Bering Strait. They crossed it and settled all of the Americas.

The crucial observation here is that the STL assumption (or its lack thereof) is *determined* by the scenarios inferred, rather than serves as a condition for inferring them. I.e., this assumption is not something generally true about stories or texts or plan abduction tasks; rather, it is a specific fact about the particular contents of particular situations. Thus, in the suicide stories, it is our knowledge of suicide, and something about the nature of narratives, that lets us infer that the getting of the rope probably happened quite soon before the suicide; in contrast, our knowledge of elected offices and migrations (plus the same narrative conventions) is responsible for the very different inferences in the subsequent examples.

Moreover, determining the speaker's intention seems far easier than obtaining or computing probabilities about the world. Readers appear to have no doubt that the author of the "Jack got a rope" texts intended to convey a hanging; but they would have grave difficulties computing the probabilities in the world on which this conclusion is ostensibly based. Thus, it seems reasonable to question whether requiring such a computation is beneficial way to approach the task.

Indeed, such conclusions about the world seem largely irrelevant to the conclusions one reaches in text understanding. For example, consider the story:

(12) Jack wanted to kill himself. He got a guillotine.

Jack's plan is presumably new to the understander's experience; its prior must be very low indeed. However, in all these cases, the interpretation intended by the speaker is compelling. That is, assuming that the text is felicitous, the probability that the speaker intended the interpretation is approximately 1.

# 4 Text Understanding is Not a Modified form of Understanding the World

We have argued that a theory of text understanding is unlikely to be achieved by a minor adjustment to a probabilistic model of understanding the real world. But why should this be the case? After all, it seems intuitively compelling that a theory of text understanding should be grounded in a theory of understanding the real world.

But text understanding and understanding the world are different in at least three important ways:

- Abstractiveness: The information available by (non-linguistic) observation is concrete, and that required inferences abstractive, while the information conveyed by language is abstract, and the required inferences specificational.
- Relevance: The components of a (well-formed) text are relevant, whereas observations about the world need not be.
- Textual Entities: The entities conveyed by the text may not have any correspondence to entities "in the world".

Note how just the assumption of relevance trivializes the computation Charniak and Goldman are performing. The assumption of coherence lets us cast the plan recognition problem they are attempting to solve as the following problem: Given some observations $E_i$, determine the following probability for all $H$'s containing all the $E_i$:

$$P(H|E1, E2, \ldots, En)$$

We can then infer an $H$ if there is one with high enough probability.

We used the relevance hypothesis to assume only that there will be a single entity that "contains" all the observations. (We will be a bit vague for the time being about the notion of containment that we are alluding to here. Intuitively, for plan recognition-related problems, $H$ would be a plan whose subparts include the $E_i$. Note that some such notion is required in probabilistic systems as well.) This assumption drastically changes the problem. By Bayes' theorem we have

$$= \frac{P(H)P(E1, E2, \ldots, En|H)}{P(E1, E2, \ldots, En)}$$

Now, by exploiting a *determinism* assumption, namely, that each scenario will always manifest all

its components (cf. Goldman (1990)), we can assume that $P(E1, E2, \ldots, En|H)$ is 1 for all explanatory hypotheses. Thus, for all potential explanations, the following holds:

$$= \frac{P(H)}{P(E1, E2, \ldots, En)}$$

But the denominator is just a standard normalizing factor. Thus the probabilities of the potential explanations are in proportion to their priors.

That is, since we are dealing with a text, rather than unconstrained observations, ***Charniak and Goldman's entire probabilistic account amounts to believing the candidate scenarios in proportion to their priors.*** Indeed, a careful inspection shows that this is all Charniak and Goldman's Bayesian networks compute with respect to the problem at hand.[2]

We have now trivially dealt with the only probabilistic motivation: In cases like the "checking-a-bag-at-the-airport" story, the relative priors of air travel versus terrorist bombing let us conclude that air travel is being discussed. Note that we do not require unlikely knowledge of esoteric probabilities, nor are we required to perform any complex calculation: We require only the (relative) frequencies of the scenarios.

(Below we will argue that the above interpretation of what to compute is incorrect, but that performing this computation is guaranteed to produce the correct result.)

Consider now the following stories:

(13) **a.** Jan likes gardens with lots of shrubs. Pat prefers rock.

**b.** Jan likes jazz. Pat prefers rock.

These texts contain a paradigmatic example of disambiguating a word in context. This is just the sort of thing that one wants to "fall out" of an interpretation process; Charniak (1986) shows one way in which this might be so for texts involving plan recognition. The problem here is that, in this example, there is nothing at all about the *world* that makes the inference compelling: Indeed, Pat is probably more likely to prefer rock music to rock gardens. Rather, "rock" is disambiguated depending upon whether the *topic of the discourse* is types of garden material or types of music. The problem is, one would presumably like the same mechanism to handle disambiguation here as well as in cases of plan recognition, i.e., cases in which there is an ostensible connection to the real world. However, if the theory of text interpretation is parasitic on a theory of understanding the real world, it does not appear possible for this to be the case.

[2]To be fair, we note that to compute actual probabilities, we need to make an additional assumption of some sort. For example, we can make the assumption of *completeness*, i.e., that all the possible explanatory scenarios are known. We will argue below that similar assumptions are in fact warranted, but that the problem is even simpler than this because computing actual probabilities is unnecessary. In any case, the completeness assumption is no stronger than those required for a Bayesian network to produce a useful result.

# 5 A Proposal

Our proposal is rather old fashion. We are doing little more than extending, and putting on a more precise foundation, the well-known knowledge-based inference approach exemplified by Schank and Abelson (1975), Cullingford (1978), Wilensky (1978), Charniak (1978) and their siblings and descendants.

The basic idea of the model is as follows: There is a set of types of items that a text can convey. These items are *conceptual*, i.e., they correspond to the types of contents that a speaker might want to communicate. Sentences of a coherent text are presumed to comprise components of such an item. This assumption trivializes (but does not eliminate entirely) the probabilistic aspect of the understanding process.

The model enables us to concisely characterize the understanding process as follows:

> Find that interpretation of the text that conforms to plausible goals of a speaker, taking into account
>
> - the type of contents a speaker might attempt to convey,
>  and
> - the speaker's effectiveness in selecting utterances that will identify the speaker's intentions.

To flesh out the theory, we need to deliver on the following:

- A characterization of the plausible contents of discourses. Let us call these *content types.*
- A speaker strategy.

We will attempt to deliver primilary on the latter point, and then only enough to sketch out why natural language interpretation is not primarily a probabilistic enterprise.

Even at this sketchy level, however, the theory embodies a number of assumptions about the nature of discourse:

- A text is coherent to the extent that its elements can be (transitively) identified as components of a single entity.

  Here we are defining explaining an utterance as construing it as part of some larger, pre-defined schema.

  The class of such entities will necessarily be broad, including such familiar objects as events (of which the plans involved in "plan recognition" comprise a special case), but also such entities as contrasts, arguments, physical objects, expectation failures, "what has been happening in the family since you've last seen us", and so on. Similarly, what it means to be a component of such an entity will have a similarly broad interpretation. In particular, simple logical implications, such as those accommodated by inheritance, are meant to be included.

  The kinds of entities responsible for coherence at a low level (e.g., an event sequence) are likely to be rather different from the kinds responsible for global coherence (e.g., a novel). We will remain agnostic for the time being as to whether this difference is significant.

- The content types are largely invariant with speaker intent.

  This idealization is meant to capture the intuition that coherency is separable from appropriateness. That is, while the choice of a contents is a function of speaker intent, as is the manner in which the contents is conveyed, and the aspects of it that are actually mentioned, that it is a coherent contents can be judged independently. Thus, a speaker intending to communicate an experience to a hearer is likely to be informative, and state only propositions believed to be unknown to the hearer, whereas, if the intent is to entertain, the speaker might tell a story that the hearer already knows. Nevertheless, the discourse can be identified as coherent or not independently of these goals.

- Discourses can be categorized as defective or felicitous by a *sensible speaker/hearer* (Wilensky 1989).

In our critique of Charniak and Goldman's approach, we noted how the assumption of relevance allows the probabilistic computation to be trivialized to the computation of determining the priors on potentially explanatory scenarios. We now reexamine this idea. In particular, we defined the problem as finding the $H$ that maximizes

$$P(H|E1, E2, \ldots, En)$$

where $H$ is a potentially explanatory scenario. Let us now re-interpret this notation so that $H$ means the scenario the speaker intended to communicate, and the $E_i$ as speech acts, i.e., the utterances of a text, rather than events in the world. Bayes' theorem still gives us

$$= \frac{P(H)P(E1, E2, \ldots, En|H)}{P(E1, E2, \ldots, En)}$$

but now the interpretation of the components is quite different: The prior probability is now that of the speaker communicating a particular content; all we have said so far is that this will be 0 for contents not on our approved list. We will furhter presume that various contents more likely than others, and that speakers and hearers know these values (at least as approximated by their relative frequency in discourse) to within an order of magnitude or two.

The conditional probability is the probability that the particular utterances were made, given the intention to communicate the contents. Here is where the bulk of out discourse assumptions are pertinent. First, note that what we mean by this notation is that the speaker uttered precisely these utterances, and no more. Our discourse theory exploits this in the following postulate:

> Cooperative selection: Speakers will utter those components of a scenario that maximally distinguish it from other scenarios.

That is, the chance that a particular component will be selected by a speaker as part of an attempt to express a contents will be inversely proportional to its frequency of occurrence in scenarios.

Cooperative selection makes contents recognition simple. Consider for example our motivating "checking-a-bag-at-the-airport" text. Before we noted that relevance allowed us to believe a contents in proportion to its prior (now interpreted as the frequency of communication of a contents rather than as the frequency of occurrence of an event). However, since putting a bomb in a suitcase is such a good distinguisher of the terrorist scenario, the probability that a (unmarked) text intending to convey the terrorist scenario would contain a reference to it is extremely high – this reference would come close to uniquely identifying the scenario, which, after all, is the point of the discourse. Accordingly, the probability that the text not containing this reference would be an attempt to communicate the terrorist scenario would be vanishingly small.

These principles help explain why text interpretation is so easy: Utterances are selected by the speaker in such a fashion so that the relative priors of plausibly explanatory contents are exceedingly efficient in determining intention.

The time course of a text is also subject to a helpful discourse principle:

> Incremental Monotone Interpretation: In an "unmarked" discourse, determining a content type is done incrementally with each utterance; moreover, additional utterances do not reverse interpretation rankings.

For example, consider the following story:

(14) John picked up an old newspaper. He walked over to the bird cage. He said "Polly want a cracker?" He sat down in the chair next to the bird cage and turned on the lamp. He waited for a fly to land and swatted it.

Here it seems like the story is intentionally misleading us. Note that, in principle, we could have waited for more information and not have jumped to a conclusion. But instead the Incremental Monotone Interpretation leads us down a garden path.

We have tacitly been embracing the following assumption:

> Perfect knowledge model: In a felicitous discourse, the speaker's model of the hearer's relevant knowledge is correct.

We need this assumption, or something like it, of course, for communication to be successful under any scenario.

Well, what of the nature of content types? Of the first, we note here only that we anticipate content types to be a very broad class, and to include lots of types that have often been conceived of as discourse items per se. As an example, an "expectation failure" might be a content type. (Indeed, we suspect it is an extremely common one.) This observation suggests how to handle the negation cases that were deemed problematic above. Recall the following text:

(15) John doesn't put his money in a bank.

Assuming no special knowledge about not putting money in a financial institution or in the side of a river, the only interpretation found is likely to be an expectation failure of a financial institution scenario.

Another content type might be topically related facts. This is meant to handle such cases as

(16) Jan likes gardens with lots of shrubs. Pat prefers rock.

where the situation being conveyed is not a sequence of actions, but instead, some facts that are topically related.

The point of both examples is that the problems discussed above are resolved by having a theory of interpretation broader than that of being true in the world.

# 6 Some Problems and Comments

Of course, real texts are much more complicated than we (or the probability theorists) have been letting on. However, we suspect that much of the additional analysis that is needed fits nicely into the framework we have been trying to work out. For example, real texts often have sentences that serve a "meta", discourse function, such as

(17) A funny thing happened on the way to work today.

(18) I just heard an interesting fact.

From the viewpoint espoused here, we suspect that a good way to analyze such utterances is as a *cueing* function. That is, they help the hearer to select the general class of content types into which the text should be assimilated, and are thus in accordance with the Cooperative Selection idea.

Another complication is that the goal of the speaker is not properly expressed as communicating a content type, but all the ancillary propositions as well. That is, the speaker's goal is to communicate (at least) a set of propositions. This fact complicates the simple probabilistic formulation that we presented earlier, but does not appear to undercut the essential idea.

We have said only a little about the nature of content types. But one point is worth noting. Our "theory of discourse" has very little to do with text per se. In particular, it posits none of the "discourse relations" found, for example, in Hobbs (1985) or Hovy (1990). Indeed, the theory of discourse is almost entirely a theory of conceptual structure, treating the contents of individual utterances as related, *not to each other*, but to some overarching *conceptual structure* whose communication comprises the goal of the speaker. Doing so requires that we postulate as conceptual objects what others seem to have been viewed previously as primarily expository devices. Thus, we have postulated content types like expectation failure and topically related facts to avoid having a discourse theory requiring relations between sentences such as contrast, violated expections, evaluation, elaboration, etc.

In sum, we take the extreme position that the theory of discourse is both (i) the crucial basis for language interpretation, and (ii) practically nonexistent, given that one has in hand the proper conceptual menagerie.

# 7 References


Charniak, E. On the use of framed knowledge in language comprehension. .I Artificial Intelligence. .R Vol. II, No. 3, 1978.

Charniak, E. Passing markers: A theory of contextual infulence in language comprehension. In *Cognitive Science 7*, (3), 1983.

Charniak, E. A neat theory of marker passing. In the *Proceedings of the Fifth National Conference on Artificial Intelligence*, pp. 584-589, 1986.

Charniak, E., and Goldman R. P. Plan Recognition in Stories and in Life. In the *Proceedings of the Workshop on Uncertainty and Probability in AI.* Morgan Kaufman Publishers, Inc., 1989.

Cullingford, R. E. Script application: Computer understanding of newspaper stories. Yale University Computer Science Research Report #116, 1978.

Goldman, R. P. A Probabilistic Approach to Language Understanding. Brown University Department of Computer Science Technical Report No. CS-90-34. December, 1990.

Goldman, R. P. and Charniak, E. A probabilistic ATMS for plan recognition. In the *Proceedings of the Plan-recognition Workshop*, 1988.

Grice, H. P. Logic and conversation. In P. Cole and J. Morgan (eds.) *Syntax and Semantics 3: Speech acts.* Academic Press, New York, 1975.

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

Hobbs, J. R., Stickel, M., Martin, P. Edwards, D. Interpretation as Abduction. In the *Proceedings of the 26th Annual Meeting of the ACL*, pp. 95-103, 1988.

Hovy, Eduard H. Parsimonious and Profligate Approaches to the Question of Discourse Structure Relations. In *Proceedings of the Fifth International Workshop on Natural Language Generation*, Dawson, Pennylvania, 1990.

Norvig, P. and Wilensky, R. A Critical Evaluation of Commensurable Abduction Models for Semantic Interpretation. In the *Proceedings of COLING 90.* Helsinki, Finland.

Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufman Publishers, Inc., Los Altos, CA, 1988.

Schank, R. C. & Abelson, R. P. .I Scripts, Plans, Goals and Understanding. .R Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1977.

Shimony, S. and Charniak, E., A New Algorithm for Finding MAP Assignments to Belief Networks. In the *Proceedings of the Sixth Conference on Uncertainty in AI.* General Electric, 1990.

Sperber, D. & Wilson, D. *Relevance.* Harvard University Press, Cambridge, Mass., 1986.

Wilensky, R. .I Understanding Goal-Based Stories. .R Yale University Department of Computer Sciences Research Report #140, 1978.

Wilensky, R. Points: A Theory of Story Content. Berkeley Electronic Research Laboratory Memorandum No. UCB/ERL/M80/17. April 24, 1980.

Wilensky, R. *Planning and Understanding: A Computational Approach to Human Reasoning..* Addison Wesley, New York, 1983.

Wilensky, R. Primal Content and Actual Content: An Antidote to Literal Meaning. In the *Journal of Pragmatics*, Vol. 13, pp.163-186. Elsevier Science Publishers B.V. (North-Holland), Amsterdam, The Netherlands. 1989.

Wu, Dekai. Probabilistic Unification-Based Integration of Syntactic and Semantic Preferences For Nominal Compounds. In COLING-90: 13th International Conference on Computational Linguistics, August 1990.

Wu, Dekai. Automatic Inference: A Probabilistic Basis for Natural Language Interpretation. UCB technical report UCB//CSD-92-692.