EXPLOITING CONDITIONAL INDEPENDENCE IN PROBABILISTIC MODELS OF LANGUAGE UNDERSTANDING

Summary

In this paper we investigate a method to build probabilistic models of language understanding by systematically exploring interdependencies between multiple linguistic features. Our method reifies conditional independence assertions implicit in sign-based linguistic theories and describes the joint distribution of interrelated features compactly without requiring additional assumptions about modularity. With this approach, we are able to model a wide array of psycholinguistic results on lexical access and the access of idioms, and to explain some processing difficulties in syntactic disambiguation.

Subject Areas: sentence processing, psychological modeling, probabilistic models.

Word Count: 3194

1. INTRODUCTION

Recent years have seen a number of probabilistic models of different aspects of language interpretation and processing. For example some models of semantics use probabilities of certain semantic combinations to compute the maximum-likelihood interpretation of an ambiguous utterance (Charniak and Goldman 1988; Wu 1992; Hobbs and Bear 1990). Probabilistic models of syntax (Resnik 1993; Resnik 1992; Jurafsky 1992b; Stolcke 1995) augment syntactic rules with various kinds of conditional probabilities. In psycholinguistics, most models of lexical access include some account of frequency results (higher-frequency words are accessed faster). And recent psycholinguistic models of syntactic and lexical disambiguation (MacDonald 1993; MacDonald et al. 1994; Trueswell and Tanenhaus 1991; Trueswell and Tanenhaus 1994) show that certain processing difficulties can be modeled by considering appropriate lexical frequencies.

In this paper we investigate the a principled way to merge linguistic structure with probabilistic reasoning. We show that exploiting conditional independence in the various dimensions of linguistic knowledge allows one to model several well known psycholinguistic results on lexical access and the access of idioms, and on syntactic disambiguation and processing difficulties. We suggest that using conditional probabilities, rather than relative frequencies explains the similarity neighborhood effects found by (Luce et al. 1990) and others. Furthermore, we suggest that a probabilistic account may explain the often-noticed fact that top-down effects on access are very commonly found in psycholinguistic studies in phonology, but are much rarer in syntax.

1.1. Psychological results

Many psychological results on lexical access as well as lexical and syntactic disambiguation suggest that access and disambiguation are parallel, evidential, and probabilistic. Tyler (1984) and Salasoo and Pisoni (1985) show that the timing of lexical access is quite dependent on the frequency of the lexical item. High-frequency lexical items have higher initial activation than low-frequency ones (Marslen-Wilson (1990)), are accessed more easily (Tyler 1984 and Zwitserlood 1989), and reach recognition threshold more quickly

(Simpson and Burgess 1985 and Salasoo and Pisoni 1985). Cacciari and Tabossi (1988) found that different idioms are accessed at quite different rates, and in particular that the rate depending on the 'predictability' of the idiom; we will suggest a relation between predictability and probability. The evidential nature of access is demonstrated by studies which show that different contexts chance the access point even for the same construction, i.e. that the access point is context-sensitive.

Studies on disambiguation have also shown effects which have been argued to demonstrate evidential, probabilistic processing. Many such studies have focused on the main-verb (MV), reduced relative (RR) ambiguity (MacDonald 1993; MacDonald et al. 1994; Trueswell and Tanenhaus 1994; Trueswell and Tanenhaus 1991; Trueswell, Tannenhaus and Garnsey 1994). In many cases the MV/RR ambiguity is resolved in favor of the Main Clause reading leading to a garden path analysis.

(1) # The horse raced past the barn fell.

This has been accounted for by the different frequencies of the preterite and participial forms of the verb *raced* (MacDonald 1993; Simpson and Burgess 1985). In other cases, lexical constraints permit the RR interpretation.

(2) The horse carried past the barn fell.

(3) The horse found past the barn fell.

Furthermore, semantic context and thematic fit can also impact disambiguation. For instance, subjects experienced difficulty at the phrase "by the lawyer" only in the first of the following three examples (Trueswell, Tannenhaus and Garnsey 1994).

(4) The witness examined by the lawyer turned out to be unreliable.

(5) The witness who was examined by the lawyer turned out to be unreliable.

(6) The evidence examined by the lawyer turned out to be unreliable.

These semantic and thematic fits can also be represented probabilistically.

In summary:

1. The timing of lexical and idiom access is dependent on evidential context and on frequency.

2. Disambiguation is similarly dependent on lexical, syntactic, thematic, and semantic evidential context, and on frequency.

2. MODEL

Our proposal is that access and disambiguation should be treated with a single probabilistic model. To determine if a specific word is to be accessed, we compute the *conditional probability* of the word given the evidence. Evidence can come from syntactic, semantic, and phonetic sources, both top-down and bottom-up. Words are accessed if they fall within a certain percentage of the most highly-ranked word. This *threshold confidence ratio* is a universal constant in the grammar.

In concert with recent sign-based linguistic theories such as HPSG (Pollard and Sag 1994), Construction Grammar (Fillmore 1988; Fillmore et al. 1988), Cognitive Grammar (Langacker 1987), and other unification-based theories (Sag et al. 1985), we assume that the accessed construction is the site for common correlations and constraints between conceptual and perceptual sources of linguistic knowledge. Computationally, this allows us to view specific constructions as *values* of *latent variables* that render conditionally independent (d-separate (Pearl 1988; Lauritzen and Spiegelhalter 1988)) top-down syntactic, lexical, and other context information from bottom-up phonological and other sensory information.

Figure 1 shows a computational realization of this idea using Probabilistic nets (Pearl 1988). Probabilistic nets are able to represent hierarchical knowledge structures where evidence combination exploits the recursive decomposability of the joint distribution. The basic idea is to compactly represent the joint distribution of several sources using local conditional distributions combined with a set of conditional independence assertions that allow us to construct a global joint distribution from such local distributions.

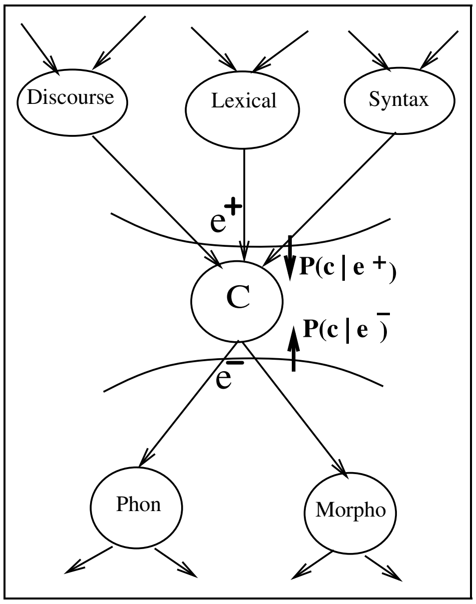


Figure 1: Sources of evidence for access, and a probabilistic network representing the role of top-down and bottom-up evidence.

In general, given the top-down and bottom up sources of evidence, we wish to compute the likelihood of the specific construction c ($c \in C$) being accessed. In other words, we wish to compute $P(c|e^+, e^-)$. The structure of the net in Figure 1 allows us to compute the influence of e^+ and e^- separately. Using Bayes' rule and the simplifying assumption that e^+ can affect e^- only through C, we get

$$\begin{aligned} P(c|e) = P(c|e^+, e^-) &= \frac{P(c, e^-|e^+)}{P(e^-|e^+)} \\ &= \frac{P(c|e^+)P(e^-|c, e^+)}{P(e^-|e^+)} \\ &= \frac{P(c|e^+)P(e^-|c)}{P(e^-|e^+)} \end{aligned} \tag{1}$$

The values of various sources are further recursively decomposable to observable and unobservable variables. For instance, an $SCFG$ grammar entails a conditional independence statement that intermediate nodes in a derivation tree (non-terminals) d-separate *outer* from *inner* nodes. Thus the influence of *outside* and *inside* probabilities are computed separately and then combined (Jelinek and Lafferty 1991; Stolcke 1995), much as in Equation 1. Lack of space prevents us from specifying the exact translation mechanism from $SCFG$ to probabilistic nets. In general, the probabilistic network formulation allows us to use standard evidence combination algorithms that exploit the various conditional independence relationships reflected in recursively decomposable structures (Lauritzen and Spiegelhalter 1988; Li and D'Ambrosio 1994; Pearl 1988).

3. RESULTS

We now turn to how our model explains explain psycholinguistic results on lexical and idiom access from a wide variety of sources, as well as being able to predict cases of processing difficulty and garden path phenomena for the data in Section 1.1.

3.1. Access is a Function of Conditional Probability, not Frequency

Probability-based and frequency-based models both explain the frequency effect in lexical access. Lexical items with a higher frequency will tend to have a higher probability, since the posterior is directly proportional to the prior. Since idioms and lexical items are treated equivalently, the model thus also accounts for experiments on idioms which show that more familiar idioms are accessed more quickly (d'Arcais 1993), and predicts that this frequency effect should extend to syntactic constructions.

But our probabilistic algorithm differs from a simple frequency-based algorithm in making the prediction that access of a construction will be *inversely* proportional to the commonality of the *evidence*. Intuitively, things which are common are not good evidence for any one construction in particular. This aspect of using posteriors rather than priors to model access explains experimental results in both lexical and idiom processing.

First, the model provides a way to formalize the intuitions of Cacciari and Tabossi's (1988) results on idiom access. Cacciari and Tabossi distinguished predictable and unpredictable idioms in Italian. Predictable idioms were those like *in seventh heaven* in which there is an early cue to idiomaticity. Unpredictable idioms were those like *go to the devil* in which the cue that the phrase was not literal comes late in the idiom. They found that the idiomatic meaning was accessed immediately for predictable idioms, but much later for unpredictable idioms (300 ms after offset). Since the authors controlled for frequency of idioms, this cannot be accounted for by assuming that more frequent idioms were accessed earlier. A further examination of their late-cue idioms showed that many of them also began with very common words such as *venire ('come')*, or *andare ('go')*. This result falls out of our model. The discriminating power of the evidence is low for cases where highly probable words or constructions begin the idiom because they provide evidence for so many other constructions as well. This means few constructions will be pruned.

Next, the probabilistic model explains the similarity neighborhood effects found by (Luce et al. 1990) and others. Luce et al. showed that words differ in their *similarity neighborhoods*, the set of words which resemble them phonetically. They showed a correlation between the time to access a word and the size of its neighborhood; words with large neighborhoods were slower on auditory naming and lexical decision. In our model, a large frequency-weighted neighborhood means that there are a lot of words which strongly resemble the target word – ie. have a common phonology. In this case, many words will remain above the *threshold confidence ratio* and not be pruned. A small neighborhood means there are very few words in the language whose phonotactics resemble the target word, and hence the string of phones used as evidence will have good discriminating power and many target words will be pruned.

3.2. Higher-Level Constraints Harder to Find?

In phonology, the conditional probability of a phoneme appearing given a word in which it occurs $P(phon|word)$ is very high, and thus top-down evidence for the phoneme will be quite high. Syntactic constraints, on the other hand, are generally specified in terms of very abstract constructions like NOUN or VERB. Thus the top-down conditional probability of any *particular* noun appearing is *typically* quite

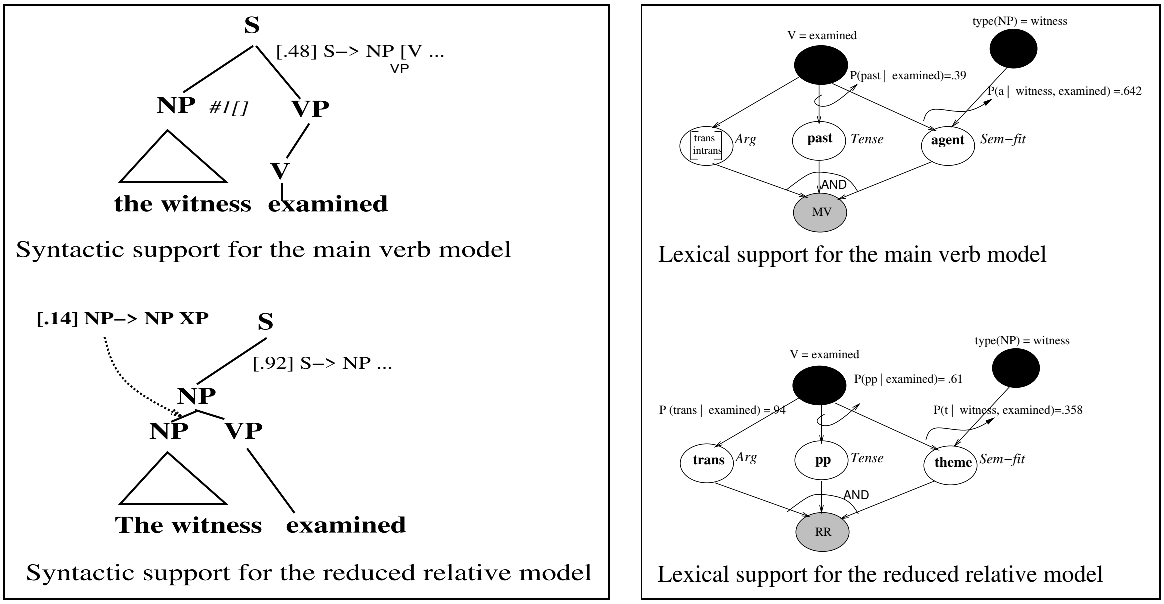
(a)  (b)

Figure 2: *(a)* shows the syntactic parse trees for the MV and the RR interpretations assuming an $SCFG$ generating grammar. *(b)* shows the probabilistic net that represents lexical support for the two interpretations for the same input. The input is from data in Section 1.1.

---

low, and top-down support $P(word|syn)$ will be much lower. Cacciari and Tabossi (1988), found just such top-down syntactic effects from idioms and Tanenhaus and Lucas (1987) notes that psycholinguistic evidence of top-down effects is very common in phonology, but much rarer in syntax.

3.3. Processing Difficulty and Disambiguation

To apply the model described above to on-line disambiguation, we assume that there are a few interpretations $(I_1, \ldots I_n)$ that are consistent with the input data. In this paper, we will focus on the values of lexical and syntactic features for the Relative (RR) and Main Verb (MV) interpretations at different stages of the input for examples in Section 1.1. Figure 2 illustrates the structure and values of syntactic and lexical features supporting the MV and RR interpretations for the input "The witness examined".

In Figure 2(a),the conditional probability of a construction given top-down syntactic evidence $P(c|e)$ is relatively simple to compute in a Construction Grammar or any other augmented-stochastic-context-free formalism. Recall that the $SCFG$ prior probability gives the conditional probability of the right hand side of a rule given the left hand side. In particular, since the parser operates left to right, the top-down probability $P(c|e)$ is the probability that the evidence construction e^+ left-expands to c:

(7)

$$P(e \overset{L*}{\Rightarrow} c)$$

In a context-free grammar, a nonterminal a left-expands to a nonterminal b if there is some derivation tree whose root is a and whose leftmost leaf is b. Using the conditional independence statements entailed by the derivation trees in Figure 2(a), we can compute the syntactic posterior confidence ratio, MV/RR. Equation 2 illustrates the basic computation.[1]

$$\begin{aligned}(\frac{MV}{RR})_{syn} = \frac{P(\textit{S} \rightarrow \textit{NP VP})P(\textit{NP} \rightarrow \text{``}\textit{w}\text{''})P(\textit{VP} \rightarrow \textit{V})P(\textit{V} \rightarrow \text{``}\textit{exam}\text{''})}{P(\textit{S} \rightarrow \textit{NP} \ldots)P(\textit{NP} \rightarrow \textit{NP XP})P(\textit{NP} \rightarrow \text{``}\textit{w}\text{''})P(\textit{VP} \rightarrow \textit{V})P(\textit{V} \rightarrow \text{``}\textit{exam}\text{''})} &= \\ \frac{P(\textit{S} \rightarrow \textit{NP VP})}{P(\textit{S} \rightarrow \textit{NP} \ldots) \times P(\textit{NP} \rightarrow \textit{NP XP})} &= \\ \frac{.48}{.92 \times .14} &= 3.73.\end{aligned} \tag{2}$$

Figure 2(b) shows the general structure and organization of lexical information sources. Our model requires conditional probability distributions specifying the preference of every verb for different argument structures, as well its preference for different tenses. We also compute the semantic fit between possible fillers in the input and different conceptual roles of a given predicate.[2]

To compute the semantic fit, we need a measure of the selectivity between conceptual roles in a thematic grid and possible input fillers. In this paper, we use typicality ratings that were generated by psychological norming studies (Trueswell, Tannenhaus and Garnsey 1994) to quantify the degree of fit between a specific

[1]The data in Figure 2(a) is from (Marcus et al. 1993).

[2]The role of other features such as *voice* and *aspect* in access and disambiguation can be systematically studied using methods developed here.

filler (such as "witness") to a specific argument slot ("agent" or "theme") *given* a predicate verb ("examined"). We could also obtain this information from an external source as was done by Resnik (Resnik 1993). One advantage of our representation is that parameter estimation methods can be applied to automatically estimate the true value of this measure, using prior knowledge generated from typicality ratings or from external sources such as Wordnet (Miller 1990).

The impact of these lexical features can be calculated using the probabilistic network shown in Figure 2(b). For a given input, some nodes are instantiated as evidence. The resulting structure is shown in Figure 2(b) for the input "The witness examined" (black nodes in the figure are evidence). Equation 3 illustrates how lexical features impact the MV/RR ratio.

$$\begin{aligned}
(\frac{MV}{RR})_{lex} &= \frac{P(past, agent|exam, w)}{P(pp, trans, theme|exam, w)} \\
&= \frac{P(past|exam) \times P(agent|exam, w)}{P(pp|exam) \times P(trans|exam) \times P(theme|exam, w)} \\
&= \frac{.39 \times .642}{.61 \times .358 \times .94} = 1.22
\end{aligned} \tag{3}$$

In equation 3, the first equality specifies the lexical constraints imposed by the MV and RR interpretations, while the second equality is based on Bayes rule utilizing the conditional independence assertions entailed by the form of the lexical net shown in Figure 2.

The overall posterior ratio requires propagating the *conjunctive* impact of syntactic and lexical sources on our model. For the data and structure shown in Figure 2, our model prediction is shown in Equation 4. [3]

[3] Our model is implemented in Common-Lisp and uses the public domain software package, IDEAL (Srinivas and Breese 1990) for belief propagation. The conjunctive requirements on lexical features (shaded nodes with the AND arc in Figure 2) are treated as *query nodes* during computation of the posterior.

$$\frac{MV}{RR} = \frac{MV_{lex \wedge syn}}{RR_{lex \wedge syn}} = 4.56 \tag{4}$$

As mentioned in Section 2, interpretations that fall below the threshold confidence ratio are *pruned*. For the examples presented here, this ratio was fixed at $5 : 1$. Thus our model predicts that while the MV interpretation is clearly preferred, the RR interpretation is not pruned and still active at this stage of the input.

We tested our model in the ambiguous region of the input for all examples in Section 1.1. Figure 3(a) shows the results obtained from our model for the data in Table 1. The data (taken from (MacDonald 1993) and from (Marcus et al. 1993) (for *found*) shows the relevant lexical probabilities for the examples "The horse raced past the barn fell"and the replacement of *raced* by *carried* or *found*. As shown in Figure 3(a), our model predicts that the MV/RR ratio exceeds the threshold immediately after the verb *raced* is accessed ($MV/RR \approx 387 \gg 5$) leading to the *pruning* of the RR interpretation. In the other cases, while the MV/RR ratio is temporarily rising, it never overshoots the threshold, allowing both the MV and the RR interpretations to be active throughout the ambiguous region.

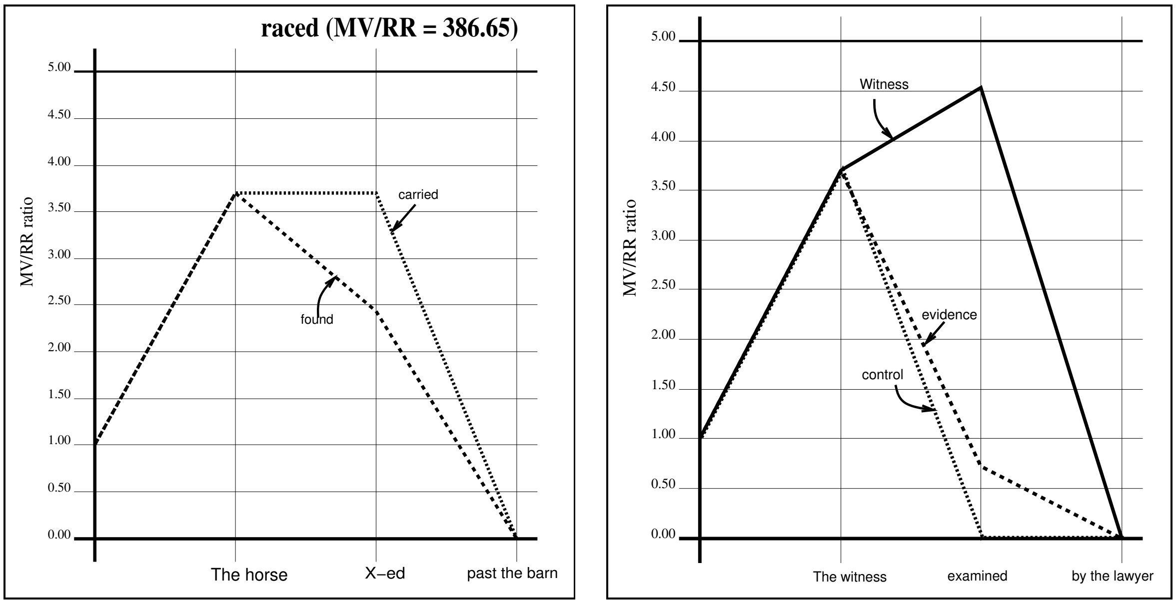
Table 1: Lexical probabilities for different verbs

		raced	**carried**	**found**
Tense	Past	92%	48%	38%
	Past participle	8%	52%	62%
Arg	Transitive	11%	94%	93%
	Intransitive	89%	6%	7%

Figure 3(b) shows the MV/RR ratio at different stages of the *examined* examples. Table 2 shows thematic fit data for these examples.[4]

As illustrated in Equation 4 after processing the input phrase "The witness examined",the RR inter-

[4]Information on thematic fit, indicated by the $P(x)$ entries for the different constructions are culled from typicality ratings used in the psychological study by (Trueswell, Tannenhaus and Garnsey 1994). The entry given by $P(a|x)$ are probabilities that the agent slot is bound to the subject position NP given x is the type of the NP.



(a) (b)

Figure 3: Results of our model on processing difficulty and disambiguation. *(a)* shows that the MV/RR posterior ratio for *raced* falls outside the threshold and the RR interpretation is pruned. For *found* and *carried*, both interpretations are active in the disambiguating region. *(b)* shows the role of thematic fit on the MV/RR ratio. Data shown for animate NP in the subject position (the witness), inanimate NP with strong semantic fit ("the evidence"), and with an unambiguous control.

Table 2: Lexical and Thematic fit probabilities for **examined**

verb	Past	PP	Trans	Intran	$P(a\|w)$	$P(t\|w)$	$P(a\|e)$	$P(t\|e)$
examined	.39	.61	.94	.06	. 642	.358	.18	.82

pretation is less preferred but not pruned. This leads to *limited* processing difficulty when encountering the next phrase "by the lawyer" which is both syntactically and semantically incompatible with the MV interpretation. No reassignment of roles is required in the case of "The evidence examined . . . ", or with the unambiguous control, hence no processing difficulty is predicted.

Thus our model garden-paths on the example "The horse raced past the barn fell" but will not garden path on the example "The horse carried past the barn fell" or on the example "The horse found past the barn fell". Our model also explains the correlations that (Trueswell, Tannenhaus and Garnsey 1994) found between thematic fit and processing difficulty. Furthermore we are able to explain garden-pathing as a graded effect, where processing difficulty and chance of garden pathing depends on how strongly the input favors a given interpretation.

4. CONCLUSION

The computational model proposed here combines two basic ideas in language processing. The first idea is to that multiple sources of linguistic knowledge, conceptual and perceptual, interact in access and disambiguation. This idea is manifest in PDP models of language processing (McClelland et al. 1989; Kawamoto 1993). The second idea is that linguistic knowledge is highly structured, and hierarchically organized (exemplified by unification grammars). Using probabilistic nets allows us to compute the joint distribution of multiple correlated features by using structural relationships to minimize the number of inter-feature interactions. This has the dual advantages of compact representation and clarity of model. Our hypothesis that linguistic structures are coded in *partially* independent dimensions allows us to model a wide array of psycholinguistic results, and offers a computational method to systematically investigate the modularity/non-modularity hypothesis.

REFERENCES

Cacciari, C. and Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, 27, 668–683.

Carlson, G. N. and Tanenhaus, M. K. (1987). Thematic roles and language comprehension. In W. Wilkins (Ed.), *Thematic relations*. Academic Press, San Diego.

Charniak, E. and Goldman, R. (1988). A logic for semantic interpretation. In *Proceedings of the 26th ACL*, Buffalo, NY.

d'Arcais, G. B. F. (1993). The comprehension and semantic interpretation of idioms. In C. Cacciari and P. Tabossi (Eds.), *Idioms: Processing, Structure, and Interpretation*. Lawrence Erlbaum Associates, New Jersey.

Fillmore, C. J. (1988). The mechanisms of "Construction Grammar". In *Proceedings of BLS 14*, pages 35–55, Berkeley, CA.

Fillmore, C. J., Kay, P., and O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, 64, 501–538.

Hobbs, J. R. and Bear, J. (1990). Two principles of parse preference. In *Proceedings of the 13th International Conference on Computational Linguistics (COLING-90)*, pages 162–167, Helsinki.

Jelinek, F. and Lafferty, J. D. (1991). Computation of the probability of initial substring generation by stochastic context-free grammars. *Computational Linguistics*, 17, 315–323.

Jurafsky, D. (1992b). An on-line computational model of human sentence interpretation. In *Proceedings of the Tenth National Conference on Artificial Intelligence (AAAI-92)*, pages 302–308, San Jose, CA.

Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity. *Journal of Memory and Language*, 32, 474–516.

Langacker, R. (1987). *Foundations of Cognitive Grammar I: Theoretical Prerequisites*. Stanford University Press, Stanford.

Li, Z. and D'Ambrosio, B. (1994). Efficient inference in Bayes networks as a combinatorial optimization problem. *International Journal of Approximate Reasoning*, 11(1), 55–81.

Luce, P. A., Pisoni, D. B., and Goldfinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive Models of Speech Processing*. MIT Press, Cambridge, MA.

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language*, 32, 692–715.

MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994). Syntactic ambiguity resolution as lexical ambiguity resolution. In *Perspectives on Sentence Processing*. Erlbaum, Hillsdale, NJ.

Marcus, M. P., Santorini, B., and Marcinkiewicz, M. A. (1993). Building a large annotated corpus of english: The Penn treebank. *Computational Linguistics*, 19, 313–330.

Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive Models of Speech Processing*. MIT Press, Cambridge, MA.

Marslen-Wilson, W., Brown, C. M., and Tyler, L. K. (1988). Lexical representations in spoken language comprehension. *Language and Cognitive Processes*, 3, 1–16.

McClelland, J. L., St. John, M., and Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes*, 4, 123–154.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufman, San Mateo, Ca.

Pollard, C. and Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago.

Resnik, P. (1992). Probabilistic tree-adjoining grammar as a framework for statistical natural language processing. In *Proceedings of the 14th International Conference on Computational Linguistics*, pages 418–424, Nantes, France.

Resnik, P. (1993). *Selection and Information: A Class-Based Approach to Lexical Relationships*. PhD thesis, University of Pennsylvania. (Institute for Research in Cognitive Science report IRCS-93-42).

Sag, I. A., Kaplan, R., Karttunen, L., Kay, M., Pollard, C., Shieber, S., and Zaenen, A. (1985). Unification and grammatical theory. In *Proceedings of the Fifth West Coast Conference on Formal Linguistics*.

Salasoo, A. and Pisoni, D. B. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language*, 24, 210–231.

Simpson, G. B. and Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 28–39.

Lauritzen, S. and Spiegelhalter, D. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *J. Royal Statistical Society* B, 50:127–224.

Srinivas, S. and Breese, J.S. (1990). IDEAL: A software package for the analysis of belief networks. Inname *Proceedings of Sixth Workshop on Uncertainty in AI*, Cambridge, Mass. 1990.

Stolcke, A. (1995). An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational Linguistics*, 21, 165–202.

Swinney, D. A. and Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, 18, 523–534.

Tanenhaus, M. K. and Lucas, M. M. (1987). Context effects in lexical processing. *Cognition*, 25, 213–234.

Trueswell, J. C. and Tanenhaus, M. K. (1991). Tense, temporal context and syntactic ambiguity resolution. *Language and Cognitive Processes*, 6, 303–338.

Trueswell, J. C. and Tanenhaus, M. K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In *Perspectives on Sentence Processing*. Erlbaum, Hillsdale, NJ.

Trueswell, J. C., Tanenhaus, M.K., and Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in parsing. *Journal of Memory and Language*, 1194, 285–317.

Tyler, L. K. (1984). The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics*, 36, 417–427.

Miller, George (1990). Wordnet: An on-line lexical database. *International Journal of Lexicaography*, 3(4)(Special Issue).

Wu, Dekai. (1992). *Automatic inference: A probabilistic basis for natural language interpretation*. PhD thesis, University of California, Berkeley. (Computer Science Department Tech. report UCB/CSD 92/692).

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, 32, 25–64.