# Bayesian Models of Human Sentence Processing

Srini Narayanan
University of California, Berkeley.
snarayan@cs.berkeley.edu

Daniel Jurafsky
University of Colorado, Boulder
jurafsky@colorado.edu

## Abstract

Human language processing relies on many kinds of linguistic knowledge, and is sensitive to their frequency, including lexical frequencies (Tyler, 1984; Salasoo & Pisoni, 1985; Marslen-Wilson, 1990; Zwitserlood, 1989; Simpson & Burgess, 1985), idiom frequencies (d'Arcais, 1993), phonological neighborhood frequencies (Luce, Pisoni, & Goldfinger, 1990), subcategorization frequencies (Trueswell, Tanenhaus, & Kello, 1993), and thematic role frequencies (Trueswell, Tanenhaus, & Garnsey, 1994; Garnsey, Pearlmutter, Myers, & Lotocky, 1997). But while we know that each of these knowledge sources must be probabilistic, we know very little about exactly how these probabilistic knowledge sources are combined. This paper proposes the use of Bayesian decision trees in modeling the probabilistic, evidential nature of human sentence processing. Our method reifies conditional independence assertions implicit in sign-based linguistic theories and describes interactions among features without requiring additional assumptions about modularity. We show that our Bayesian approach successfully models psycholinguistic results on evidence combination in human lexical, idiomatic, and syntactic/semantic processing.

## Introduction

Many modern psychological models of language processing are based on the on-line interaction of many kinds of linguistic knowledge, (Clifton, Speer, & Abney, 1991; Ferreira & Clifton, 1986; MacDonald, 1994; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; Trueswell *et al.*, 1994; Tyler, 1989). Although the exact time-course of the use of these different knowledge sources is not yet fully understood, it is clear that the processing of this knowledge is sensitive to frequency, from lexical frequencies (Tyler, 1984; Salasoo & Pisoni, 1985; Marslen-Wilson, 1990; Zwitserlood, 1989; Simpson & Burgess, 1985), idiom frequencies (d'Arcais, 1993), phonological neighborhood frequencies (Luce *et al.*, 1990), subcategorization frequencies (Trueswell *et al.*, 1993), and thematic role frequencies (Trueswell *et al.*, 1994; Garnsey *et al.*, 1997). Probabilistic versions of linguistic knowledge are also becoming common in linguistics (Resnik, 1993, 1992; Jurafsky, 1996).

But while we know that each of these knowledge sources must be probabilistic, and in fact we have preliminary probabilistic models of some specific linguistic levels, we know very little about exactly how these probabilistic knowledge sources are combined. This is particularly true with higher level knowledge, where the association of probabilities with sophisticated linguistic structural representation has only recently begun. Can a coherent probabilistic interpretation be given for the problem of language interpretation at different levels? What kinds of conditional independence assumptions can we make in combining knowledge, and how can we represent these assumptions? How can sophisticated linguistic structural knowledge be combined with probabilistic augmentations? The automatic speech processing (ASR) and natural language processing (NLP) literature (Bahl, Jelinek, & Mercer, 1983; Fujisaki, Jelinek, Cocke, & Black, 1991; Charniak & Goldman, 1988; Hobbs & Bear, 1990) have argued that language processing must be evidential and Bayesian. This paper proposes the use of Bayesian decision trees to address the issues in modeling the probabilistic, evidential nature of human sentence processing.

## Basic Result

The idea that lexical access is parallel is well-accepted (Swinney, 1979), and it is also widely assumed that at least some aspects of syntactic processing are parallel (Gorrell, 1989; MacDonald, 1993). Similarly well-accepted is the role that frequency plays in lexical (Marslen-Wilson, 1990; Salasoo & Pisoni, 1985; Simpson & Burgess, 1985; Zwitserlood, 1989), idiomatic (d'Arcais, 1993), syntactic (Trueswell *et al.*, 1993), and thematic processing (Trueswell *et al.*, 1994; Garnsey *et al.*, 1997).

Jurafsky (1996) argued that a Bayesian model (i.e. using posterior probabilities rather than frequencies) was also able to account for a number of effects that were not explainable by a frequentist model, including the intuitions of the Cacciari and Tabossi (1988) results on idiom access, the Luce *et al.* (1990) results on similarity neighborhoods, and the insight of Tanenhaus and Lucas (1987) that psycholinguistic evidence of top-down effects is very common in phonology, but much rarer in syntax.

But complete probabilistic models of syntactic and semantic processing have been much harder to build. For example, a number of studies have focused on the main-verb (MV), reduced relative (RR) ambiguity (Frazier & Rayner, 1987; MacDonald, 1993; MacDonald, Pearlmutter, & Seidenberg, 1994; Spivey-Knowlton & Sedivy, 1995; Trueswell & Tanenhaus, 1994, 1991; Trueswell *et al.*, 1994; Spivey-Knowlton & Sedivy, 1995). 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.

Proponents of the *constraint satisfaction* model have argued that this can be accounted for by the different lexical/morphological frequencies of the preterite and participial forms of the verb *raced* (MacDonald, 1993; Simpson & Burgess, 1985).

But in other cases, constraints on verb subcategorization permit the RR interpretation. The verb *found*, for example, is transitive, and so doesn't cause as strong a garden path in the RR interpretation (Pritchett, 1988; Gibson, 1991):

2. a. The horse carried past the barn fell.
   b. The bird found in the room died.

Studies have also found probabilistic effects of verb subcategorization preferences (Jurafsky, 1996; Trueswell *et al.*, 1993). For example Jurafsky (1996) suggested that the garden-path effect could be caused by a combination of lexical, syntactic, and verb subcategorization probabilities. More recent studies have suggested that semantic context and thematic fit can also impact disambiguation. For instance Trueswell *et al.* (1994) showed that strong thematic constraints were also able to ameliorate garden path effects in RR/MV ambiguities; subjects experienced difficulty at the phrase "by the lawyer" only in the first of the following three examples: [1]

3. a. The witness examined by the lawyer turned out to be unreliable.
   b. The witness who was examined by the lawyer turned out to be unreliable.
   c. The evidence examined by the lawyer turned out to be unreliable.

Thus assorted previous work has argued that various probabilistic knowledge sources each play a role in processing; but how exactly are these probabilities to be combined? Our model is based on 3 assumptions: linguistic knowledge is represented probabilistically, multiple interpretations are maintained in parallel, and the probabilities of these interpretations can be computed via a Belief net ('probabilistic independence net'). Given the probabilities and the Bayes formalism, the model explains a number of psychological results. The next section explains what we mean by 'assigning probabilities to linguistic structure'. We then introduce the probabilistic independence net formalism for combining different probabilities. Finally, we examine how well the model stands up to various psychological results.

## Prior Probabilities

We assume that linguistic knowledge is represented as a collection of signs or constructions, each of which represents a conventionalized pairing of meaning and form, and each of which is represented as signs in typed unification-based augmented context-free rules (Pollard & Sag, 1987; Fillmore, 1988). Thus words, morphological structures (like the *-ed* past tense morpheme), and syntactic constructions (like the passive construction) are each represented as 'constructions'. Each of these constructions is associated with a prior probability, which can be computed from relative frequencies from corpora or norming studies. [2] For example, in order to compute the probability of the simplified Stochastic Context-Free Grammar (SCFG) rule in (1), we can use the Penn Treebank (Marcus *et al.*, 1993) to get a frequency for all NPs (52,627), and then for those NP's which consist of a Det and an N (33,108). The conditional probability is then 33,108/52,627=.63.

(1) [.63] *NP* $\rightarrow$ *Det N*

Similarly, verb subcategorization probabilities can be computed from the Treebank or from norming studies like Connine *et al.* (1984). Thematic probabilities can be computed by normalizing verb bias norms, for example from Garnsey *et al.* (1997). Table 1 shows some lexical probabilities, for the verb *examine*, including morphological, subcategorization, and thematic probabilities. The thematic probabilities were computed by using psychological norming studies (Trueswell *et al.*, 1994) to quantify the degree of fit between a specific filler (such as "witness") to a specific argument slot ("agent" or "theme") *given* a predicate verb ("examined"). This information can also be obtained from a semantic database (like WordNet) as was done by Resnik (Resnik, 1993). See Jurafsky (1996) for further details of the probability computations.

Table 1: Lexical and Thematic fit probabilities for **examined**. Note "A" refers to Agent, "e" to examined, "ev" to evidence, "w" to witness, and "T" to theme.

| Past | PP | Trans | Intran |
|---|---|---|---|
| .39 | .61 | .94 | .06 |
| $P$(A\|w,e) | $P$(T\|w,e) | $P$(A\|ev,e) | $P$(T\|ev,e) |
| .642 | .358 | .18 | .82 |

## Construction Processing via Belief nets

Bayesian belief networks are data-structures that represent probability distributions over a collection of random variables. The basic network consists of a set of *variables* and directed *edges* between variables. Each variable can take on one of a finite set of states. The variables and edges together form a *directed acyclic graph*(DAG). For each variable $A$ (a node in the graph) with parents $B_1, \ldots B_n$, there is an attached conditional probability table $P(A|B_1, \ldots, B_n)$. Importantly, the network structure reflects conditional indepen-

[1] Although the original study by Ferreira and Clifton (1986) had not found semantic effects, Trueswell *et al.* (1994) used a stronger manipulation of thematic constraint.

[2] See (Roland & Jurafsky, 1998; Merlo, 1994; Gibson & Pearlmutter, 1994) for comparisons of experimental and corpus-based frequencies.

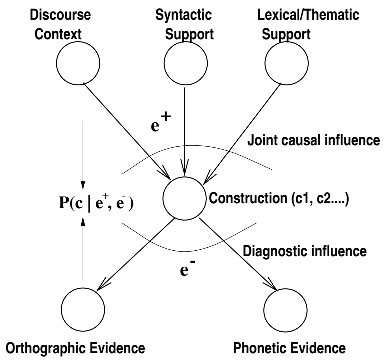


Figure 1: Sources of evidence for access, and a Belief network representing the role of top-down and bottom-up evidence.

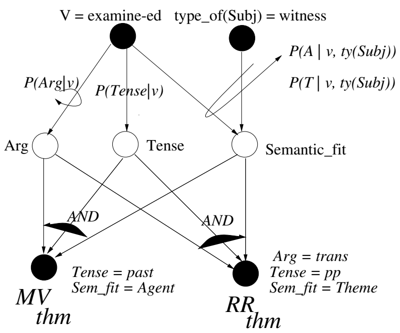


Figure 2: The Belief net that represents lexical support for the two interpretations for the same input. The input is from data in Table 1.

dence relations between variables, which allow a decomposition of the joint distribution into a product of conditional distributions. The following theorem sets up the basic chain rule which is used for computing the joint distribution from the conditional distribution.[3]

**Theorem 1** . *Jensen (1995) Let* $B$ *be a Belief network over* $U = \{A_1, \ldots, A_m\}$ . *Then the joint probability distribution P(U) is the product of the local conditional probability distributions specified in* $B$ *:*

$$P(U) = \prod_i P(A_i | pa(A_i)) \tag{1}$$

*where* $pa(A_i)$ *is the parent set of* $A_i$ .

The crucial insight of our Belief net model is to view specific constructions as *values* of *latent variables* that render top-down ( $e^+$ ) and bottom-up evidence ( $e^-$ ) conditionally independent (d-separate them (Pearl, 1988)). Thus syntactic, lexical, argument structure, and other contextual information acts as *prior* or *causal* support for a construction/interpretation, while bottom-up phonological or graphological and other perceptual information acts as *likelihood*, *evidential*, or *diagnostic* support. Figure 1 shows a computational realization of this idea.

Using Belief nets to model human sentence processing allows us to a) quantitatively evaluate the impact of different independence assumptions in a uniform framework, b) directly model the impact of highly structured linguistic knowledge sources with local conditional probability tables, while well known algorithms for updating the Belief net (Jensen (1995)) can compute the global impact of new evidence, and c) develop an on-line interpretation algorithm, where partial input corresponds to partial evidence on the network, and the update algorithm appropriately marginalizes over unobserved nodes. So as evidence comes in incrementally, different nodes are instantiated and the posterior probability of different constructions changes appropriately.

[3]For a comprehensive exposition see Pearl (1988), Jensen (1995).

To apply our model to on-line disambiguation, we assume that there are a set of constructions ( $(c_1, \ldots c_n) \in C$ ) that are consistent with the input data. At different stages of the input, we compute the posterior probabilities of the different interpretations given the top down and bottom-up evidence seen so far. We then apply the *beam-search* algorithm of Jurafsky (1996): prune out all constructions whose posterior probability is less than a certain ratio of the best construction (highest posterior). We will refer to this ratio as the Threshold Confidence Ratio (TCR). (i.e. prune out all $c \in C$ where $\frac{P(c_{best})}{P(c)} \geq TCR$ ). [4]

## Modeling Lexical and Thematic support

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.[5] Figure 2 shows the general structure and organization of lexical and thematic information sources. The thematic probabilities and their method of computation were shown in Table 1. As shown in Figure 2, the $MV$ and $RR$ interpretations require the conjunction of specific values corresponding to tense, semantic fit and argument structure features. Note that only the $RR$ interpretation requires the transitive argument structure.

## Modeling syntactic support

In Figure 3, the conditional probability of a construction given top-down syntactic evidence $P(c|e)$ is relatively simple to compute in an augmented-stochastic-context-free formalism (parse trees shown in Figure 3. Recall that the

[4]In this paper, we will focus on the support from thematic, and syntactic features for the Reduced Relative (RR) and Main Verb (MV) interpretations at different stages of the input for the examples we saw earlier. So we will have two constructions $c_1, c_2 \in C$ where $P(c_1|e^+, e^-) = MV, P(c_2|e^+, e^-) = RR$. For all examples reported here we set $TCR = 5$ (prune out the $RR$ interpretation if $\frac{MV}{RR} \geq 5$ ).

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

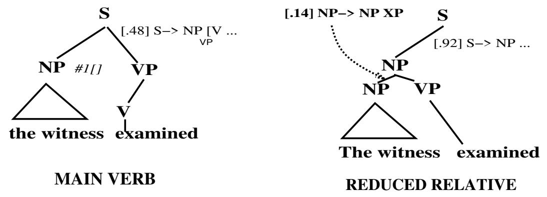


Figure 3: The syntactic parse trees for the $MV$ and the $RR$ interpretations assuming an $SCFG$ generating grammar.

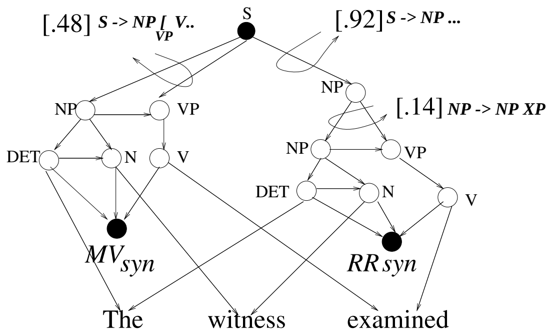


Figure 4: The Belief network corresponding to the syntactic support.

$SCFG$ 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 $P(c|e_{syn})$ is the probability that the evidence left-expands to $c$:

$$P(e \stackrel{L*}{\Rightarrow} c) \tag{2}$$

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 $b$.

Figure 4 illustrates the Belief network representation that corresponds to the syntactic parse trees in Figure 3. Note that the context-freeness property translates into the conditional independence statements entailed by the network. [6]

## Computing the joint influence

The overall posterior ratio requires propagating the *conjunctive* impact of syntactic and lexical/thematic sources on our model. Figure 5 shows our Belief net architecture for combining the two sources. The Belief net in Figure 5 embodies the assumption that the syntactic and thematic influences are dependent only the value of the specific construction, which in this case is either the Main Verb (MV) or the Reduced Relative (RR) construction. In other words, inter-source dependencies are explicitly captured by specific constructions. Furthermore, in computing the conjunctive impact of the lexical/thematic and syntactic support to compute $MV$ and $RR$, we use the well studied NOISY-AND model (Pearl, 1988)

[6]For exact technical details, including an automatic network construction technique, refer to (Narayanan, 1998)

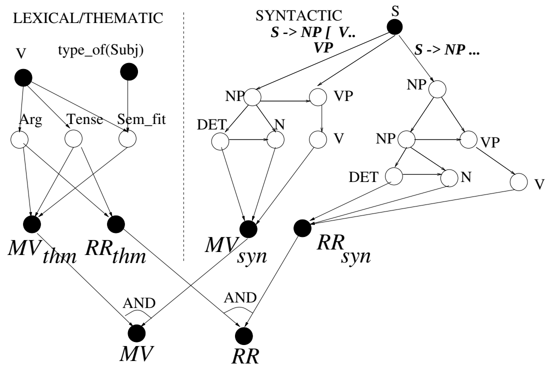


Figure 5: The Belief net that combines the thematic and syntactic support for a specific construction.

for combining conjunctive sources where it is assumed that whatever *inhibits* a specific source (syntactic) from indicating support for a construction, is independent of mechanisms that inhibit other sources (lexical) from indicating support for the same construction. This is called the assumption of exception independence, and is used widely with respect to both disjunctive (NOISY-OR) and conjunctive sources.

## Model results

There are a number of psycholinguistic results which argue for a Bayesian model of sentence processing. See Jurafsky (1996), for example, for a summary of the argument that conditional probabilities are a more appropriate metric than frequencies. The main result we will discuss here is evidence from on-line disambiguation studies that shows that a Bayesian implementation of probabilistic evidence combination accounts for garden-path disambiguation effects.

We tested our model in the ambiguous region of the input for all example sentences presented earlier, by computing the ratio $\frac{MV}{RR}$ of the posterior at different stages of the input. Note that under partial input the Belief net inference automatically marginalizes over the values of the unseen input. So in the case when only the subject has been input ("the horse" in the examples in Figure 6) the thematic influence is minimal and the $\frac{MV}{RR}$ ratio is basically a result of the syntactic support.

The data in Figure 6 was taken from (MacDonald, 1993) and from (Marcus *et al.*, 1993) (for *found*). Figure 6 shows the relevant posterior probabilities for the examples "The horse raced past the barn fell"and the replacement of *raced* by *carried* or *found* at different stages of the input. As shown in Figure 6, 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.

Figure 6 and Figure 7 show the $MV/RR$ ratio at different

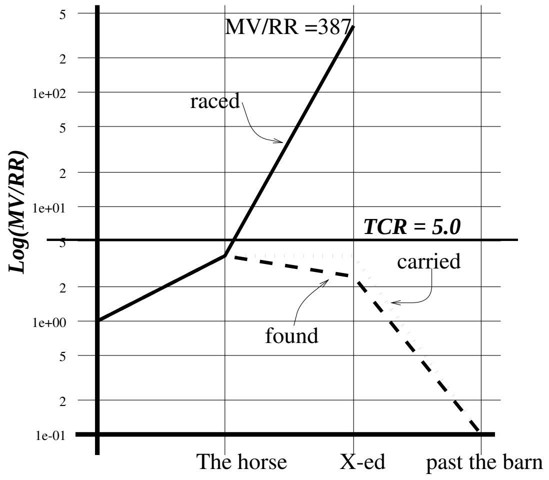


Figure 6: Disambiguation with Lexical Probabilities showing that the $MV/RR$ posterior ratio for *raced* falls above the threshold and the $RR$ interpretation is pruned. For *found* and *carried*, both interpretations are active in the disambiguating region.

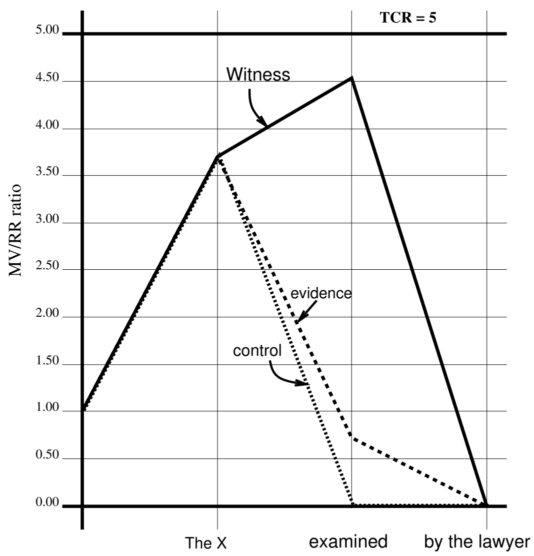


Figure 7: The role of thematic fit on the $MV/RR$ ratio. Data shown for animate $NP$ in the subject position (the witness), inanimate $NP$ with strong semantic fit ("the evidence"), and with an unambiguous control.

stages of the *examined* examples. Information on thematic fit, was culled from Typicality ratings used in the psychological study by (Trueswell *et al.*, 1994).

As illustrated in Figure 7 after processing the input phrase "The witness examined",the $RR$ interpretation is less preferred but not pruned. This leads to *limited* processing difficulty (limited because it approaches the $TCR$, but never exceeds it) 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 *et al.*, 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.

## Conclusion

The computational model proposed here combines two basic ideas in language processing. The first idea is that multiple sources of linguistic knowledge, conceptual and perceptual, interact in access and disambiguation. This idea is manifest in the psychological literature on lexical access and sentence processing, as well as in PDP and dynamical systems models of language processing (Tabor, Juliano, & Tanenhaus, 1997). The second idea is that linguistic knowledge is highly structured, and hierarchically organized (exemplified by syntactic and argument-structure knowledge). 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.

## Acknowledgements

Thanks to Jerry Feldman, Stuart Russell and Nancy Chang for valuable comments on various aspects of this work.

## References

Bahl, L. R., Jelinek, F., & Mercer, R. L. (1983). A maximum likelihood approach to continuous speech recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *5*(2), 179–190.

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

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

Clifton, Jr, C., Speer, S., & Abney, S. (1991). Parsing arguments: Phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language*, *30*, 251–271.

Connine, C., Ferreira, F., Jones, C., Clifton, C., & Frazier, L. (1984). Verb frame preference: Descriptive norms. *Journal of Psycholinguistic Research*, *13*(4), 307–319.

d'Arcais, G. B. F. (1993). The comprehension and semantic interpretation of idioms. In Cacciari, C., & Tabossi, P. (Eds.), *Idioms: Processing, Structure, and Interpretation*, pp. 79–98. Lawrence Erlbaum Associates, New Jersey.

Ferreira, F., & Clifton, Jr, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*, 348–368.

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

Frazier, L., & Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language*, *26*, 505–526.

Fujisaki, T., Jelinek, F., Cocke, J., & Black, E. (1991). A probabilistic parsing method for sentence disambiguation. In Tomita, M. (Ed.), *Current Issues in Parsing Technology*, pp. 139–152. Kluwer, Boston.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*, 58–93.

Gibson, E. (1991). *A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA.

Gibson, E., & Pearlmutter, N. J. (1994). A corpus-based analysis of psycholinguistic constraints on preposition-phrase attachment. In *Perspectives on Sentence Processing*, pp. 181–198. Erlbaum, Hillsdale, NJ.

Gorrell, P. G. (1989). Establishing the loci of serial and parallel effects in syntactic processing. *Journal of Psycholinguistic Research*, *18*(1), 61–73.

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

Jensen, F. (1995). *Bayesian Networks*. Springer-Verlag.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*, 137–194.

Luce, P. A., Pisoni, D. B., & Goldfinger, S. D. (1990). Similarity neighborhoods of spoken words. In Altmann, G. T. M. (Ed.), *Cognitive Models of Speech Processing*, pp. 122–147. 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. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, *9*(2), 157–201.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Syntactic ambiguity resolution as lexical ambiguity resolution. In *Perspectives on Sentence Processing*, pp. 123–154. Erlbaum, Hillsdale, NJ.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn treebank. *Computational Linguistics*, *19*(2), 313–330.

Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In Altmann, G. T. M. (Ed.), *Cognitive Models of Speech Processing*, pp. 148–172. MIT Press, Cambridge, MA.

Merlo, P. (1994). A corpus-based analysis of verb continuation frequencies for syntactic processing. *Journal of Psycholinguistic Research*, *23*(6), 435–457.

Narayanan, S. (1998). Graphical models of stochastic grammars. Tech. rep. TR-98-012, ICSI, Berkeley, CA.

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

Pollard, C., & Sag, I. A. (1987). *Information-Based Syntax and Semantics: Volume 1: Fundamentals*. University of Chicago Press, Chicago.

Pritchett, B. (1988). Garden path phenomena and the grammatical basis of language processing. *Language*, *64*(3), 539–576.

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*, pp. 418–424 Nantes, France.

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

Roland, D., & Jurafsky, D. (1998). How verb subcategorization frequencies are affected by corpus choice. submitted to *ACL-98*.

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

Simpson, G. B., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(1), 28–39.

Spivey-Knowlton, M., & Sedivy, J. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *In press*.

Spivey-Knowlton, M., Trueswell, J., & Tanenhaus, M. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses.. *Canadian Journal of Experimental Psychology*, *47*, 276–309.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*, 645–659.

Tabor, W., Juliano, C., & Tanenhaus, M. (1997). Parsing in a dynamical system. *Language and Cognitive Processes*, *12*, 211–272.

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

Trueswell, J. C., & Tanenhaus, M. K. (1991). Tense, temporal context and syntactic ambiguity resolution. *Language and Cognitive Processes*, *6*(4), 303–338.

Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In *Perspectives on Sentence Processing*, pp. 155–179. Erlbaum, Hillsdale, NJ.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of of thematic role inforomation in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285–318.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical prefernce form garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*(3), 528–553.

Tyler, L. K. (1984). The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics*, *36*(5), 417–427.

Tyler, L. K. (1989). The role of lexical representations in language comprehension. In Marslen-Wilson, W. (Ed.), *Lexical Representation and Process*, pp. 439–462. MIT Press, Cambridge, MA.

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