Tanenhaus, M., Leiman, J., & Seidenberg, M. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic context. *Journal of Verbal Learning and Behaviour, 18*, 427–441.

Voutilainen, A. (1995). A syntax-based part-of-speech analyser. In *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics, Dublin*, pp. 157–164. San Francisco, CA. Morgan Kaufmann.

Williams, C. & Hinton, G. (1990). Mean field networks that learn to discriminate temporally distorted strings. In D. Touretsky, J. Elman, T. Sejnowski, & G. Hinton (eds.), *Connectionist Models: Proceedings of the 1990 Summer School*, pp. 18–22.

Willshaw, D. (1981). Holography, association and induction. In G. Hinton & J. Anderson (eds.), *Parallel Models of Associative Memory*, pp. 83–104. Hillsdale, NJ: Erlbaum.

Willshaw, D., Buneman, P., & Longuet-Higgins, C. (1969). Non-holographic associative memory. *Nature, 222*, 960–962.

A computational model of the grammatical aspects of word recognition as supertagging

Albert E. Kim, Bangalore Srinivas and John C. Trueswell

University of Washington / AT&T Labs-Research / University of Pennsylvania

We describe a Constraint-Based Lexicalist model of human sentence processing. Highlighting a convergence of developments in multiple fields toward lexicalist and statistical processing perspectives, we argue that much of the syntactic ambiguity of language can be understood as lexical ambiguity, which is resolved during word recognition. The model is a connectionist system, which acquires wide coverage grammatical knowledge from supervised training on highly variable, naturally occurring text. The model learns to map each of the words in a sentence to an elementary tree from Lexicalized Tree Adjoining Grammar (Joshi & Schabes, 1996). These elementary trees are rich in grammatical information, encoding, among other things, the number and type of complements taken by a verb. The syntactic richness of these lexical representations results in substantial lexico-syntactic ambiguity. At the same time, statistical mechanisms for lexical ambiguity resolution are shown to effectively resolve this ambiguity. Simulations show that the model accounts for previously reported patterns in human sentence processing, including frequency-shaped processing of verb subcategory (e.g., Juliano & Tanenhaus, 1994) and effects of subtle contextual cues in lexical category ambiguity resolution (e.g., MacDonald, 1993).

In the last fifteen years, there has been a striking convergence of perspectives in the fields of linguistics, computational linguistics, and psycholinguistics regarding the representation and processing of grammatical information. First, the lexicon has played an increasingly important role in the representation of the syntactic aspects of language. This is exemplified by the rise of grammatical formalisms that assign a central role to the lexicon for characterizing syntactic forms, e.g., LFG (Bresnan & Kaplan, 1982), HPSG (Pollard & Sag, 1994), CCG (Steedman, 1996), Lexicon-Grammars (Gross, 1984), LTAG (Joshi &

Schabes, 1996), Link Grammars (Sleator & Temperley, 1991) and the Minimalist Program within GB (Chomsky, 1995). Second, theories of language processing have seen a shift away from "rule-governed" approaches for grammatical decision-making toward statistical and constraint-based approaches. In psycholinguistics, this has been characterized by a strong interest in connectionist and activation-based models (e.g., Lewis, 1993; McRae, Spivey-Knowlton & Tanenhaus, 1998; Stevenson, 1994; Tabor, Juliano & Tanenhaus, 1996). In computational linguistics, this is found in the explosion of work with stochastic approaches to structural processing (cf. Church & Mercer, 1993). In linguistics, this interest is most apparent in the development of Optimality Theory (Prince & Smolensky, 1997).

In this chapter, we highlight how the shift to lexical and statistical approaches has affected theories of sentence parsing in both psycholinguistics and computational linguistics. We present an integration of ideas developed across these two disciplines, which builds upon a specific proposal from each. Within psycholinguistics, we discuss the development of the Constraint-Based Lexicalist (CBL) theory of sentence processing (MacDonald, Pearlmutter & Seidenberg, 1994; Trueswell & Tanenhaus, 1994). Within computational linguistics, we discuss the development of statistical approaches to processing Lexicalized Tree-Adjoining Grammar (LTAG, Joshi & Schabes, 1996). Finally, we provide a description of the CBL theory, which is based on LTAG.

A constraint-based theory of sentence processing

Psycholinguistic thinking about the syntactic aspects of language comprehension has been deeply influenced by theories that assign a privileged role to supra-lexical syntactic representations and processes. This view has been most extensively developed in the theory of Frazier (1979, 1989), which proposed that syntactic processing is controlled by a two-stage system. In the first stage, a single syntactic representation of the input is computed using a limited set of phrase structure rules and basic grammatical category information about words. When syntactic knowledge ambiguously allows multiple analyses of the input, a single analysis is selected using a small set of structure-based processing strategies. In a second stage of processing, the output of this structure-building stage is integrated with and checked against lexically specific knowledge and contextual information, and initial analyses are revised if necessary. The basic proposal of this theory – that syntactic processing is, at least in the earliest stages, independent from lexically specific and contextual influences – has been one of the dominant ideas of sentence processing theory (e.g., Ferreira & Clifton, 1986; Perfetti, 1990; Mitchell, 1987, 1989; Rayner, Carlson & Frazier, 1983).

A diverse group of recent theories has challenged this two-stage structure-building paradigm by implicating some combination of lexical and contextual constraints and probabilistic processing mechanisms in the earliest stages of syntactic processing (Crocker, 1994; Corley & Crocker, 1996; Ford, Bresnan & Kaplan, 1982; Gibson, 1998; Jurafsky, 1996; MacDonald et al., 1994; Pritchett, 1992; Stevenson, 1994; Trueswell & Tanenhaus, 1994). We focus in this chapter on the body of work known as the Constraint-Based Lexicalist theory (MacDonald et al., 1994; Trueswell & Tanenhaus, 1994), which proposes that all aspects of language comprehension, including the syntactic aspects, are better described as the result of pattern recognition processes than the application of structure building rules. Word recognition is proposed to include the activation of rich grammatical structures (e.g., verb argument structures), which play a critical role in supporting the semantic interpretation of the sentence. These structures are activated in a pattern shaped by frequency, with grammatically ambiguous words causing the temporary activation of multiple structures. The selection of the appropriate structure for each word, given the context, accomplishes much of the work of syntactic analysis. That is, much of the syntactic ambiguity in language is proposed to stem directly from lexical ambiguity and to be resolved during word recognition.[1] The theory predicts that initial parsing preferences are guided by these grammatical aspects of word recognition.

The CBL framework can be illustrated by considering the role of verb argument structure in the processing of syntactic ambiguities like the Noun Phrase/Sentence Complement (NP/S) ambiguity in sentences like (1a) and (1b).

(1) a. *The chef forgot the recipe was in the back of the book.*
 b. *The chef claimed the recipe was in the back of the book.*

In (1a), a temporary ambiguity arises in the relationship between the noun phrase *the recipe* and the verb *forgot*. Due to the argument structure possibilities for *forgot*, the noun phrase could be a direct object or the subject of a sentence complement. In sentences like this, readers show an initial preference for the direct object interpretation of the ambiguous noun phrase, resulting in increased reading times at the disambiguating region *was in...* (e.g., Holmes, Stowe & Cupples, 1989; Ferreira & Henderson, 1990; Rayner & Frazier, 1987). On the CBL theory, the direct object preference in 1a is due to the lexical representation of the verb *forgot*, which has a strong tendency to take a direct object rather than a sentence complement. The CBL theory proposes that word recognition includes the activation of not only semantic and phonological representations of a word, but also detailed syntactic representations. These lexico-syntactic representations, and the processes by which they are activated, are proposed to play critical roles in the combinatory commitments of language comprehension. The direct object preference should therefore be eliminated when the verb *forgot* is replaced with a verb like *claimed*, which has a strong tendency to take a sentence complement rather than a direct object. These predictions have been confirmed experimentally (Trueswell, Tanenhaus & Kello, 1993; Garnsey, Pearlmutter, Myers & Lotocky, 1997), and connectionist models have captured these preferences (Juliano & Tanenhaus, 1994; Tabor et al., 1996).

Experimental work has also indicated that the pattern of processing commitments is not determined solely by individual lexical preferences, but involves an interaction between argument structure preference and lexical frequency. NP-biased verbs result in strong direct object commitments regardless of the lexical frequency of the verb. S-bias verbs, on the other hand, show an effect of frequency, with high frequency items resulting in strong S-complement commitments and low frequency items resulting in much weaker S-complement commitments (Juliano & Tanenhaus, 1993; though see Garnsey et al., 1997). This interaction between frequency and structural preference is explained by Juliano & Tanenhaus (1993) as occurring because the argument structure preferences of S-bias verbs must compete for activation with the regular pattern of the language – that an NP after a verb is a direct object. The ability of the S-bias verbs to overcome this competing cue depends upon frequency. Juliano & Tanenhaus (1994) present a connectionist model that shows that such interactions emerge naturally from constraint-based lexicalist models, since the models learn to represent more accurately the preferences of high frequency items. In later sections, we return to the issue of interactions between lexical frequency and "regularity" and discuss its implications for the architecture of computational models of language processing.

The CBL theory has provided an account for experimental results involving a wide range of syntactic ambiguities (e.g., Boland, Tanenhaus, Garnsey & Carlson, 1995; Garnsey et al., 1997; Juliano & Tanenhaus, 1993; Trueswell & Kim, 1998; MacDonald, 1993, 1994; Spivey-Knowlton & Sedivy, 1995; Trueswell et al., 1993; Trueswell, Tanenhaus & Garnsey, 1994; cf. MacDonald et al., 1994). As this body of experimental results has grown, there has been a need to expand the grammatical coverage of computational modeling work to match that of the most comprehensive descriptions of the CBL theory, which

have been wide in scope, but have not been computationally explicit (MacDonald et al., 1994; Trueswell & Tanenhaus, 1994). Existing computational models have focused on providing detailed constraint-based accounts of the pattern of processing preferences for particular sets of experimental results (McRae et al., 1998; Tabor et al., 1996; Spivey-Knowlton, 1996; Juliano & Tanenhaus, 1994). These models have tended to be limited syntactic processors, with each model addressing the data surrounding a small range of syntactic ambiguities (e.g., the NP/S ambiguity). This targeted approach has left open some questions about how CBL-based models "scale up" to more complicated grammatical tasks and more comprehensive samples of the language. For instance, the Juliano & Tanenhaus model learns to assign seven different verb complement types based on co-occurrence information about a set of less than 200 words. The full language involves a much greater number of syntactic possibilities and more complicated co-occurrence relationships. It is possible that the complexities of computing the fine-grained statistical relationships of the full language may be qualitatively greater than in these simple domains, or even intractable (Mitchell, Cuetos, Corley & Brysbaert, 1995). It is also possible that these targeted models are so tightly focused on specific sets of experimental data that they have acquired parameter settings that are inconsistent with other data (see Frazier, 1995). Thus, there is a need to examine whether the principles of the theory support a model that provides comprehensive syntactic coverage of the language but which still predicts fine-grained patterns of argument structure availability.

Lexicalized grammars and supertagging

In developing a broader and more formal account of psycholinguistic findings, we have drawn insights from work on statistical techniques for processing over LTAG (Srinivas & Joshi, 1999). This section introduces LTAG and representational and processing issues within it.

The idea behind LTAG is to localize the computation of linguistic structure by associating lexical items with rich descriptions that impose complex combinatory constraints in a local context. Each lexical item is associated with at least one "elementary tree" structure, which encodes the "minimal syntactic environment" of a lexical item. This includes such information as head-complement requirements, filler-gap information, tense, and voice. Figure 1 shows some of the elementary trees associated with the words of the sentence *The police officer believed the victim was lying*.[2] The trees involved in the correct parse of the sentence are highlighted by boxes. Note that the highlighted tree for *believed* specifies each of the word's arguments, a sentential complement and a noun phrase subject.

Encoding combinatory information in the lexicon rather than in supralexical rules has interesting effects on the nature of structural analysis. One effect is that the number of different descriptions for each lexical item becomes much larger than when the descriptions are less complex. For instance, the average elementary tree ambiguity for a word in Wall Street Journal text is about 47 trees (Srinivas & Joshi, 1999). In contrast, part-of-speech tags, which provide a much less complex description of words, have an ambiguity of about 1.2 tags per word in Wall Street Journal text. Thus, lexicalization increases the local ambiguity for the parser, complicating the problem of lexical ambiguity resolution. The increased lexical ambiguity is partially illustrated in Figure 1, where six out of eight words have multiple elementary tree possibilities. The flip-side to this increased lexical ambiguity, however, is that resolution of lexical ambiguity yields a representation that is effectively a parse, drastically reducing the amount of work to be done after lexical ambiguity is resolved (Srinivas & Joshi, 1999). This is because the elementary trees impose such complex combinatory constraints in their own local contexts that there are very few ways for the trees to combine once they have been correctly chosen. The elementary trees can be understood as having "compiled out" what would be rule applications in a context-free grammar system, so that once they have been correctly assigned, most syntactic ambiguity has been resolved. Thus, the lexicalization of grammar causes much of the computational work of structural analysis to shift from grammatical rule application to lexical ambiguity resolution. We refer to the elementary trees of the grammar as "supertags", treating them as complex analogs to part-of-speech tags. We refer to the process of resolving supertag ambiguity as "supertagging". One indication that the work of structural analysis has indeed been shifted into lexical ambiguity resolution is that the run-time of the parser is reduced by a factor of thirty when the correct supertags for a sentence are selected in advance of parsing.[3]

Importantly for the current work, this change in the nature of parsing has been complemented by the recent development of statistical techniques for lexical ambiguity resolution. Simple statistical methods for resolving part-of-speech ambiguity have been one of the major successes in recent work on statistical natural language processing (cf. Church & Mercer, 1993). Several algorithms tag part-of-speech with accuracy between 95% and 97% (cf. Charniak, 1993). Applying such techniques to the words in a sentence before parsing can substantially reduce the work of the parser by preventing the construction

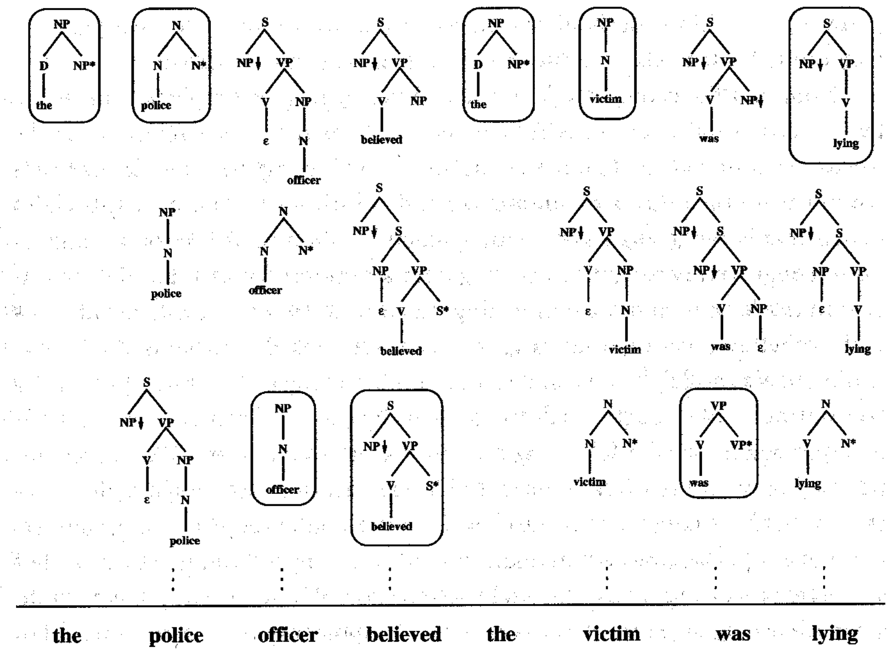

Figure 1. A partial illustration of the elementary tree possibilities for the sentence *the police officer believed the victim was lying*. Trees involved in the correct parse of the sentence are highlighted in boxes

of spurious syntactic analyses. Recently, Srinivas and Joshi (1999) have demonstrated that the same techniques can be effective in resolving the greater ambiguity of supertags. They implemented a tri-gram Hidden Markov Model of supertag disambiguation. When trained on 200,000 words of parsed Wall Street Journal text, this model produced the correct supertag for 90.9% of lexical items in a set of held out testing data.

Thus, simple statistical techniques for lexical ambiguity resolution can be applied to supertags just as they can to part-of-speech ambiguity. Due to the highly constraining nature of supertags, these techniques have an even greater impact on structural analysis when applied to supertags than when applied to part-of-speech tags. These results demonstrate that much of the computation of linguistic analysis, which has traditionally been understood as the result of structure building operations, might instead be seen as lexical disambiguation. This has important implications for how psycholinguists are to conceptualize structural analysis. It expands the potential role in syntactic analysis of simple pattern recognition mechanisms for word recognition, which have played a very limited role in classical models of human syntactic processing.

Note that the claim here is not that supertagging accomplishes the entire task of structural analysis. After elementary trees have been selected for the words in a sentence, there remains the job of connecting the trees via the LTAG combinatory operations of adjunction and substitution. The principal claim here is that in designing a system for syntactic analysis there are sound linguistic and engineering reasons for storing large amounts of grammatical information in the lexicon and for performing much of the work of syntactic analysis with something like supertagging. If such a system is also to be used as a psycholinguistic model, it is natural to predict that many of the initial processing commitments of syntactic analysis are made by a level of processing analogous to supertagging. In the following section, we discuss how an LTAG-based supertagging system resolves at the lexical level many of the same syntactic ambiguities that have concerned researchers in human sentence processing, suggesting that a supertagging system might provide a good psycholinguistic model of syntactic processing. Thus, although the question of how such a system fits into a complete language processing system is an important one, it may be useful to begin exploring the psychological implications of supertagging in advance of a complete understanding of how to design the rest of the system.[4]

A model of the grammatical aspects of word recognition using LTAG

In the remaining sections of this paper, we describe an ongoing project which attempts to use LTAG to develop a more fully-specified description of the CBL theory of human sentence processing. We argue that the notion of supertagging can become the basis of a model of the grammatical aspects of word recognition, provided that certain key adjustments are made to bring it in line with the assumptions of psycholinguistic theory (Kim, Srinivas & Trueswell, in preparation). Before introducing this model, we outline how LTAG can be used to advance the formal specification of the CBL theory.[5] We then turn to some of the findings of the model, which capture some of the major phenomena reported in the human parsing literature.

LTAG lexicalizes syntactic information in a way that is highly consistent with descriptions of the CBL theory, including the lexicalization of head-complement relations, filler-gap information, tense, and voice. The value of LTAG as a formal framework for a CBL account can be illustrated by the LTAG treatment of several psycholinguistically interesting syntactic ambiguities, e.g., prepositional phrase attachment ambiguity, the NP/S ambiguity, the reduced relative/main clause ambiguity, and the compound noun ambiguity. In all but one of these cases, the syntactic ambiguity is characterized as stemming from a lexical ambiguity.

Figure 2 presents the LTAG treatment of these ambiguities. Each of the sentence fragments in the figure ends with a syntactically ambiguous word and is accompanied by possible supertags for that word. First, the prepositional phrase attachment ambiguity is illustrated in Figure 2a. The ambiguity lies in the ability of the prepositional phrase *with the* … to modify either the noun phrase *the cop* (e.g., *with the red hair*) or the verb phrase headed by *saw* (e.g., *with the binoculars*). Within LTAG, prepositions like *with* indicate lexically whether they modify a preceding noun phrase or verb phrase. This causes prepositional phrase attachment ambiguities to hinge on the lexical ambiguity of the preposition. Similarly, the NP/S ambiguity discussed in the Introduction arises directly from the ambiguity between the elementary trees shown in Figure 2b. In this case, these trees encode the different complement-taking properties of the verb *forgot* (e.g., *the recipe* vs. *the recipe was* …). Figure 2c shows a string that could be parsed as a Noun-Noun compound (e.g., *the warehouse fires were extinguished.*) or a Subject-Verb sequence (e.g., *the warehouse fires older employees.*). In non-lexicalist grammars, this ambiguity is treated as arising from the major category ambiguity of *fires*. In LTAG, this ambiguity involves not only the category ambiguity but also a more fine-grained ambiguity regarding the previous noun *warehouse*. Due to the nature of combinatory operations of LTAG, nouns that appear as phrasal heads or phrasal modifiers are assigned different types of elementary trees (i.e., the *Alpha-/Beta-* distinction in LTAG, see Doran, Egedy, Hockey, Srinivas & Zaidel, 1994). Figure 2d illustrates the reduced relative/main clause ambiguity (e.g., *the defendant examined by the lawyer was* … vs. *the defendant examined the pistol.*). Here again, the critical features of the phrase structure ambiguity are lexicalized. For instance, the position of the gap in an object-extraction relative clause is encoded at the verb (right-hand tree in Figure 2d). This is because LTAG trees encode the number, type, and position of all verb complements, including those that have been extracted. Finally, Figure 2e illustrates a structural ambiguity that is not treated lexically in LTAG. As in Figure 2a, the preposition *with* is associated with two elementary trees, specifying verb phrase or noun phrase modification. However, in this example, both attachment possibilities involve the same tree (NP-attachment), which can modify either *general* or *secretary*. The syntactic information that distinguishes between local and non-local attachment is not specified lexically. So, within LTAG, this final example is a case of what we

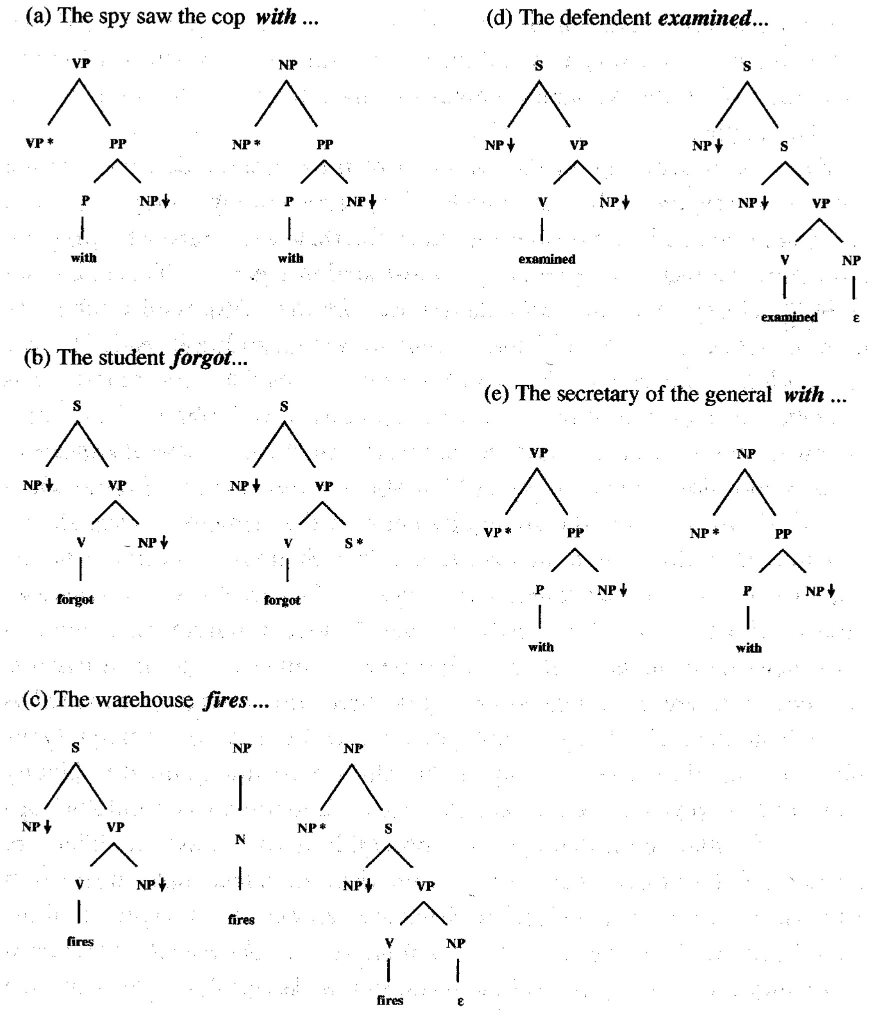

Figure 2. LTAG treatment of several psycholinguistically interesting syntactic ambiguities: (a) PP-attachment ambiguity; (b) NP/S ambiguity; (c) N/V category ambiguity; (d) reduced relative/main clause ambiguity; (e) PP-attachment ambiguity with both attachment sites being nominal.

might call true attachment ambiguity. This example illustrates the point made earlier that even when a lexical tree is selected, syntactic processing is not complete, since lexical trees need to be combined together through the operations of substitution and adjunction. In the first four examples, the selection of lexical trees leaves only a single way to combine these items. In the final example, however, multiple combinatory possibilities remain even after lexical selection.

The examples in Figure 2 illustrate the compatibility of LTAG with the CBL theory. The two frameworks lexicalize structural ambiguities in similar ways, with LTAG providing considerably more linguistic detail. This suggests that LTAG can be used to provide a more formal statement of the representational claims of the CBL theory. For instance, one can characterize the grammatical aspects of word recognition as the parallel activation of possible elementary trees. The extent to which a lexical item activates a particular elementary tree is determined by the frequency with which it has required that tree during an individual's linguistic experience. The selection of a single tree is accomplished through the satisfaction of multiple probabilistic constraints, including semantic and syntactic contextual cues. The CBL theory has traditionally focused on the activation of verb argument structure. The introduction of a wide-coverage grammar into this theory generates clear predictions about the grammatical representations of other classes of words. The same ambiguity resolution processes occur for all lexical items for which LTAG specifies more than one elementary tree.

The grammatical predictions of LTAG are worked out in an English grammar, which is the product of an ongoing grammar development project at the University of Pennsylvania (Doran et al., 1994). The grammar provides lexical descriptions for 37,000 words and handles a wide range of syntactic phenomena, making it a highly robust system. The supertagging work described in this chapter makes critical use of this grammar. The comprehensiveness of the grammar makes it a valuable tool for psycholinguistic work, by allowing formal statements about the structural properties of a large fragment of the language. In our case, it plays a critical role in our attempt to "scale up" CBL models in order to investigate the viability of such models on more complex grammatical situations than they have been tested on before.

Implementation

In this section, we describe preliminary results of a computational modeling project exploring the ability of the CBL theory to integrate the representations of LTAG. We have been developing a connectionist model of the grammatical aspects of word recognition (Kim et al., in preparation), which attempts to account for various psycholinguistic findings pertaining to syntactic ambiguity resolution. Unlike previous connectionist models within the CBL approach (McRae et al., 1998; Tabor et al., 1997; Spivey-Knowlton, 1996; Juliano

& Tanenhaus, 1994), this model has wide coverage in that it has an input vocabulary of 20,000 words and is designed to assign 304 different LTAG elementary trees to input words. The design of the model was not guided by the need to match a specific set of psycholinguistic data. Rather, we applied simple learning principles to the acquisition of a wide coverage grammar, using as input a corpus of highly-variable, naturally occurring text. Certain patterns of structural preferences and frequency effects, which are characteristic of human data, fall directly out of the model's system of distributed representation and frequency-based learning.

The model resembles the statistical supertagging model of Srinivas & Joshi (1999), which we briefly described above. We have, however, made key changes to bring it more in line with the assumptions behind the CBL framework. The critical assumptions are that human language comprehension is characterized by distributed, similarity-based representations (cf. Seidenberg, 1992) and by incremental processing of a sentence. The Srinivas and Joshi model permits the use of information from both left and right context in the syntactic analysis of a lexical item (through the use of Viterbi decoding). Furthermore, their model has a "perfect" memory, which stores the structural events involving each lexical item separately and without error. In contrast, our model processes a sentence incrementally, and its input and internal representations are encoded in a distributed fashion. Distributed representations cause each representational unit to play a role in the representation of many lexical items, and the degree of similarity among lexical items to be reflected in the overlap of their representations.

These ideas were implemented in a connectionist network, which provided a natural framework for implementing a distributed processing system.[6] The model takes as input information about the orthographic and semantic properties of a word and attempts to assign the appropriate supertag for the word given the local left context. The architecture of the model consists of three layers with feed-forward projections, as illustrated in Figure 3.

The model's output layer is a 95 unit array of syntactic features which is capable of uniquely specifying the properties of 304 different supertags. These features completely specify the components of an LTAG elementary tree: 1) part-of-speech, 2) type of "extraction," 3) number of complements, 4) category of each complement, and 5) position of complements. Each of these components is encoded with a bank of localist units. For instance, there is a separate unit for each of 14 possible parts of speech, and the correct activation pattern for a given supertag activates only one of these units (e.g., "Noun").

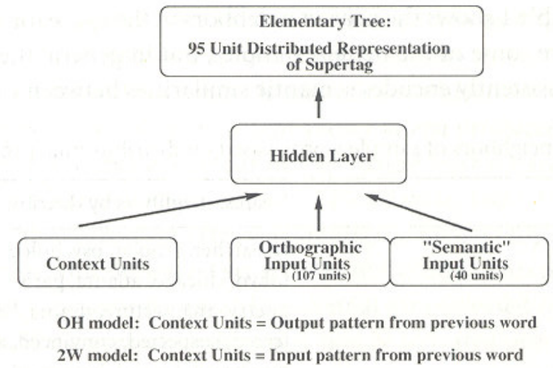

Figure 3. Architecture of the model

The model was given as input rudimentary orthographic information and fine-grained distributional information about a word. 107 of the units encoded orthographic features, namely the 50 most common three-letter word-initial segments (e.g., *ins*), the 50 most common two-letter word-final segments (e.g., *ed*), and seven properties such as capitalization, hyphenation, etc. The remaining 40 input units provide a "distributional profile" of each word, which was derived from a co-occurrence analysis.

The orthographic encoding scheme served as a surrogate for the output of morphological processing, which is not explicitly modeled here but is assumed to be providing interactive input to lexico-syntactic processes that are modeled. The scheme was chosen primarily for its simplicity – it was automatically derived and easily applied to the training and testing corpus, without requiring the use of a morphological analyzer. It was expected to correlate with the presence of common English morphological features.

Similarly, the distributional profiles were used as a surrogate for the activation of detailed semantic information during word recognition. Although space prevents a detailed discussion, we note that several researchers have found that co-occurrence-based distributional profiles provide detailed information about the semantic similarity between words (cf. Burgess & Lund, 1997; Landauer & Dumais, 1997; Schuetze, 1993). The forty-dimensional profiles used here were created by first collecting co-occurrence statistics for a set of 20,000 words in a large corpus of newspaper text.[7] The co-occurrence matrix was compressed by extracting the 40 principal components of a Singular Value Decomposition (see Kim et al., in preparation, for details). An informal inspection of the space reveals that it captures certain grammatical and semantic information. Table 1 shows the nearest neighbors in the space for some selected words. These are some of the better examples, but in general the information in the space consistently encodes semantic similarities between words.

Table 1. Nearest neighbors of sample words based on distributional profiles

Word	Nearest neighbors by distributional profile
scientist	researcher, scholar, psychologist, chemist
london	tokyo, chicago, atlanta, paris
literature	poetry, architecture, drama, ballet
believed	feared, suspected, convinced, admitted
bought	purchased, loaned, borrowed, deposited
smashed	punched, cracked, flipped, slammed
confident	hopeful, optimistic, doubtful, skeptical
certainly	definitely, obviously, hardly, usually
from	with, by, at, on

We implemented two variations on the basic architecture described above, which gave the model an ability to maintain information over time so that its decisions would be context sensitive. The first variation expanded the input pattern to provide on each trial a copy of the input pattern from the previous time step along with the current input. This allowed the network's decisions about the current input to be guided by information about the preceding input. We will call this architecture the "two-word input" model (2W). The second variation provided simple recurrent feedback from the output layer to the hidden layer so that on a given trial the hidden layer would receive the previous state of the output layer. This again allowed the model's decision on a given trial to be contingent on activity during the previous trial. We call this architecture the "output-to-hidden" architecture (OH). For purposes of brevity, we discuss only the results of the 2W architecture. In all statistical analyses reported here, the OH architecture produced the same effects as the 2W architecture.

The model was trained on a 195,000 word corpus of Wall Street Journal text, which had been annotated with supertags. The annotation was done by translating the annotations of a segment of the Penn Treebank (Marcus, Santorini & Marcinkiewicz, 1993) into LTAG equivalents (Srinivas, 1997). During training, for each word in the training corpus, the appropriate orthographic units and distributional profile pattern were activated in the input layer. The input activation pattern was propagated forward through the hidden layer to the output layer. Learning was driven by back propagation of the error between the model's output pattern and the correct supertag pattern for the current word (Rumelhart, Hinton & Williams, 1986).

We tested the overall performance of the model by examining its supertagging accuracy on a 12,000 word subset of the training corpus that was held out of training. The network's syntactic analysis on a given word was considered to be the supertag whose desired activation pattern produced the lowest error with respect to the model's actual output (using least squares error). On this metric, the model guessed correctly on 72% of these items. Using a slightly relaxed metric, the correct supertag was among the model's top three choices (the three supertags with the lowest error) 80% of the time. This relaxed metric was used primarily to assess the model's potential for increased overall accuracy in future work; if the correct analysis was highly activated even when it was not the most highly activated analysis, then future changes might be expected to increase the model's overall accuracy (e.g., improvements to the quality of the input representation). Accuracy for basic part of speech on the relaxed metric was 91%. The performance of the network can be compared to 79% accuracy for a "greedy" version of the tri-gram model of Srinivas & Joshi (1999), which was trained on the same corpus. The greedy version eliminated the previously mentioned ability of the original model to be influenced by information from right context in its decisions about a given word.

Although these results indicate that the model acquired a substantial amount of grammatical knowledge, the main goal of this work is to examine the relationship between the model's operation and human behavioral patterns, including the patterns of misanalysis characteristic of human processing. In pursuing this goal, we measure the model's degree of commitment to a given syntactic analysis by the size of its error to that analysis relative to its error to other analyses. We make the linking hypothesis that reading time elevations due to misanalysis and revision in situations of local syntactic ambiguity should be predicted by the model's degree of commitment to the erroneous syntactic analysis at the point of ambiguity. For example, in the NP/S ambiguity of Example 1, the model's degree of commitment to the NP-complement analysis over the S-complement analysis should predict the amount of reading time elevation at the disambiguating region *was in…* . Examination of the model's processing of syntactic ambiguities revealed patterns characteristic of human processing.

Modeling the NP/S ambiguity

One pattern of behavioral data that our model aims to account for is the pattern of processing difficulty around the NP/S ambiguity, illustrated by *The chef forgot the recipe was in the back of the book* (discussed in the Introduction as

(1)). In (1a), comprehenders can initially treat the noun phrase *the recipe* as either a NP-complement of *forgot* or the subject of a sentential complement. Numerous experiments have found that readers of locally ambiguous sentences like 1a often erroneously commit to a NP-complement interpretation (Holmes et al., 1989; Ferreira & Henderson, 1990; Trueswell et al., 1993; Garnsey et al., 1997).

Several experiments have found that the general processing bias toward the NP-complement is modulated by the structural bias of the main verb (Trueswell et al., 1993; Garnsey et al., 1997). Erroneous commitments to the NP-complement interpretation are weakened or eliminated when the main verb has a strong S-bias (e.g., *claimed*). Similar effects have also been found when verb bias information is introduced to processing through a lexical priming technique (Trueswell & Kim, 1998). Thus, the language processing system appears to be characterized simultaneously by an overall bias toward the NP-complement analysis and by the influence of the lexical preferences of S-bias verbs.

The coexistence of these two conflicting sources of guidance may be explained in terms of "neighborhoods of regularity" in the representation of verb argument structure (Seidenberg, 1992; Juliano & Tanenhaus, 1994). NP-complement and S-complement verbs occupy neighborhoods of representation, in which the NP-complement neighborhood dominates the "irregular" S-complement neighborhood, due to greater membership. The ability of S-complement items to be represented accurately is dependent on frequency. High frequency S-complement items are accurately represented, but low frequency S-complement items are overwhelmed by their dominant NP-complement neighbors. Juliano & Tanenhaus (1993) found evidence in support of this hypothesis in a study in which the ability of verb bias information to guide processing was characterized by an interaction between the frequency and the subcategory of the main verb. The ability of S-complement verbs to guide processing commitments was correlated with the verb's lexical frequency. Low frequency S-complement verbs allowed erroneous commitments to the NP-complement analysis in spite of the verb's bias, while high frequency S-complement items caused rapid commitments to the correct S-complement analysis.

Our model provides such a neighborhood-based explanation of the human processing data for NP/S ambiguities. We presented the model with NP/S ambiguous fragments, such as *The economist decided ...*, which contained either a verb that strongly tended to take S-complements in the training corpus or strongly tended to take NP-complements. The model assigned either a NP- or S-complement analysis to 96% of such verbs, indicating that it clearly recognized NP/S verbs. In resolving the NP/S ambiguity, the model showed a general bias toward the NP-complement structure, which can be overcome by lexical information from high frequency S-complement verbs. All NP-biased verbs were correctly analyzed, but S-biased verbs were misanalyzed on 9 of 14 items, with 8 of 9 misanalyses being to the NP-complement. The dominance of the NP-complement analysis, however, is modulated by the frequency of exposure to S-complement items. The model accurately subcategorized S-biased verbs when they were high in frequency (5 of 7) but was highly inaccurate on low frequency items (none were correctly classified; 6 of 7 were mis-analyzed as NP-complement verbs).

The model's frequency-by-subcategory interaction arises from its system of distributed representation and frequency sensitive learning. S-complement verbs and NP-complement verbs have a substantial overlap in input representation, due to distributional and orthographic similarities (-ed, -ng, etc.) between the two types of verbs and the fact that S-complement verbs are often NP/S ambiguous. NP-complement tokens dominate S-complement tokens in frequency by a ratio of 4 to 1, causing overlapping input features to be more frequently associated with the NP-complement output than the S-complement output during training. The result is that a portion of the input representation of S-complement verbs becomes strongly associated with the NP-complement output, causing a tendency for the model to misanalyze S-complement items as NP-complement items. The model is able to identify non-overlapping input features that distinguish S-complement verbs from their dominant neighbors, but its ability to do so is affected by frequency. When S-complement verbs are seen in high frequencies, their distinguishing features are able to influence connection weights enough to allow accurate representation; however, when S-complement verbs are seen in low frequencies, their NP-complement-like input features dominate their processing. The explanation here is similar to the explanation given by Seidenberg & McClelland (1989) for frequency-by-regularity interactions in word naming (e.g., the high frequency irregularity of *have* vs. the regularity of *gave, wave, save*) and past tense production.

The theoretical significance of this interaction lies partly in its emergence in a comprehensive model, which is designed to resolve a wide range of syntactic ambiguities over a diverse sample of the language. These results provide a verification of conclusions drawn by Juliano & Tanenhaus (1994) from a much simpler model, which acquired a similar pattern of knowledge about NP-complement and S-complement verbs from co-occurrence information about verbs and the words that follow them. It is important to provide such follow-up work for Juliano & Tanenhaus (1994), because their simplifications of the domain were extreme enough to allow uncertainty about the scalability of their results. Although their training materials were drawn from naturally occurring text (Wall Street Journal and Brown corpus), they sampled only a subset of the verbs in that text and the words occurring after those verbs. S-complement tokens were more common in their corpus than in the full language, and only past-tense tokens were sampled. This constitutes a substantial simplification of the co-occurrence information available in the full language. In our sample of the Wall Street Journal corpus, non-auxiliary verbs account for only 10.8% of all tokens, suggesting that the full language may contain many co-occurrence events that are "noise" with respect to the pattern detected by the Juliano & Tanenhaus (1994) model. For instance, as Juliano & Tanenhaus observe, their domain restricts the range of contexts in which the determiner *the* occurs, obscuring the fact that in the full language, *the* often introduces a subject noun phrase rather than an object noun phrase. It is conceivable that the complexity of the full language would obscure the pattern of co-occurrences around the NP/S ambiguity sufficiently to prevent a comprehensive constraint-based model from acquiring the pattern of knowledge acquired by the Juliano & Tanenhaus (1994) model. Our results demonstrate that the processing and representational assumptions that allow constraint based models to naturally express frequency-by-regularity interactions are scalable – they continue to emerge when the domain is made very complex.

Modeling the noun/verb lexical category ambiguity

Another set of behavioral data that our model addresses is the pattern of reading times around lexical category ambiguities like that of *fires* in (4).

(4) a. *the warehouse* fires *burned for days.*
    b. *the warehouse* fires *many workers every spring.*

The string *warehouse fires* can be interpreted as a subject-verb sequence (4a) or a compound noun phrase (4b). This syntactic ambiguity is anchored by the lexical ambiguity of *fires*, which can occur as either a noun or a verb.

Several experiments have shown that readers of sentences like (4a) often commit erroneously to a subject-verb interpretation, as indicated by processing difficulty at the next word (*burned*), which is inconsistent with the erroneous interpretation and resolves the temporary ambiguity. Corley (1998) has shown that information about the category bias of the ambiguous word is rapidly employed in the resolution of this ambiguity. When the ambiguous word is

one that tends statistically to be a verb, readers tend to commit erroneously to the subject-verb interpretation, but when the word tends to occur as a noun, readers show no evidence of misanalysis. MacDonald (1993) has found evidence of more subtle factors, including the relative frequency with which the preceding noun occupies certain phrase-structural positions, the frequency of co-occurrence between the preceding noun and ambiguous word, and semantic fit information. Most important for the current work, MacDonald found that when the ambiguous word was preceded by a noun that tended to occur as a phrasal head, readers tended to commit to the subject-verb interpretation. However, when the preceding noun tended to occur as a noun modifier, readers tended to commit immediately to the correct noun-noun compound analysis. The overall pattern of data suggests a complex interplay of constraints in the resolution of lexical category ambiguity. Lexically specific information appears to be employed very rapidly and processing commitments appear to be affected by multiple sources of information, including subtle cues like the modifier/head likelihood of a preceding noun.

Like human readers, our model shows sensitivity to both lexical category bias and fine-grained contextual cues when processing locally ambiguous fragments like *the warehouse fires.* We presented the model with fragments ending in noun/verb ambiguous verbs (e.g., *the emergency plans*). The ambiguous words were either noun biased (e.g., *plans*), verb-biased (e.g., *pay*), or equi-biased (e.g., *bid*). The preceding noun was either one that tended to occur as a phrasal head in the training corpus (e.g., *division*) or one that tended to occur as a noun modifier in the corpus (e.g., *emergency*). Lexical bias was determined by frequency properties in the training corpus.

The model clearly recognized the target words as nouns and verbs, as indicated by the fact that 97% of the test items were assigned either a noun supertag or a verb supertag. More subtle aspects of the model's operation were revealed by an examination of the activation values of the noun and verb part-of-speech units separately from the rest of the output layer. The model showed strong commitments to the contextually supported category when that category was either the dominant sense of a biased word or when the word was equi-biased – the contextually supported unit had superior activation in 90% of such cases. In contrast, the model had difficulty activating the contextually supported category when it was the subordinate category of a biased word – showing superior activation for the contextually supported category in only 35% of such cases. Thus, context and lexical bias interacted such that the model showed a strong tendency to activate a contextually-supported pattern when it was either the dominant pattern or had an equally frequent alternative, but when context supported the subordinate pattern, the model was unable to activate this pattern.

Interestingly, this interaction resembles the "subordinate bias" effect observed in the semantic aspects of word recognition (Duffy, Morris & Rayner, 1988). When semantically ambiguous words are encountered in biasing contexts, the effects of context depend on the nature of the word's bias. When preceding discourse context supports the subordinate sense of a biased ambiguous word, processing difficulty occurs. When context supports the dominant sense or when it supports either sense of an equi-biased word, no processing difficulty occurs. Our model shows a qualitatively identical effect with respect to category ambiguity. We take this as further support for the idea, central to lexicalist theories, that lexical and syntactic processing obey many of the same processing principles. On the basis of this kind of effect in the model, we predict that human comprehenders should show subordinate bias effects in materials similar to the ones used here. Furthermore, because the subordinate bias effects found here are quite natural given the model's system of representation and processing, we would expect similar effects to arise in the model and in humans with respect to other syntactic ambiguities that are affected by local left context (see Trueswell, 1996, for similar predictions about subordinate bias effects involving the main clause/relative clause ambiguity).

General discussion

We have attempted to advance the grammatical coverage and formal specification of Constraint-based Lexicalist models of language comprehension. A convergence of perspectives between CBL theory in psycholinguistics and work in theoretical and computational linguistics has supported and guided our proposals. We have attempted to give a concrete description of the syntactic aspects of the CBL theory by attributing to human lexical knowledge the grammatical properties of a wide coverage Lexicalized Tree Adjoining Grammar (Doran et al., 1994). In developing a processing model, we have drawn insight from work on processing with LTAG which suggests that statistical mechanisms for lexical ambiguity resolution may accomplish much of the computation of parsing when applied to rich lexical descriptions like those of LTAG (Srinivas & Joshi, 1999). We have incorporated these ideas about grammar and processing into a psychologically motivated model of the grammatical aspects of word recognition, which is wide in grammatical coverage.

The model we describe is general in purpose; it acquires mappings between a large sample of the lexical items of the language and a large number of rich grammatical representations. Its design does not target any particular set of syntactic ambiguities. Nevertheless, it qualitatively captures subtle patterns of human processing data, such as the frequency-by-regularity interaction in the NP/S ambiguity (Juliano & Tanenhaus, 1993) and the use of fine-grained contextual cues in resolving lexical category ambiguities (MacDonald, 1993).

The wide range of grammatical constructions faced by the model and the diversity of its sample of language include much of the complexity of the full language and support the idea that constraint-based models of sentence processing are viable, even on a large grammatical scale. The model provides an alternative to the positions of Mitchell et al. (1995) and Corley & Crocker (1996), which propose statistical processing models with only coarse-grained parameters such as part-of-speech tags. Their argument is that the sparsity of some statistical data causes the fine-grained parameters of constraint-based models to be "difficult to reliably estimate" (Corley & Crocker, 1996) and that the large number of constraints in constraint-based models causes the management of all these constraints to be computationally intensive. Such arguments assume that a coarse-grained statistical model is more viable and more "compact" than a fine-grained model.

The issue of whether fine-grained statistical processing is viable may hinge on some basic computational assumptions. The observation that sparsity of statistical data affects the performance of statistical processing systems is certainly valid. But there are a number of reasons why this does not support arguments against fine-grained statistical processing models. First, there is a large class of statistical processing models, including connectionist systems like the one used here, that are well suited to the use of imperfect cues. For instance, a common strategy employed by statistical NLP systems to deal with sparse data is to "back off" to statistics of a coarser grain. This is often done explicitly, as in verb subcategorization methods, where decisions are conditionalized on lexical information (individual verbs) when the lexical item is common, but are conditionalized on (backed off to) basic category information (all verbs), when the lexical item is rare (Collins, 1995). In connectionist systems like ours, statistical back-off is the flip-side of the network's natural tendency to generalize but also to be guided by fine-grained cues when those cues are encountered frequently. Fine grained features of a given input pattern are able to influence behavior when they are encountered frequently, because they are given repeated opportunities to influence connection weights. When such fine-grained features are not encountered often enough, they are overshadowed by coarser-grained input features, which are by their very nature more frequent. Systems like our model can be seen as discovering back-off points. We argue that systems that do such backing off are the appropriate class of system for modeling much of sentence processing. As a back-propagation learning system with multiple grammatical tasks competing for a limited pool of processing resources, our model is essentially built to learn to ignore unreliable cues.

Thus, the interaction between frequency and subcategory that we have discussed emerges naturally in the operation of statistical processing devices like the model described here. Fine-grained information about S-complement verbs is able to guide processing when it is encountered often enough during training to influence connection weights in spite of the dominance of NP-complement signals. The ability of Head/Modifier likelihood cues about nouns to influence connection weights is similarly explained.

In general, we view the sparsity of data as an inescapable aspect of the task of statistical language processing rather than as a difficulty that a system might avoid by retreating to more easily estimable parameters. Even part-of-speech tagging models like Corley & Crocker's (1996) include a lexical component, which computes the likelihood of a lexical item given a candidate part-of-speech for that word, and their model is therefore affected by sparsity of data for individual words – this is true for any tagger based on the dominant Hidden Markov Model framework. Furthermore, as mentioned earlier, work in statistical NLP has increasingly indicated that lexical information is too valuable to ignore in spite of the difficulties it may pose. Techniques that count lexically specific events have generally out-performed techniques that do not, such as statistical context-free grammar parsing systems (see Marcus, this volume). It seems to us that, given a commitment to statistical processing models in general, there is no empirical or principled reason to restrict the granularity of statistical parameters to a particular level, such as the part-of-speech tags of a given corpus. Within the engineering work on part-of-speech tagging, there are a number of different tag-sets, which vary in the granularity of their tags for reasons unconnected to psychological research, so that research does not motivate a psychological commitment to any particular level of granularity. Furthermore, the idea that the language processing system should be capable of counting statistical events at only a single level of granularity seems to be an assumption that is inconsistent with much that is known about cognition, such as the ability of the visual processing system to combine probabilistic cues from many levels of granularity in the recognition of objects. The solution to the data sparsity problem, as manifested in humans and in successful engineer-

ing systems, is to adopt the appropriate learning and processing mechanisms for backing off to more reliable statistics when necessary.

We have argued that the complexities of statistical processing over fine grained lexical information do not warrant the proposal of lexically-blind processing mechanisms in human language comprehension. Although the complexities may be unfamiliar, they are tractable, and there are large payoffs to dealing with them. An increasingly well understood class of constraint-satisfaction mechanisms is well suited to recognizing fine-grained lexical patterns and also to backing off to coarser-grained cues when fine-grained data is sparse. The modeling work described here and research in computational linguistics suggests that such mechanisms, when applied to the rich lexical representations of lexicalized grammars, can accomplish a substantial amount of syntactic analysis. Furthermore, the kind of mechanism we describe shows a pattern of processing that strongly resembles human processing data, suggesting that such mechanisms are good models of human language processing.

Acknowledgements

This work was partially supported by National Science Foundation Grant SBR-96-16833; the University of Pennsylvania Research Foundation; and the Institute for Research in Cognitive Science at the University of Pennsylvania (NSF-STC Cooperative Agreement number SBR-89-20230). The authors thank Marian Logrip for assistance in the preparation of this paper and thank Paola Merlo, Suzanne Stevenson, and two anonymous reviewers for helpful comments on the paper.

Notes

1. The amount of syntactic structure that is lexically generated goes beyond the classical notion of argument structure. In lexicalized grammar formalisms such as LTAG, the entire grammar is in the lexicon. For instance, the attachment site of a preposition can be treated as a lexically specific feature. Noun attaching prepositions and verb attaching prepositions have different senses. We will discuss this in further detail in the following sections.

2. The down-arrows and asterisks in the trees mark nodes at which trees make contact with each other during the two kinds of combinatory operations of Tree Adjoining Grammar, substitution and adjunction. Down-arrows mark nodes at which the substitution operation occurs, and asterisks mark footnodes, which participate in the adjunction operation. The

details of the combinatory operations of TAG are beyond the scope of this chapter. See Joshi and Schabes (1996) for a discussion.

3. This is based on run-times for a sample of 1300 sentences of Wall Street Journal text, reported by Srinivas and Joshi (1999). Running the parser without supertagging took 120 seconds, while running it with correct supertags pre-assigned took 4 seconds.

4. Srinivas (1997) has suggested that this can be done by a process that is simpler than full parsing. He calls this process "stapling".

5. Of course, formal specification of this theory can be achieved by using other lexicalized grammatical frameworks, e.g., LFG (Bresnan & Kaplan, 1982), HPSG (Pollard & Sag, 1994), CCG (Steedman, 1996).

6. This is not to say that left-to-right processing and overlapping representations cannot be incorporated into a symbolic statistical system. However, most attempts within psycholinguistics to incorporate these assumptions into a computationally explicit model have been made within the connectionist framework (e.g., Elman, 1990; Juliano & Tanenhaus, 1994; Seidenberg & McClelland, 1989). By using a connectionist architecture for the current model, we are following this precedent and planning comparisons with existing modeling results.

7. For each of the 20,000 target words, we counted co-occurrences with a set of 600 high frequency "context" words in 14 million words of Associated Press newswire. Co-occurrences were collected in a six-word window around each target word (three words to either side of the word).

References

Boland, J.E., Tanenhaus, M.K., Garnsey, S.M. & Carlson, G.N. (1995). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory & Language, 34*, 774–806.

Burgess, C. & Lund, K. (1997). Modeling parsing constraints with high-dimensional context space. *Language & Cognitive Processes, 12*, 177–210.

Charniak, E. (1993). *Statistical language learning*. Cambridge, MA: MIT Press.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Church, K. & Mercer, R. (1993). Introduction to the special issue on computational linguistics using large corpora. *Computational Linguistics, 19*, 1–24.

Corley, S. (1998). A Statistical Model of Human Lexical Category Disambiguation. Unpublished doctoral dissertation, University of Edinburgh, Edinburgh, UK.

Corley, S. & Crocker, M.W. (1996). Evidence for a tagging model of human lexical category disambiguation. In *Proceedings of the 18th Annual Conference of the Cognitive Science Society*.

Crocker, M.W. (1994). On the nature of the principle-based sentence processor. In C. Clifton, L. Frazier & K. Rayner (Eds.), *Perspectives on sentence processing*, pp. 245–266. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Doran, C., Egedi, D., Hockey, B.A., Srinivas, B. & Zaidel, M. (1994). XTAG system – a wide coverage grammar for English. In *Proceedings of the 17th International Conference on Computational Linguistics (COLING '94)*. Kyoto, Japan.

Duffy, S.A., Morris, R.K. & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory & Language, 27*, 429–446.

Elman, J. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

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

Ferreira, F. & Henderson, J.M. (1990). The use of verb information in syntactic parsing: A comparison of evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 555–568.

Ford, M., Bresnan, J. & Kaplan, R.M. (1982). A competence-based theory of syntactic closure. In J. Bresnan (Ed.), *The mental representation of grammatical relations*, pp. 727–796. Cambridge, MA: MIT Press.

Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research, 24*, 437–468.

Frazier, L. (1989). Against lexical generation of syntax. In W.D. Marslen-Wilson (Ed.), *Lexical Representation and Process*. Cambridge, MA: MIT Press.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. Bloomington, IN: Indiana University Linguistics Club.

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. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1–76.

Gross, M. (1984). Lexicon-grammar and the syntactic analysis of French. In *Proceedings of the 10th International Conference on Computational Linguistics (COLING '84)*, Stanford, CA.

Holmes, V.M., Stowe, L. & Cupples, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language, 28*, 668–689.

Joshi, A. & Schabes, Y. (1997). *Tree-adjoining grammars*. In Rozenberg, G. and Salomaa, A. (Eds.). Handbook of Formal Languages, Volume 3, pp. 63–124. Berlin, New York: Springer-Verlag.

Juliano, C. & Tanenhaus, M.K. (1994). A constraint-based lexicalist account of the subject/object attachment preference. *Journal of Psycholinguistic Research, 23*, 459–471.

Juliano, C. & Tanenhaus, M.K. (1993). Contingent frequency effects in syntactic ambiguity resolution. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

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

Kaplan, R. & Bresnan, J. (1982). Lexical functional grammar: A formal system of grammatical representation. In J. Bresnan (Ed.), *The Mental Representation of Grammatical Relations*, pp. 173–281. Cambridge, MA: The MIT Press.

Kim, A., Srinivas, B. & Trueswell, J. (in preparation). A computational model of lexico-syntactic processing during language comprehension: syntactic analysis through word recognition.

Landauer, T.K. & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction & representation of knowledge. *Psychological Review, 104*, 211–240.

Lewis, R.L. (1993). *An architecturally-based theory of human sentence comprehension.* Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Lund, K., Burgess, C. & Atchley, R. (1995). Semantic and associative priming in high-dimensional semantic space. In *Proceedings of the 17th Annual Conference of the Cognitive Science Society.*

MacDonald, M. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 9*, 157–201.

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

MacDonald, M.C., Pearlmutter, N.J. & Seidenberg, M.S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.

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

McRae, K., Spivey-Knowlton, M.J. & Tanenhaus, M.K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 38*, 283–312.

Mitchell, D.C. (1989). Verb-guidance and other lexical effects in parsing. *Language and Cognitive Processes, 4*, 123–154.

Mitchell, D.C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mitchell, D.C., Cuetos, F., Corley, M.M.B. & Brysbaert, M. (1995). Exposure-based models of human parsing. *Journal of Psycholinguistic Research, 24*, 469–488.

Perfetti, C.A. (1990). The cooperative language processors: Semantic influences in an autonomous syntax. In G.B. Flores d'Arcais, D.A. Balota & K. Rayner (Eds.), *Comprehension processes in reading*. Hillsdale, NJ: Erlbaum.

Pollard, C. & Sag, I. (1994). *Head-driven Phrase Structure Grammar*. Chicago, IL: University of Chicago Press.

Prince, A. & Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science, 275*, 1604–1610.

Pritchett, B.L. (1992). *Grammatical competence and parsing performance*. Chicago, IL: The University of Chicago Press.

Rayner, K., Carlson, M. & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing. *Journal of Verbal Learning and Verbal Behavior, 22*, 358–374.

Rayner, K. & Frazier, L. (1987). Parsing temporarily ambiguous complements. *Quarterly Journal of Experimental Psychology, 39A*, 657–673.

Rumelhart, D., Hinton, G. & Williams, R. (1986). Learning representations by back-propagating errors. *Nature, 323*, 533–536.

Schütze, H. (1993). Word space. In S. Hanson, J. Cowan, & C. Giles (Eds.), *Neural Information Processing Systems 5*. San Mateo, CA: Morgan Kaufmann.

Seidenberg, M.S. (1992). Connectionism without tears. In S. Davis (Ed.), *Connectionism: Theory & Practice*. New York, NY: Oxford University Press.

Seidenberg, M.S. & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523–568.

Sleator, D. & Temperley, D. (1991). Parsing English with a link grammar. *Technical report CMU-CS-91-196, Department of Computer Science, Carnegie Mellon University.*

Spivey-Knowlton, M.J. (1996). *Integration of visual and linguistic information: Human data and model simulations.* Unpublished doctoral dissertation, University of Rochester, Rochester, NY.

Spivey-Knowlton, M.J. & Sedivy, J. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition, 55*, 227–267.

Srinivas, B. (1997). *Complexity of lexical descriptions and its relevance to partial parsing.* Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia, PA.

Srinivas, B. & Joshi, A.K. (1999). Supertagging: An approach to almost parsing. *Computational Linguistics, 25*(2), 237–265.

Steedman, M. (1996). *Surface Structure and Interpretation*. Cambridge, MA: MIT Press.

Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research, 23*, 295–322.

Tabor, W., Juliano, C. & Tanenhaus, M. (1996). A dynamical system for language processing. In *Proceedings of the 18th Annual Conference of the Cognitive Science Society.*

Trueswell, J. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language, 35*, 566–585.

Trueswell, J.C. & Kim, A.E. (1998). How to prune a garden-path by nipping it in the bud: Fast-priming of verb argument structures. *Journal of Memory and Language, 39*, 102–123.

Trueswell, J. & Tanenhaus, M. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In C. Clifton, K. Rayner & L. Frazier (Eds.), *Perspectives on sentence processing*, pp. 155–180. Hillsdale, NJ: Lawrence Erlbaum Associates.

Trueswell, J., Tanenhaus, M., & Garnsey, S. (1994). Semantic Influences on Parsing: Use of Thematic Role Information in Syntactic Ambiguity Resolution. *Journal of Memory and Language, 33*, 285–318.

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