Multiple Constraints in Syntactic Ambiguity Resolution: A Connectionist Account of Psycholinguistic Data

Curt Burgess
Department of Psychology
University of California-Riverside
Riverside, CA 92521-0426
curt@cassandra.ucr.edu

Kevin Lund
Department of Psychology
University of California-Riverside
Riverside, CA 92521-0426
kevin@cassandra.ucr.edu

Abstract

We implement a constraint satisfaction connectionist style model that accounts for data from three psycholinguistics experiments investigating the gardenpath effect with reduced relative constructions. Normative data was collected on the stimuli used in experiments by Burgess and Tanenhaus (1992) and Ferreira and Clifton (1986) and this data served as the input for the simulation. We have demonstrated with this set of simulations that a plausible theoretical framework for a range of these results is a hierarchical connectionist network which is sensitive to a number of constraints inherent in the input stimuli. The model accounts for the top-down effect of context, the contribution of the bottom-up morphological frequency asymmetry of the verb, and the probabilistic nature of the disambiguating preposition. These effects are sensitive to the timecourse of processing as well. The pattern of results from the psycholinguistic data suggest that syntactic processing is a confluence of multiple constraints that represent both bottom-up and top-down influences in processing. These results are incompatible with a deterministic parsing model. The hierarchical connectionist style model presented in this paper is sensitive to the range of constraints discussed above and is offered as a more adaptive theoretical model that can capture the domain of effects found in the literature encompassing local syntactic ambiguity resolution.

Introduction

The difficulty in developing a model of language comprehension is specifying the various components of the system and the nature of interaction among these components. The resolution of syntactic and lexical ambiguity has played a central role in addressing this difficulty because of the special problems that ambiguity poses for the efficient functioning of the language processor. With respect to syntactic processing, the language comprehension system must make a commitment to a particular syntactic analysis at some point in time. The problem with syntactic ambiguity is not so much that a sentence may have multiple interpretations, but that most sentences have local ambiguities, that is, there are points in the sentence where more than one syntactic interpretation is possible.

There has been an ongoing controversy about which point during the comprehension process can contextual information, in the form of semantic or discourse constraints, influence the parsing process. Two general views have predominated this controversy. McClelland (1988) and Marslen-Wilson (1975) claim that the language comprehension system operates in an interactive fashion, such that all sources of information serve to constrain the syntactic analysis of the sentence. Alternatively, Frazier (1978; also Ferreira & Clifton, 1986) has argued that the parser makes a commitment to a syntactic structure according to syntactic strategies, such as Minimal Attachment, and without any guidance from non-syntactic processors. According to Minimal Attachment, the parser builds the surface structure by attaching each new constituent to the preceding phrase marker using the fewest possible nodes (see Figure 1a). A reduced-relative construction, *The evidence examined by...*, will gardenpath because, ultimately, it will require the more complex structure (see Figure 1b). Thus, according to this model, the parser is inherently serial and autonomous in its processing. Ferreira and Clifton

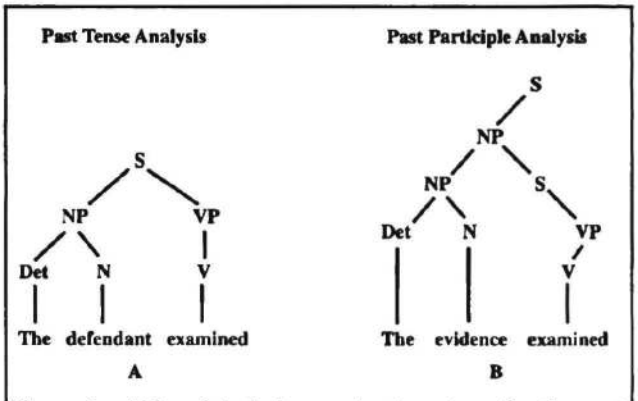

Figure 1. Abbreviated phrase structure trees for the past-tense and past-participle analyses.

reasoned that a strong constraint such as the implausibility of an inanimate subject in a reduced-relative sentence like (1b) should eliminate the processing load (gardenpath) usually associated with reduced relative sentences like (1a).

(1a.) The defendant examined by the lawyer turned out to be unreliable.
(1b.) The evidence examined by the lawyer turned out to be unreliable.

However, their results were consistent with the Minimal Attachment proposal; animacy of the subject had no effect on reading times in the disambiguating by-phrase. Burgess and Hollbach (1988) proposed a computational model based on earlier theoretical work (Cottrell, 1988; Simpson & Burgess, 1985; Tanenhaus & Carlson, 1988; Tanenhaus, Burgess,

1985; Tanenhaus & Carlson, 1988; Tanenhaus, Burgess, D'Zmura, & Carlson, 1987) that demonstrated the plausibility of accounting for the gardenpath phenomena as a function of the frequency asymmetry of the past-tense and past-participle morphological verb forms, as well as other constraints. In a series of experiments, Burgess and Tanenhaus (1992) used sentences that minimized the past-tense bias more effectively than those used by Ferreira and Clifton. Using both sets of materials in a self-paced single-word reading time task, it was found that the gardenpath occurred in both animate and inanimate conditions. However, when the sentences were presented two words at a time (e.g., *The ransom / paid by ...*) they found that gardenpaths occurred in the inanimate, past-participle biased condition with Ferreira and Clifton's materials, but not with the more constraining inanimate materials. They concluded that the disambiguating preposition must have an opportunity to be integrated with the verb (as one would expect in normal parafoveal reading) in order for the disambiguating information to constrain processing. In a subsequent experiment, Burgess, Tanenhaus, and Hoffman (1994) confirmed the importance of the timeliness of integrating the disambiguating preposition with the ambiguous verb. Syntactic gardenpaths that did not occur previously with the more constraining, past-participle biased sentences (*The ransom paid by...*) now occurred when the verb was presented with a longer preposition that would delay the disambiguation process (*The ransom paid along...*). These results suggest that the parser utilizes a variety of constraints, including semantic bias, morphological verb-form information, the nature of the disambiguating preposition, and that the timecourse of the availability of this information is crucial (also see Burgess & Hollbach, 1988; McDonald, 1994; Trueswell, Tanenhaus, & Garnsey, 1993).

In this paper we describe a set of computer simulations that replicate results from three experiments using a constraint satisfaction connectionist network. These results implicate three sources of information that are crucial to syntactic ambiguity resolution: 1) thematic fit of noun and verb, 2) morphological asymmetries of the ambiguous verb, and 3) the disambiguating action of the preposition.

Structure of a Parallel System

The implementation is a localist style of connectionist model, implemented on the Rochester Connectionist Simulator. The model is a hierarchically structured five layer network with layers corresponding to the major language subsystems (see Figure 2). The lexical level receives input and the link weights (L1 ... L8) connecting the semantic (nodes 12 & 13), syntactic (nodes 14 & 15), and discourse level (21, 22, 23, 24; although only 22 & 23 are linked to the preposition) are modified as a function of the input. The input takes the form of values that

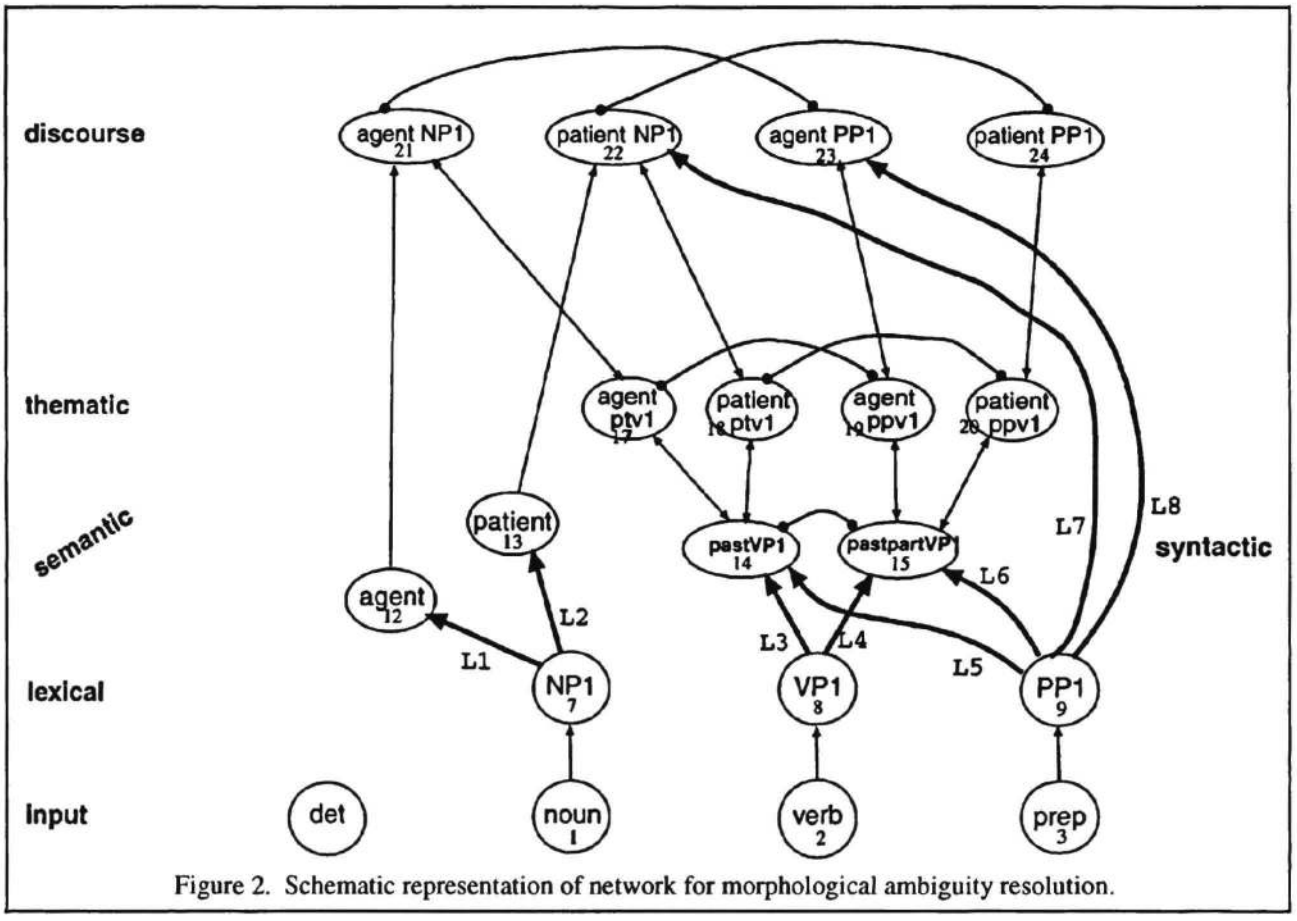

Figure 2. Schematic representation of network for morphological ambiguity resolution.

correspond to either human judgements about the relationship of the sentence constituents or the statistical occurrence of the verb (this is described in detail later). Link weights are free to vary from -1 to +1. Laterally competing representations have weak inhibitory connections (-0.1). The model is designed with a slight bias toward bottom-up priority. All excitatory links are weighted +1, except as follows: bottom-up links from the thematic roles (17, 18, 19, 20) +.750; top-down links from discourse elements (21, 22, 23, 24) +.500; top-down links from thematic roles, +.500. The bottom-up activation from the preposition (9) is weighted +1 on L5 and L6 and +.500 on L7 and L8. The constraint offered by the preposition (see below) is, thus, stronger for the syntactic nodes (14 & 15) than on the discourse nodes (22 & 23).

There are two major types of nodes in the network: input nodes and processing nodes (which are further subdivided into pure processing and processing/criterion nodes). Activation is sent to the input nodes in a sequence determined by the experimental condition. Once activated, their activation, A, is multiplied by a constant, d, at every time step; this constant, when less than one, results in a decay of input (see Equation 1). The remaining nodes have their activation determined at every time step as the sum of the activations of the nodes to which they are connected. The input into any node is weighted by the link strengths, w, of that input (see Equation 2).

$$\text{Equation 1: } A_{it} = dA_{it-1}$$
$$\text{Equation 2: } A_{it} = \sum A_{jt} w_{ij}$$

Criterion nodes have the additional property of being used to calculate a distance metric for the network. The state of the network can be conceptualized as a point in a state space with dimensionality equal to the number of criterion nodes (five in this simulation: nodes 14, 15, 21, 22, 23). A volume of this space is declared to be a "solution volume." If the state point of the network enters this volume, the network is considered to have reached a solution. The number of time steps until this occurs, or until the distance between the current state and the solution volume ceases to decrease, is the time needed to disambiguate the input. Syntactic disambiguation was operationally defined as follows: five criterion nodes crucial to the past-participle interpretation are activated to > 500 units (nodes 15, 22, 23), and nodes likely to have been strongly activated due to past-tense bias (nodes 21 & 14) have activation decreased to < 500 units.

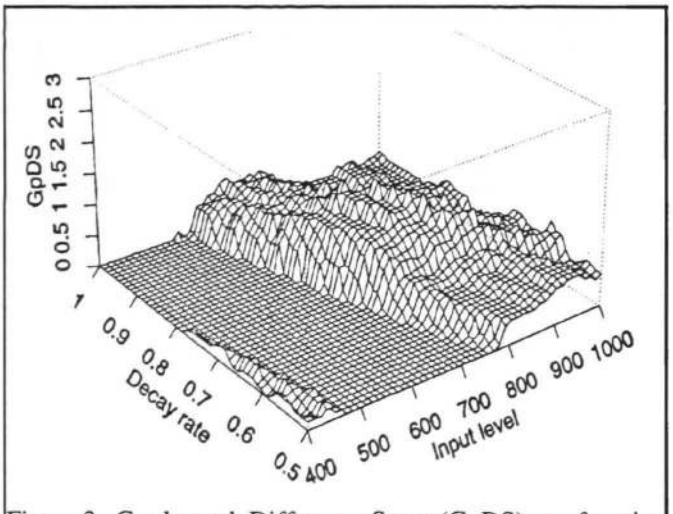

Figure 3. Gardenpath Difference Score (GpDS) as a function of decay rate and input activation level.

The specific decay rate (d = .8) and the specific input rate (A = 760) were determined by running a set of simulations and computing a Gardenpath Difference Score (GpDS) to plot out the range of input by decay parameters that produced the desired gardenpath pattern. The basic empirical result that the network needed to replicate, given the set of inputs from stimuli from both the Burgess and Tanenhaus (BT) (1992) and the Ferreira and Clifton (FC) (1986) studies, was a processing advantage for the BT past-participle sentences over the past-tense sentences, compared to any processing advantage the FC past-participle sentences had over their past-tense sentences. Figure 3 plots out the GpDS as a function of decay rate and input level. Each point on the figure represents the degree of difference in the two sets of stimuli's gardenpathing. The larger the value of the GpDS, the stronger the past-participle effect of the BT stimuli as compared to the FC stimuli. It can be seen that this effect is quite robust in that a large portion of the landscape conforms to this pattern given the input sets. Lower input levels were not effective, since an insufficient amount of activation would accrue. Conversely, smaller decay rates were not effective, because the network would become saturated with activation.

The analysis of the GpDS was used to set the input activation level and decay rate, thus, calibrating the system. It is important to note that once the network was calibrated, all subsequent constraints on network performance are solely determined by the normative aspects of the input data.

Morphologically Ambiguous Verbs

At the heart of the system is the morphologically ambiguous verb. The theoretical motivation for such a representation stems from the lexical ambiguity research. The activation of word meanings occur in parallel as a function of their respective associative strength (Burgess & Simpson, 1988; Simpson & Burgess, 1985). Similarly, the morphologically ambiguous verb would make available its past-tense (node 14) and past-participle (node 15) forms at different rates. The bias for each form would depend on the usage of that particular verb in natural language. As these morphological forms become active in parallel, so does each of their sets of associated thematic roles (Carlson & Tanenhaus, 1988).

Morphological frequencies of past-tense and past-participle verb forms were extracted from the raw data set used to compute the word frequency count of Francis and Kucera (1982). All sentences containing the verbs used by both sets of materials were extracted from the data set. The relative proportion of occurrence (past-tense or past-participle) was computed by calculating how often the verb was used in a simple past-tense construction and in a past-participle construction. These proportions were used as the weights for the two forms of the verb, labeled (L3) and (L4). Thus, if a verb occurred rather frequently as a past participle, that

morphological form would receive greater activation than a verb whose past-participle form was more rare.

Goodness of Fit Between Agent, Patient, and the Verb

We assume that the associative quality of the agent and patient to the ambiguous verb are relevant in determining the likelihood of the gardenpathing phenomenon. Consider the sentence fragment *The man paid....* Our intuition suggests a preference to continue this fragment with a simple past-tense completion, such as in (2a). The less frequent past-participle form (2b) seems to result in a temporary gardenpath, since there is an expectation that *The man* has paid someone, rather than having been paid by someone.

(2a.) The man paid his workers with cash.
(2b.) The man paid by the parents was unreasonable.
(2c.) The ransom paid by the parents was unreasonable.

However, intuition suggests that the past-participle version in (2c) does not produce the same gardenpath effect as (2b). That is, our knowledge of the world suggests that "The ransom" could not have "paid" anyone, and, thus, must have been what was paid.

Thematic Plausibility - Plausibility data was collected from 160 subjects for the stimuli from the Burgess and Tanenhaus (1992) and the Ferreira and Clifton (1986) experiments. Subjects were asked to rate plausibility by responding to questions like the following (using a one to ten scale):

(3a.) Can ransom pay anything (or someone)? [Agent Bias]
(3b.) Can ransom be paid by someone? [Patient Bias]

These ratings were converted to link weights that correspond to the connections L1 and L2 which connect NP1 with *agent* and *patient*.

Role of the Disambiguating Preposition

The results of the Burgess and Tanenhaus (1992) experiment, in which the verb and preposition were presented separately, suggest that thematic constraints alone are not sufficient to avoid a gardenpath. A comparison of the single-word and two-word reading experiments compels us to involve the grammatical cue of the disambiguating preposition. Of course, even the integration of the preposition while processing the verb was not, in itself, sufficient to avoid the gardenpath, given that the less constraining sentences still gardenpathed. The past-participle form of an ambiguous verb frequently occurs in conjunction with a disambiguating preposition. This suggests a specific association between the preposition and the agent thematic role which is activated from the past-participle morphological verb form.

However, the preposition is not absolutely disambiguating. Sixteen subjects were presented with sentence fragments (NP+Verb+prep) from both sets of stimuli and asked to complete the sentence. Fillers were included so the ambiguities would be less obvious. The following two sentences illustrate the variability in disambiguation with the preposition.

(4a.) The author read by the class was Pearl Buck [past participle]
(4b.) The author read by the fireplace. [past tense]

The preposition does not always indicate an agent in the prepositional phrase. In (4b), for example, it indicates a location. Other thematic roles are possible as well. The past-tense and past-participle completion proportions were used as inputs to PP1 (node 9) in the model in order to set weights on links L5, L6, L7, and L8. Links L5 and L7 serve to verify that the preposition serves to confirm a past-tense interpretation (node 14) and that a patient (or some non-agentive role) is carried by the prepositional phrase (node 22). Links L6 and L8 serve to verify that the preposition confirms a past-participle interpretation (node 15) and that the prepositional phrase contains the sentence's agent (node 23).

Simulation Experiments

The first simulation used parameters from two sets of stimuli: sentences from Burgess and Tanenhaus (1992) and Ferreira and Clifton (1986). For both Figures 4 and 5, the y-axis corresponds to either reading time or number of iterations for the network to settle. For Figures 4, the x-axis corresponds to past-tense or past-participle biased sentences. The independent variable plotted in the interior of the figure reflects the stimuli set used, either those of Burgess and Tanenhaus (BT) or of Ferreira and Clifton (FC).

Replication of Two-word Segmentation Reading Experiment

Using the two word segmentation reading task, Burgess and Tanenhaus (1992) replicated the result found by Ferreira and Clifton (1986), i.e., no reading time advantage for the past-participle biased sentences. However, with more constraining stimuli, the past-participle biased sentences were understood more quickly at the point of disambiguation (Verb+prep) than were the past-tense biased items. Figure 4a illustrates their results.

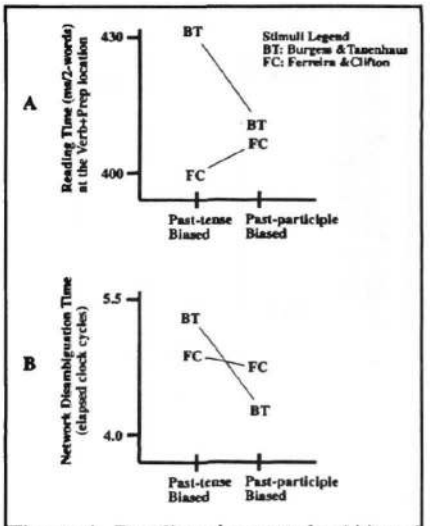

Figure 4. Reading time results (A) and simulation results (B) at the Verb+Prep sentence location.

The same set of effects can be seen in the simulation

outcome: an advantage in processing time for the past-participle biased sentences compared to the past-tense, but only for the more constraining sentences (see Figure 4b). The simulated overall reading time for Ferreira and Clifton's sentences is longer than their actual results.

The compelling aspect of these results is that three potential sources of information need to converge in the comprehension process to avoid a processing load associated with the past-participle biased sentences. The NP has to be sufficiently constraining in combination with the morphological availability of the verb. While this is a necessary constraint, it does not appear sufficient to avoid the processing load. This information has to be able to combine with the disambiguating effect of the preposition. Presumably there is a narrow window of opportunity in the timecourse of processing for this to occur given that past-participle forms are relatively infrequent. If true, one would expect a delay in the integration of the preposition to result in a gardenpath. This is implied in the difference in the one-word and two-word segmentation manipulation by Burgess and Tanenhaus; however, it is directly tested in the following experiment.

Manipulating Preposition Length

In the previous studies, the preposition length was short since *by* was used in all sentences by Burgess and Tanenhaus, and in half the sentences by Ferreira and Clifton. Using only the strongly past-participle biased sentences that did not show the gardenpath effect (from Burgess and Tanenhaus), Burgess, Tanenhaus, and Hoffman (1994) manipulated the length of the preposition so that when subjects read the sentence, presumably, a single eye fixation either would or would not capture both the verb and preposition. As discussed earlier, these sentences gardenpath in the long preposition condition, but do not show this effect in the short preposition condition (replicating their earlier result). This can be seen in the left side of Figure 5.

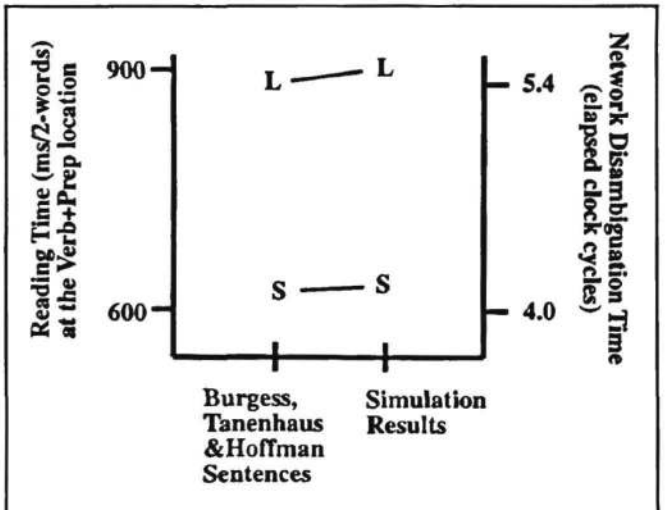

Figure 5. Reading time results and simulation results for long and short preposition conditions using past-participle biased sentences at the Verb+Prep sentence location.

Of interest here, is whether the simulation model we present will capture this effect. This was simulated by manipulating the temporal nature of the input of the verb and the preposition. Assuming that in the short preposition condition, an eye fixation is able to capture the verb and preposition together, we simulated that by passing activation to both those lexical units simultaneously. In the long preposition simulation, activation was passed to the preposition one time step later than when the verb was activated. These results are presented on the right side of Figure 5 and match the results of found by Burgess, Tanenhaus, and Hoffman (1994) in that the long preposition condition takes longer to settle on an interpretation than does the short preposition condition.

General Discussion

A variety of recent experiments have demonstrated that several factors can influence how the parser makes local online commitments during the comprehension process (Burgess and Tanenhaus, 1992; Burgess, Tanenhaus & Hoffman, 1994; MacDonald, 1994; Trueswell, Tanenhaus & Garnsey, 1993). We have demonstrated with this set of simulations that a plausible theoretical framework for a range of these results is a hierarchical connectionist network which is sensitive to a number of constraints inherent in the input stimuli.

The model accounts for the top-down effect of context, the contribution of the bottom-up morphological frequency asymmetry of the verb, and the probabilistic nature of the disambiguating preposition. These effects are sensitive to the timecourse of processing as well. The model makes some counterintuitive predictions. For example, it predicts that subjects would gardenpath on the minimal attachment structure when the verb bias is strongly past-participle. We suspect that the same model can account for the context by dominance interactions found in the lexical ambiguity literature as well.

While norming procedures have been used frequently in psycholinguistics experiments, they have been utilized infrequently in computational modeling (viz., Rumelhart, Smolensky, McClelland & Hinton, 1986). The current simulations validate the basic methodology of using normative aspects of stimuli to set certain input parameters in connectionist models of natural language processing. These normative values are another form of representing regularities in language use.

The pattern of results from the psycholinguistic data suggest that syntactic processing is a confluence of multiple constraints that represent both bottom-up and top-down influences in processing. These results are incompatible with a deterministic parsing model such as Minimal Attachment. The hierarchical connectionist style model presented in this paper is sensitive to the range of constraints discussed above and is offered as a more adaptive theoretical model that can capture the domain of effects found in the literature encompassing local syntactic ambiguity resolution.

Acknowledgements

We want to thank Jennifer Melbinger for her assistance in coding the norming study and Michael Tanenhaus, Catherine

Decker and three anonymous reviewers for their helpful comments. This study was supported in part by a University of California Academic Senate Research Grant awarded to the first author.

References

Burgess, C., & Hollbach, S.C. (1988). A computational model of syntactic ambiguity as a lexical process. In *Proceedings of the Tenth Annual Cognitive Science Society Meeting* (pp. 263-269). Hillsdale, NJ: Lawrence Erlbaum Associates.

Burgess, C., & Simpson, G. B. (1988). Cerebral hemispheric mechanisms in the retrieval of ambiguous word meanings. *Brain and Language, 33,* 86-103.

Burgess, C., & Tanenhaus, M. K. (1992). *Semantic, syntactic and visual factors in syntactic ambiguity resolution.* Unpublished manuscript.

Burgess, C., Tanenhaus, M. K., & Hoffman, M. (1994). Parafoveal and semantic effects on syntactic ambiguity resolution. In *Proceedings of the Sixteenth Annual Cognitive Science Society Meeting.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Cottrell, G. (1988). A model of lexical access of ambiguous words. In Small, S., Cottrell, G., & Tanenhaus, M. (Eds.), *Lexical Ambiguity Resolution* (pp. 195-228). San Mateo, CA: Morgan Kaufmann Publishers.

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

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies.* Ph.D. Thesis, University of Connecticut. Indiana University Linguistics Club.

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

Marslen-Wilson, W. (1975). Sentence perception as an interactive parallel process. *Science, 189,* 226-228.

McClelland, J.L. (1987). The case for interactionism in language processing In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 3-36). Hillsdale, NJ: Erlbaum Associates.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In J. McClelland & D. Rumelhart (Eds.), *Parallel distributed processing, Part 2: Psychological and biological models* (pp. 7-57). Cambridge, MA: MIT Press.

Tanenhaus, M. K., & Carlson, G. N. (1988). In W. D. Marslen-Wilson (Ed.), *Lexical representation and processing.* Cambridge, Ma: MIT Press.

Tanenhaus, M.K., Burgess, C., Hudson-D'Zmura, S, & Carlson, G. (1987). Thematic roles in language processing. In *Proceedings of the Ninth Annual Cognitive Science Society Meetings* (587-596). Hillsdale, NJ: Lawrence Erlbaum Associates.

Trueswell, J.C., Tanenhaus, M.K., & Garnsey, S.M. (1993). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language.*