Journal of Psycholinguistic Research, Vol. 23, No. 6, 1994

A Constraint-Based Lexicalist Account of the Subject/Object Attachment Preference

Cornell Juliano[1,2] and Michael K. Tanenhaus[1]

When a noun phrase could either be the object of the preceding verb or the subject of a new clause or a sentence complement, readers and listeners show a strong preference to parse the noun phrase as the object of the verb. This can result in clear garden paths for sentences such as The student read the book was stolen *and* While the student read the book was stolen. *Even when the verb does not permit a noun phrase complement, some processing difficulty is still found. This has led some theorists to propose models in which initial attachments are lexically blind, with lexical information subsequently used as a filter to evaluate and revise initial analyses. In contrast, we show that these results emerge naturally from constraint-based lexicalist models. We present a modeling experiment with a simple recurrent network that was trained to predict upcoming complements for a sample of verbs taken from the Penn Treebank corpus. The model exhibits an object bias and it also shows effects of verb frequency which are similar to those found in the psycholinguistic literature.*

This article is a preliminary report on a modeling experiment we are conducting to evaluate the plausibility of a *constraint-based lexicalist* account of attachment preferences involving subject/object ambiguities. Consider the examples in 1a and 1b:

1a. The student read the book was stolen.
1b. While the student read the book was stolen.

We wish to thank Michael Spivey-Knowlton for helpful comments. This work was supported by NIH grant HD-27206.

[1] University of Rochester, Rochester, New York.
[2] Address all correspondence to Cornell Juliano, Department of Psychology, Meliora Hall, University of Rochester, Rochester, New York 14627-0038.

In both of these sentences, the noun phrase *the book* is initially taken to be the object of the verb *read*, resulting in processing complexity, and for many people, a conscious feeling of being garden-pathed when *was* is encountered (e.g., Frazier & Rayner, 1982). The garden-path effect is clearly much stronger in sentence 1b than it is in sentence 1a.

A surprising result is that readers experience processing difficulty even when the verb does not permit an NP complement, as is illustrated in example 2a:

2a. The student implied the book was stolen.
2b. The student implied that the book was stolen.
2c. The student read the book was stolen.

Under these conditions processing difficulty occurs at the noun phrase (Garnsey, Lotocky, & McKonkie, 1992; Juliano & Tanenhaus, 1993; Trueswell, Tanenhaus, & Kello, 1993), though it may spill over onto the verb in the complement (Ferreira & Henderson, 1990). Reading times at the noun phrase *the book* would be longer compared to either a control sentence like 2b in which there is a complementizer or a sentence like 2c in which the verb is typically used with an object.

The sentences in 3 illustrate the effect demonstrated by Mitchell (1987, 1989; also Adams, Mitchell & Clifton, 1993):

3a. While the student hesitated the book was stolen.
3b. While the student hesitated, the book was stolen.
3c. While the student bought the book her parents waited in the car.

In the absence of a comma, readers have difficulty processing the subject of a main clause when it immediately follows the verb in the subordinate clause, *even* when the verb does not subcategorize for an NP complement. The comparisons which reveal this difficulty are between sentences like 3a and either 3b, which has a comma after the verb, or 3c in which the noun phrase is the object of the verb. The only possible situation where difficulty does not obtain in the commaless sentences is when the main clause begins with a case marked pronoun (Trueswell, personal communication).

Both the results for sentence complements and subordinate/main clause constructions might appear to present serious problems for any parsing model that makes immediate use of lexical information (e.g., Abney, 1989; Pritchett, 1993; Gibson, 1991). Rather, they seem to provide striking support for two-stage parsing models in which initial attachments are made on the basis of category information (Clifton, Speer, & Abney, 1991; Frazier, 1987; 1989; Mitchell, 1987;1989). Argument structure information then comes into play later as a *lexical filter*. In these models, the difficulty that readers experience at the noun phrase is attributed to a garden path which occurs because the initial structure built by the first-stage parser is being revised using lexical information.

The claim that we advance here is that these data are better explained within the emerging *constraint-based lexicalist* framework (cf. MacDonald, Pearlmutter, & Seidenberg, in press; Pearlmutter, Daugherty, MacDonald, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994; Trueswell, Tanenhaus & Garnsey, 1994). In this framework, the system makes full use of lexical information within a constraint-based architecture. Parsing preferences such as *object preference* emerge as a result of contingent frequencies (see also Mitchell & Cuetos, 1991) rather than as the result of parsing strategies that reflect the architecture of a parser.

In developing our account, we will focus on sentence complements. We first review some recent data showing that the difficulty that readers have with an NP after a sentence complement verb is strongly correlated with the frequency of the verb. We then develop a lexicalist model trained from a corpus that shows this effect as well as the standard object bias. Finally, we briefly discuss why the model should generalize to the subordinate/main construction.

THE DATA

Several recent studies have found that the magnitude of the difficulty that readers have with noun phrases after strongly biased sentence complement verbs (S-bias verbs) varies as a function of the properties of individual verbs. Trueswell et al. (1993) found that reading times to the noun phrase in a *that*-less complement after a strong S-bias verb were longer than comparable times to the subject of the complement when it contained the complementizer *that* or to reading times to the same noun phrase when it followed an NP-biased verb. However, for the S-bias verbs, there was a strong correlation between the magnitude of the complementizer effect at the noun phrase and how often the verb occurred with a complementizer, as measured by completion norms. For verbs like *implied*, which typically occur with a *that* complementizer, there was a clear reading time penalty in sentences like 2a compared to sentences like 2b in which the complementizer *that* was present. However, for those verbs that typically occur without a *that* there was little or no penalty. Importantly, the regression equation showed no complexity effect for the verbs that are most likely to occur without a complementizer. This finding is crucial. Lexical filtering models

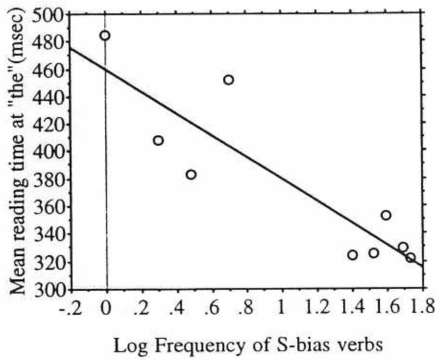

Fig. 1. Correlation between verb frequency and reading times to the word *the* following an S-bias verb.

might attribute frequency effects to the speed with which lexically driven revision takes place. However, these models *must* posit revision costs for even the highest frequency verbs.

Juliano and Tanenhaus (1993) replicated these results and further showed a strong inverse correlation between processing difficulty at the noun phrase after a strong S-bias verb and the verb's frequency of occurrence in the language. This is consistent with the Trueswell et al. (1993) results because frequency and *that*-preference themselves have a strong inverse correlation. Figure 1 presents the correlation between reading times to the word *the* in a self-paced reading experiment and the log frequency of the preceding S-bias verb taken from the data presented in Juliano and Tanenhaus[3] ($r^2 = .78406$, $F = 25.42$, $p < .01$).

This correlation is strikingly similar to an effect that is well documented in word recognition. Words with exceptional pronunciations (e.g., *caste*) take longer to pronounce than comparable words with regular pronunciations (e.g., *waste*). However the magnitude of this exception effect is correlated with the frequency of the exception words. High-frequency exception words such as *have* take no longer to pronounce than regular words of comparable frequency such as *gave*. This results in the well-known frequency by regularity interaction, as is illustrated in Fig. 2, adapted from Seidenberg and McClelland (1989).

In Fig. 3 we have grouped the highest- and lowest-frequency S-bias verbs together and contrasted them with noun phrase complement bias verbs of similar frequency (NP-bias verbs). The figure shows reading times at the determiner following the verb for high- and low-frequency verbs of each

[3] In this self-paced reading experiment, effects began at the determiner. In some other experiments, effects are delayed by a word. We find similar correlations when reading times to the entire noun phrase are used in the regression.

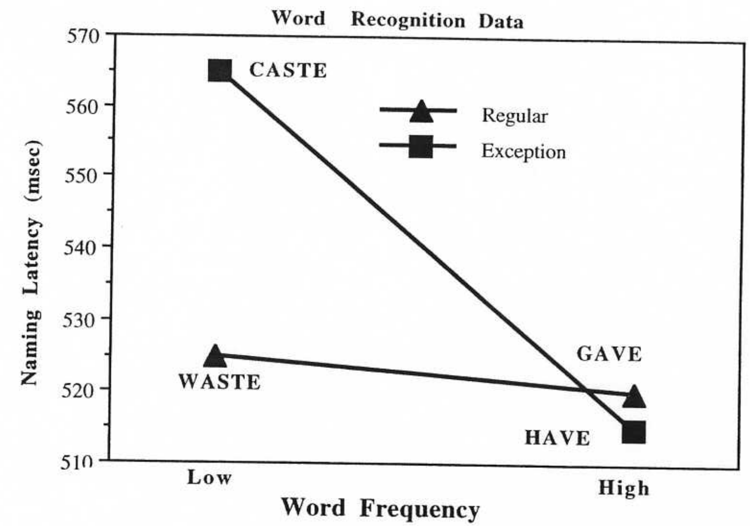

Fig. 2. Frequency by regularity Interaction in word naming. (Adapted from Seidenberg & McClelland, 1989.)

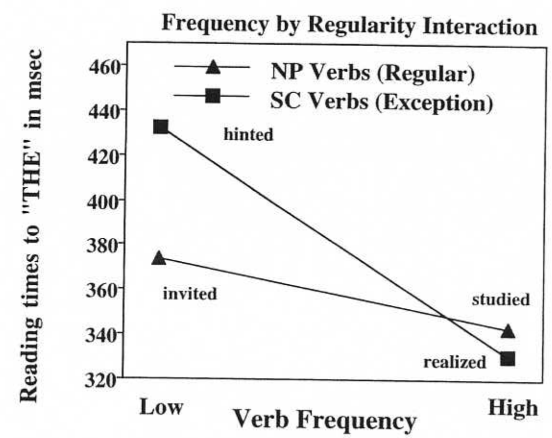

Fig. 3. Reading times to the determiner the for high and low frequency S-bias and NP-bias verbs.

type. There is a large difference in reading times between the low-frequency S-bias and NP-bias verbs, but not for the high-frequency verbs. Since a noun phrase that follows a verb is typically its object we will refer to this as the *regular* pattern. Thus, the reading time data at the determiner show the same interaction as is seen in the word recognition literature.

It is now well documented that this type of interaction will emerge from learning models with distributed representations in which the model learns the mapping between the spellings of individual words and their pronunciations. Thus the regular spelling pattern is not separately represented as a rule or strategy in the network. However, it is implicitly represented as a *bias* in the network, because the regular patterns occur more frequently. For frequent words, the bias has negligible effects on the computations involved in processing that word. However, the computations for lower-frequency words that run counter to the common pattern are delayed and the computations of lower-frequency words that correspond to the pattern are facilitated. Thus, the regularity effect need not be separately represented as a rule, a heuristic, or a principle that the word recognition system follows. Rather, it is a frequency or evidence-based bias in the system that emerges from the characteristics of the instances to which the system has been exposed.

THE MODEL

In order to see whether a similar explanation is viable for the NP-bias, we conducted a simple modeling experiment. Our goal was to test the claim that a constraint-based system that ony learns about verb/complement cooccurrences and subsequent evidence for complement types will exhibit the effects illustrated in Fig. 1, a correlation between verb frequency and complement preference, and Fig. 3, a frequency by regularity interaction for high- and low-frequency verbs.

The model learned to predict the complement type that follows a past tense verb with representative input taken from the corpora included in the Penn Treebank project (Marcus, Santorini, & Marcinkiewicz, 1993). The model uses a simple recurrent network (SRN) architecture consisting of an input layer, a hidden layer, and an output layer (Elman, 1990). The input consists of a total of 214 units. The first 156 units each represent a particular verb that the model recognizes. Thus we used a completely localist coding for each verb. The next 50 units each represent an item that follows a verb. This item could be one of 33 words (frequent articles, pronouns, and prepositions), punctuation, 10 category types or *other*, the case when the item that follows a verb does not belong to any of the categories just described. The last eight units of the input store the most recent activation of the hidden units. Thus they provide internal feedback about the past state of the model and are a way of implementing the equivalent of recurrent connections among the hidden units. The hidden and output layers both contain eight units. Each of the output units represents a particular complement type. The architecture of the model is shown in Fig. 4.

A training instance was made up of a past tense verb, the word that followed it and the type of complement that followed it. These word pairs and complement types were extracted from the sentences in which they occurred in the corpus. An example is presented in 4:

4a. (S (NP (DT The) (NN lawyer)) (VP (VBD insisted) (SBAR (IN that) (S (NP (PRP$ his) (NN client)) (VP (VBD was) (NP (NN innocent)))))
4b. insisted that SBAR

A sentence from the corpus would look like the one shown in example 4a and what was used to train the model is shown in example 4b. For this example, the word *insisted* and then the word *that* were presented as the input. The model was trained to predict a sentential complement (SBAR) because that was the complement type that occurred with this fragment in the corpus.

The training set was taken from the combined Brown Corpus and the *Wall Street Journal* Corpus, both tagged and parsed by the Treebank Project (Marcus et al., 1993). The training sequence was carried out in two steps. First, a verb unit that corresponds to the verb in the training instance was activated and the model was trained to produce the desired phrase type. Training used the backpropagation algorithm (Rumelhart, Hinton, & Williams, 1986). Next, the verb input was set to zero and the unit corresponding to the item that followed the verb was set. This was done to ensure that the only effects of the verb could come from the verb-complement mappings that the model had established. The model was again trained to produce the desired output. Note that there is a sequential component to the model's architecture. After the first presentation, a copy of the hidden unit activation

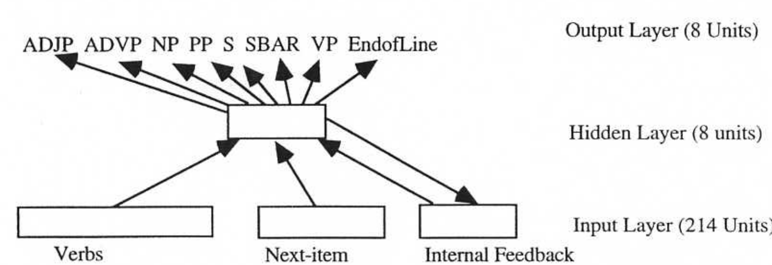

Fig. 4. The simple recurrent network architecture. Each rectangle represents a set of units. Arrows indicate the direction of flow of activation.

is fed back to the input. Thus the model has a memory for past responses that can influence succeeding events.

The training set consisted of a subset of all the past tense verb instances found in the corpus. The set of verbs that the model recognizes consist of 76 verbs that were of interest to us (i.e., verbs in which we had existing data from either norms or on-line experiments) along with an additional 76 verbs that matched the critical items in frequency and which were used to balance the critical verbs. Overall the training set consisted of 13,051 instances. The model saw each instance three times. Our goal in selecting the training set was to create as representative a sample of the entire corpus as possible. Table I presents a comparison of the training set and the entire corpus on some critical dimensions. In general, the match is quite good, though sentence complements are somewhat overrepresented because we included most of the limited number of verbs that permit sentence complements.

RESULTS

We tested the model by presenting it with a test set consisting of the verb alone. The model clearly developed appropriate verb-specific complement preferences. Because we can look at the error score for a complete set of complement types on any trial, Table II presents the summed square of the error for the NP and SBAR complement outputs for the verbs used by Trueswell et al. (1993) and Juliano and Tanenhaus (1993). These authors had selected their verbs based on completion norms. The error score reflects the confidence that the model has in a complement type. The lower the error

Table I. Comparison of Complement Types Found in the Treebank Corpus and in the Model's Training Set

	Complement statistics			
Source and phrase type	Past tense Verb	the (determiner)	that (determiner)	that (complementizer)
Treebank				
Verb NP	31,935	5632	100	0
Verb SBAR	8,502	1240	148	2230
Other	89,625	0	0	0
Training set				
Verb NP	5,686	1172	25	0
Verb SBAR	1,997	108	65	1029
Other	5,368	0	0	0

Table II. Error Scores for NP and SBAR Predictions for Different Verb Types

	Error scores	
Verb type	NP prediction	SBAR prediction
NP-only verbs	.038	1.646
NP-bias verbs	.505	1.297
S-bias verbs	.750	.474

score, the more confident the model is that it is processing a particular complement. Table II shows that the error scores clearly reflect a gradient of verb bias.

Note that the error score for the NP complement of the NP-biased verbs taken from Trueswell et al. (1993) is actually quite high. This is because some of these verbs were used frequently with sentence complements, primarily in the *Wall Street Journal* corpus. We suspect that this corpus is not representative in a number of crucial respects and that completion norms actually provide a better estimate of the true frequency of the complement occurrences for these verbs in the types of sentence frames used in our experiments. In work in progress, we are comparing the performance of models trained from completions and models trained from corpora to reading time data.

Second, we tested the hypothesis that there is an NP complement bias in the system. We did this by presenting no input and looking at the error scores of the output. These scores reflect the residual bias in the network. The bias was clearly toward an NP complement. The error score for the NP complement was .375, whereas the next lowest score was .929 for the SBAR complement, followed by the .993 for the PP complement type.

Next, we focused on a test of the verb followed by *the*. We compared the behavior of the model on just those S-bias verbs that were used in the Juliano and Tanenhaus (1993) experiments. We conducted a regression in which the log frequency of the verbs as found in the training set was used to predict the sum of the squared error for the S-complement output. The lower the error score, the greater the confidence the model has in a prediction. There is a strong correlation between the error score and verb frequency ($r^2 = .69592$, $F = 15.95$, $p < .01$). Figure 5 presents a scatter plot that includes the experimental verbs plus all the verbs in the corpus that had the same range of NP and SBAR complement frequency as the experimental verbs. Again, there is a strong correlation between verb frequency and error scores ($r^2 = .61977$, $F = 34.23$, $p < .01$). In Fig. 6, we reconstruct the frequency by regularity interaction by taking the error score for the SBAR complement for the S-bias verb with the highest frequency (*thought*) and