Modeling the Role of Plausibility and Verb-bias in the Direct Object/Sentence Complement Ambiguity

Todd R. Ferretti (todd@sunrae.sscl.uwo.ca)
Ken McRae (mcrae@uwo.ca)

Department of Psychology, Social Science Centre, University of Western Ontario
London, Ontario, Canada N6A 5C2

Abstract

We provide a computational account of the integration of various constraints proposed to be involved in the resolution of the direct object/sentence complement ambiguity. In the first part, competition-integration simulations show that a constraint-based model accounts for the results of Garnsey, Pearlmutter, Myers, and Lotocky (1997) at least as well as the garden-path model. In the second part, we compare the efficacy of norming techniques for capturing plausibility effects. Simulations show that norms designed to tap people's conceptual knowledge of events better capture plausibility effects than do norms that are biased toward tapping linguistic knowledge. We conclude that local information concerning event plausibility is an important constraint for understanding ambiguity resolution.

Part 1: Distinguishing Between Theories of Sentence Processing

The human language comprehension system is impressive in its ability to integrate multiple sources of information when comprehending sentences (Marslen-Wilson, 1975). Most theories of sentence comprehension include the idea that understanding sentences involves the rapid integration of general syntactic information (Frazier & Rayner, 1982), lexically-specific syntactic information (MacDonald, Pearlmutter, & Seidenberg, 1994; Garnsey et al., 1997), discourse information (Spivey & Tanenhaus, 1998), and knowledge of thematic roles (McRae, Ferretti, & Amyote, 1997; Trueswell, Tanenhaus, & Garnsey, 1994). They differ, however, in their claims about precisely when the comprehension system exploits various types of information. Two prominent theories of sentence processing have emerged in this debate.

The garden-path model (Frazier, 1987; Frazier & Rayner, 1982) claims the comprehension system naively constrains initial interpretation by using a limited subset of the relevant information, reserving other information sources for evaluating and, if necessary, revising the initial interpretation. The first stage of comprehension uses only the major syntactic category of each word (noun, verb, etc.), phrase-structure rules, and a small set of syntactic decision principles. One of these principles, minimal attachment, states that the initial interpretation *always* corresponds to the simplest structure that can be built. The second stage of comprehension temporally lags behind the first, with all potentially relevant sources of knowledge being used to evaluate, and if necessary, revise the initial structure.

In contrast, the constraint-based approach views syntactic ambiguity resolution as a continuous process in which the most likely syntactic alternatives are evaluated with respect to all available evidence (MacDonald et al., 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998). Thus, the distinguishing features of this model are that multiple constraints are combined to compute alternative interpretations in parallel, and that these alternatives compete with one another during processing. Unlike the garden-path model, lexically-specific knowledge and relevant discourse information is available to guide initial interpretations rather than solely to revise the initial interpretation if it is incorrect.

The Direct Object/Sentence Complement Ambiguity

The direct object/sentence complement ambiguity occurs when a noun phrase (NP) following a main verb is temporarily ambiguous with respect to its relationship to the verb. Consider the following examples:

1a) The gossipy neighbor heard (that) the story had never actually been true.
1b) The gossipy neighbor heard (that) the house would never be flooded again.

In (1), the NPs "the story" and "the house" are ambiguous as to whether they are the direct object of the verb (i.e., something that someone has heard), or the subject of a sentence complement (something that someone has heard *about*). In (1a), for example, when the complementizer "that" is removed from these sentences, readers receive confirmation that the gossipy neighbor has not directly heard the story when they read the auxiliary verb "had" or "would". Readers' use of verb-bias and plausibility information to constrain resolution of the direct object/sentence complement ambiguity has recently been investigated by Garnsey et al. (1997). They manipulated the plausibility of the post-verbal noun as a direct object. For example, in (1a) "the story" is a plausible direct object of the verb "heard" (i.e., people tend to hear stories directly), whereas "the house" is not as plausible (people do not commonly hear houses directly). Of particular interest to Garnsey et al. was how this type of plausibility

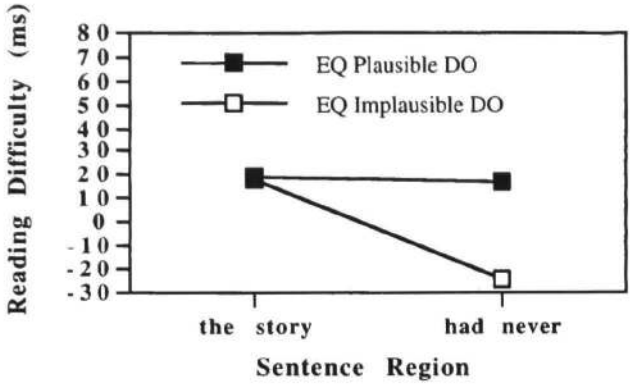

Figure 1: Reading difficulty (ambiguous unambiguous reading times) for equi-bias verbs in the two critical regions of the sentences for both implausible and plausible direct objects.

information influences ambiguity resolution when the main verb of the sentence frequently occurs with either direct objects (e.g., heard), sentence complements (indicated), or both structures with about equal probability (felt). The main prediction of this study was that direct object plausibility would have its strongest influence when verbs appeared with the alternative syntactic structures with about equal probability. This prediction was supported by their first-pass eyetracking results, illustrated in Figure 1. When a plausible NP ("the story") followed an equi-bias verb, readers had difficulty at the disambiguating region ("had never"), but not when the NP was implausible ("the house"). This pattern of results appears to be most compatible with the constraint-based model. However, because the possibility remains that revision processes may have accounted for the data, the results may be captured by the garden-path theory as well.

Competition-Integration Modeling

To distinguish between these alternatives, two competition-integration models were constructed to implement the predictions of the opposing theories. Figure 2a presents a schematic of the garden-path version of the model, whereas Figure 2b is the constraint-based version. In these models, the syntactic alternatives compete with one another during processing until one alternative reaches a criterion value. Competition arises because multiple constraints provide support for various interpretations. A major assumption of this model is that the amount of competition produced in a region of a sentence is proportional to the reading difficulty people experience at the region, with reading difficulty measured as the reading-time difference between ambiguous (no "that") and unambiguous ("that" present) versions of the sentence.

The Relevant Constraints

The two versions of the model differ most significantly in that the garden-path version includes an additional constraint, labeled general verb bias, that represents the minimal

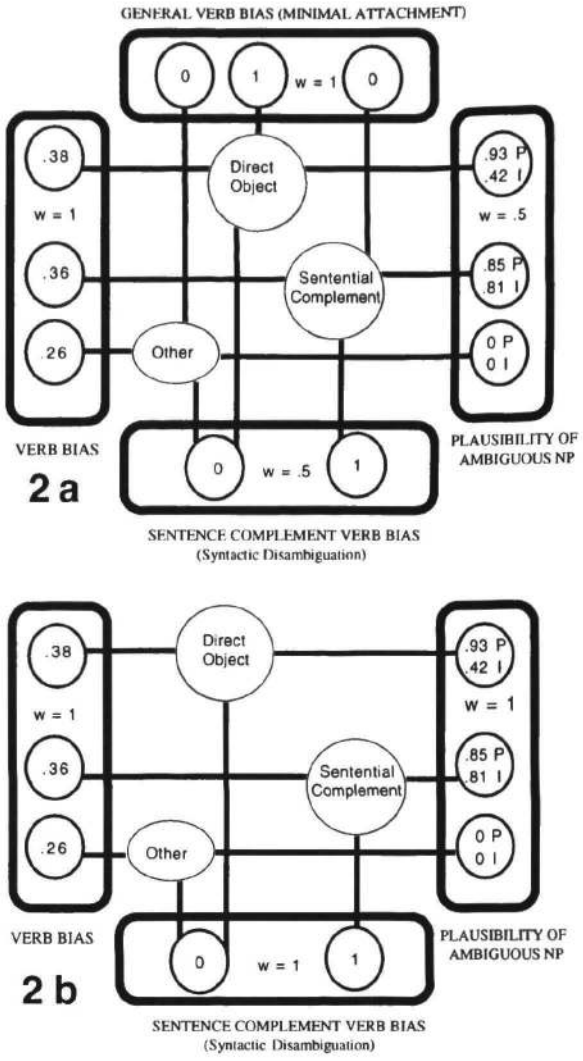

Figure 2: 2a is a schematic of the garden-path version of the model, whereas 2b is the constraint-based version.

attachment parsing principle. This constraint provides full support for the direct object interpretation in the main verb region because this is the simplest syntactic structure that can be constructed from the input to that point. In addition, there are three constraints common to both versions: verb-bias, the plausibility of the postverbal NP, and the sentence complement bias provided by the auxiliary verb. The verb-bias constraint captures the lexically-specific bias of the main verb in terms of the frequency with which it is followed by a direct object, a sentence complement, or some "other" structure such as a prepositional phrase or infinitive. The values for this constraint were taken from sentence completions collected by Garnsey et al. and the means are listed in the small circles inside the rectangle representing verb-bias.). The "plausibility of ambiguous NP" constraint captures plausibility information provided by the noun appearing in the ambiguous NP region. Note that the "P" and "I" values in the small circles represent the mean plausibility of the nouns (divided by the highest rating possible) labeled as plausible and implausible by Garnsey et al. These values

were collected in a norming study which is described in more detail in Part 2. Finally, the "sentence complement verb bias" constraint fully supports the sentence complement structure because it reflects the fact that the auxiliary verb disambiguates the sentence toward the complement interpretation.

There were no free parameters in the constraint-based version of the model. Each constraint was given a weight of 1 because the verb-bias entered into competition at the initial verb region, plausibility entered at the NP region ("the story") and the sentence complement verb-bias entered at the "had never" region. In the garden-path version, both the lexically-specific verb-bias and plausibility were delayed by one word. Thus, minimal attachment entered competition alone at the initial verbs and the verb-bias entered alone at the NP region, so both were given a weight of 1. Because the remaining two constraints entered at the "had never" region, the most straightforward way to weight them was .5 each. Note however that similar results were obtained using weights in the range of .2/.8 to .8/.2.

Processing

As in McRae et al. (1998), constraints were integrated at each region using a three-step normalized recurrence mechanism. First, each of the c informational constraints was normalized.

$$S_{c,a}(\text{norm}) = S_{c,a} / \Sigma S_{c,a} \quad (1)$$

$S_{c,a}$ represents the activation of each constraint node (i.e., the c^{th} constraint that is connected to the a^{th} interpretation node). $S_{c,a}(\text{norm})$ is the normalized constraint activation. Constraints were then integrated at each interpretation node.

$$I_a = \Sigma [w_c * S_{c,a}(\text{norm})] \quad (2)$$

I_a represents the activation of the a^{th} interpretation node. The weight linking the c^{th} constraint node to interpretation node I_a is represented by w_c. Finally, the interpretation nodes sent positive feedback to the constraints based on Equation 3. Note that the weights were equal in both directions.

$$S_{c,a} = S_{c,a}(\text{norm}) + I_a * w_c * S_{c,a}(\text{norm}) \quad (3)$$

These three steps were computed in sequence for each cycle of competition. The network iterates at each region, with the difference between the interpretation nodes gradually increasing until a criterion is reached. This criterion changes over the iterations within a region so that it becomes more lax as competition ensues. A dynamic criterion is necessary for simulating reading across multiple regions of a sentence due to the fact that fixation durations are partially determined by a preset timing program (Rayner & Pollatsek, 1989); that is, a reader will spend only so long on a fixation before making a saccade. As in Spivey and Tanenhaus' (1998) eyetracking simulations, we used $\Delta crit = .01$.

$$\textit{dynamic criterion} = 1 - \Delta crit * \textit{cycle} \quad (4)$$

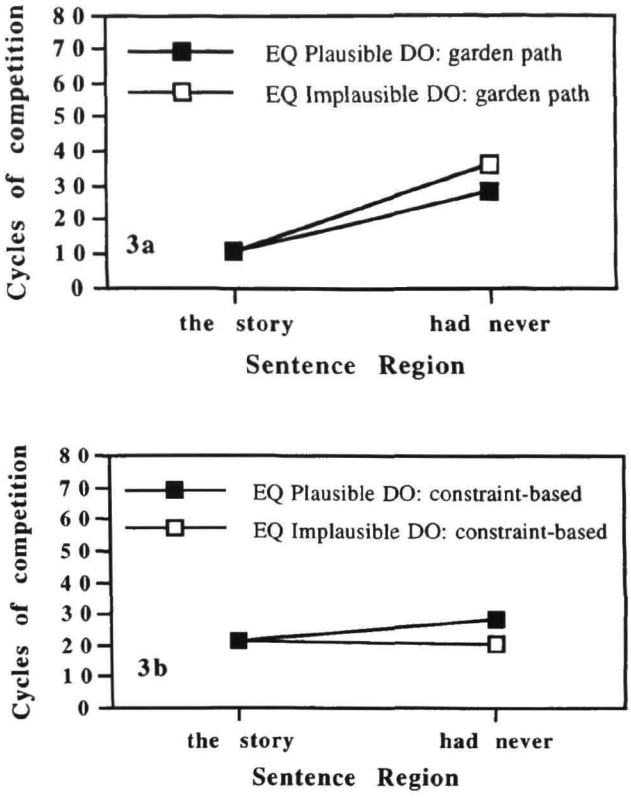

Figure 3: Cycles of competition to reach the dynamic criterion for plausible and implausible nouns in the two critical regions for the sentences containing equi-bias verbs.

Simulation Results

As illustrated in Figure 3a, the garden-path simulation predicts a different pattern of ambiguity effects than those found in the human reading-time data shown in Figure 1. The main difference between the simulations can be seen in the pattern of competition at the auxiliary verb region. The major discrepancy between the human data and the garden-path simulation is that the implausible nouns produced more competition than the plausible nouns in this region. This pattern of results is opposite to that found in the Garnsey et al.'s eyetracking data. Basically, the delay of the verb-bias and plausibility information could not be overcome. In contrast, the constraint-based simulation produced less competition for implausible nouns than for plausible nouns in the auxiliary verb region, as was found in the reading-time experiment. These simulations add to those of McRae et al. (1998) and Spivey and Tanenhaus (1998) in suggesting that architecturally-determined delays of lexically-specific syntactic and plausibility information are not present in the human sentence-comprehension system.

A look at the results of the constraint-based simulation, however, shows that although it captured the human data better than did the garden-path model, it does not completely capture the pattern of results. In particular, the pattern of reading difficulty at the verb versus at the auxiliary region was not mimicked. This is addressed in Part 2.

Table 1: Mean (and standard error) of the ratings for the items in the Equi-bias conditions of Garnsey et al. for each norming technique. Values represent the actual rating divided by the highest rating possible for each technique. "Plausible DO" and "Implausible DO" signify nouns that were classified as plausible or implausible direct objects by Garnsey et al.

	Garnsey et al. (1997)		*Role/Filler (no subject)*		*Role/Filler (with subject)*	
	Plausible DO	Implausible DO	Plausible DO	Implausible DO	Plausible DO	Implausible DO
Rating As Direct Object	.93 (.01)	.42 (.03)	.75 (.04)	.33 (.01)	.74 (.04)	.34 (.03)
Rating As Head of Sentence Complement	.85 (.02)	.81 (.02)	.74 (.03)	.42 (.03)	.83 (.04)	.52 (.05)

Part 2: Estimating Plausibility Constraints

One possible reason that the constraint-based simulation did not simulate the human data more closely is that the measure used to index plausibility did not properly capture it. Garnsey et al. indexed plausibility by asking participants to rate sentences or sentence fragments. To estimate the plausibility of the postverbal nouns as direct objects, participants rated sentences similar to (2a) on a 7-point scale (7=very plausible). To estimate the plausibility of postverbal nouns as the subject of a sentence complement, participants rated sentence fragments such as (2b).

(2a) The account executive concluded the speech.
(2b) The account executive concluded that the bank had

One potential problem with this type of norming is that rating sentences and sentence fragments for overall plausibility may not be the best indicator of local plausibility contingencies between lexical concepts in sentences. An alternative method for capturing the plausibility that specific objects or entities play various roles in specific events has been developed by McRae and colleagues (McRae et al, 1997; McRae et al., 1998). The underlying assumption of this norming method is that noun and verb combinations denote situations that happen in the world. Some of these situations tend to be common and others less common. Role/filler typicality norms index the plausibility of nouns as fillers for thematic roles of verbs by measuring how commonly the denoted situation occurs in the world. These norms differ from sentence ratings in that they abstract away from the linguistic structures and focus directly on knowledge about events or situations. McRae et al. (1998) showed that this world knowledge is computed and used immediately.

The following experiments and simulation examined the possibility that using a situation-based measure of plausibility might shed some light on the pattern of plausibility effects found by Garnsey et al.

Method

Participants

Sixty-eight native English-speaking psychology undergraduates from the University of Western Ontario received course credit for their participation.

Materials and Procedure

The 16 equi-bias verbs and their corresponding plausible and implausible nouns were rated in two norming studies. The first indexed the plausibility of each noun as either the direct object or subject of a sentence complement with the initial noun phrase left unspecified, thus indexing more locally the plausibility of the verb-noun combination. The second norming study did the same, but the initial noun phrase was included. This study tapped how plausibility changes as a result of certain entities being involved in the situations denoted by the verbs (i.e., noun-verb-noun combinations).

Ratings for Verb-Noun Combinations

To estimate plausibility of the post-verbal noun as a direct object, participants were asked questions such as:

How common is it for a
story _____
house _____
to be heard by someone?

Note that in the study the target nouns were placed among several filler nouns that varied in plausibility.

Similarly, to estimate plausibility of the post-verbal nouns as the subject of a sentence complement, participants rated the same nouns following questions such as:

How common is it for someone to hear something about a ?

Participants rated each combination on a 9-point scale, where 1 corresponded to very uncommon and 9 to very common. For each question type, the 16 items were placed in two separate booklets, with the second booklet containing the same question format as the first but with the items in

reversed order. Participants were not under time pressure but were asked not to spend too much time on any one item, and to write down the number that corresponded to their first impression. Participants rated nouns as a direct object or as a head of a sentence complement, but not both. The booklets took approximately 30 minutes to complete.

Ratings for Noun-Verb-Noun combinations

To index plausibility when the initial noun phrases were included, participants were asked questions such as:

How common is it for a

story _____

house _____

to be heard by a gossipy neighbor?

or

How common is it for a gossipy neighbor to hear something about a ?

All other aspects were identical to the experiment performed without the initial NPs.

Results and Discussion

The results for both direct object and sentence complement plausibility are shown in Table 1. Separate analyses were performed on the direct object and sentence complement ratings. Each set of ratings were entered into an analysis of variance using items as the random variable. The dependent variable was the mean rating (by item) divided by the maximum rating, and the independent variables were norming technique (3 levels: Garnsey et al. sentence norms vs role/filler without subject NP vs role/filler with subject NP) and plausibility (2 levels: plausible DO vs implausible DO). Norming technique and plausibility were within-items.

Ratings as Direct Objects

The interaction between plausibility and norming technique did not reach significance, $F(2,45) = 2.37$, $p<.11$. However, planned comparisons revealed that the nouns labeled as plausible were rated as more plausible in the Garnsey et al. norms than in either the role/filler norms without the subject included, $F(1,85) = 10.42$, $p<.01$, or the role/filler norms with the subject included, $F(1, 85) = 10.91$, $p<.01$. There was no difference between the latter two, $F<1$. The nouns labeled as implausible were rated numerically but not significantly more plausible in the Garnsey et al. norms than in the role/filler norms without the subject, $F(1,85) = 1.45$, $p>.1$, and with the subject, $F(1,85) = 1.83$, $p>.1$. Finally, the two role/filler norms did not differ, $F<1$.

There was a main effect of norming technique, $F(2,45) = 14.08$, $p<.001$. There was also a main effect of plausibility such that nouns classified as plausible direct objects were indeed rated as more plausible, $F(1,45) = 449.82$, $p<.001$.

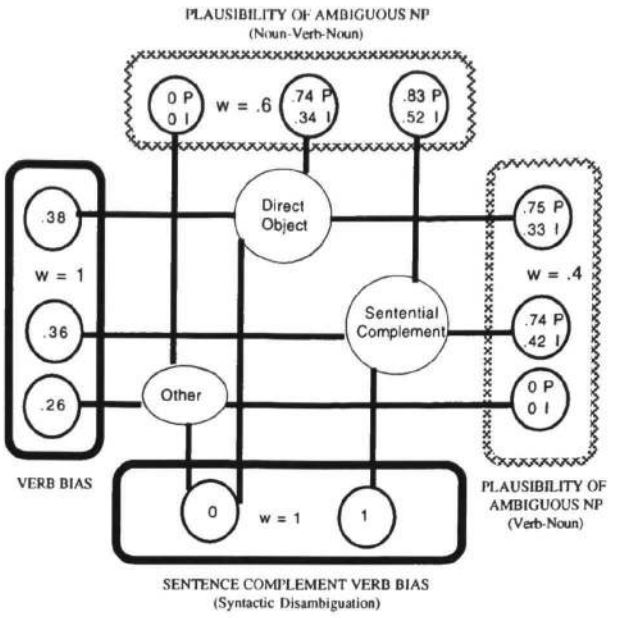

Figure 3: The model used to simulate the role/filler norms.

Ratings as Subject of Sentence Complement

Norming technique interacted with plausibility, $F(2,45) = 12.02$, $p<.001$. Planned comparisons revealed that this interaction occurred because there were no differences among the ratings for plausible nouns: Garnsey et al. versus role/filler with subjects included, $F(1,89) = 2.21$, $p>.1$; Garnsey et al. versus role/filler without subjects, $F<1$; and the two role/filler norms, $F(1,89) = 1.42$, $p>.1$. In contrast, the nouns labeled as implausible DOs were rated as more plausible in the Garnsey et al. norms than in the role/filler norms without subjects, $F(1,89) = 32.68$, $p<.001$, and with subjects, $F(1,89) = 17.58$, $p<.001$. Finally the two role filler norms did not differ, $F(1,89) = 2.32$, $p>.1$.

As above, there were main effects of norming technique, $F(2,45) = 14.08$, $p<.001$, and plausibility, $F(1,45) = 449.82$, $p<.001$.

The role/filler norming results indicate why Garnsey et al.'s plausibility manipulation produced relatively weak results. The strongest effect of this variable would be expected if the ratings had shown a high-low versus low-high pattern. Garnsey et al.'s norms suggest at least a high-high (plausible DOs) low-high (implausible DOs) pattern. However, the role/filler norms suggest a weaker manipulation of plausibility was achieved (high-high vs low-medium). In the following simulations, we examine whether these differences account for the inability of the constraint-based simulation in Part 1 to simulate the human data. The values of the role filler norming studies were entered into the constraint-based model shown in Figure 4, and the results were contrasted with the constraint-based simulation that used the norms collected by Garnsey et al.

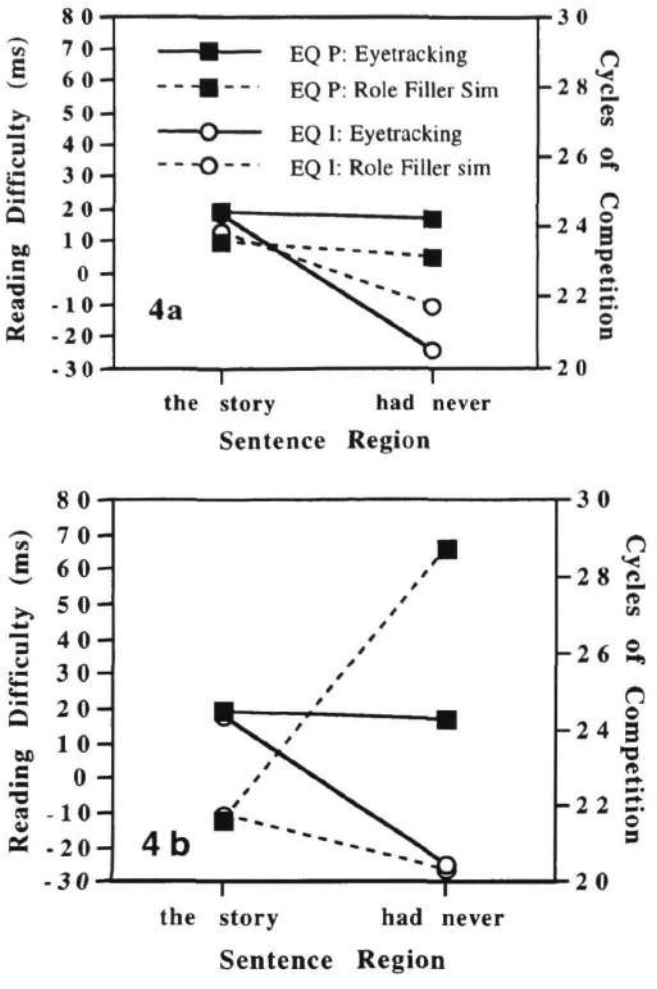

Figure 4: The mean cycles of competition produced by the constraint-based simulations are plotted with the eyetracking results of Garnsey et al. 4a shows the simulation that used plausibility as measured by the role/filler norms, whereas 4b shows the simulation that used the Garnsey et al. norms.

Simulation Results

As shown in Figure 4, the simulation using the role/filler norms to index plausibility more closely captured the eyetracking data. The main reason for the discrepancy is that the Garnsey et al. norms overestimate the difference between the plausible and implausible DOs, thus resulting in too large of a difference between these types of items at the "had never" disambiguating region. However, the norming techniques clearly make different predictions about the amount of difficulty readers should have for plausible versus implausible nouns at the disambiguation. Specifically, for the ratings collected by Garnsey et al., the model is producing more competition for the plausible nouns at the disambiguation than at the ambiguous NP. For both the role filler norms and the human data, slightly less competition (or reading difficulty) occurred for plausible nouns at the disambiguation than at the ambiguous NP. This suggests that the role/filler norms are a better measure of plausibility.

Conclusion

The simulations contrasting the garden-path and constraint-based approaches suggest that the eyetracking data collected by Garnsey et al. is better captured by a model in which all sources of information are used immediately. Moreover, the human eyetracking results were best simulated by a model that used a plausibility measure based on world knowledge abstracted away from linguistic knowledge. Understanding plausibility constraints that operate on a moment-by-moment basis during on-line sentence comprehension is best indexed by norming methods that tap knowledge of situations/events, and by an architecture in which all sources of information are immediately evaluated.

Acknowledgements

This research was supported by an NSERC Doctoral Award to the first author, and NSERC Grant 155704-98 to the second. We thank Susan Garnsey for graciously providing us with all the data necessary to conduct this research.

References

Frazier, L. (1987). Theories of syntactic processing. In J. Garfield (Ed.), *Modularity in Knowledge Representation and Natural Language Processing*. Cambridge, MA: MIT Press.

Frazier, L. & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178-210.

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.

MacDonald, M. Pearlmutter N. & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*,676-703.

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

McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes, 12*, 137-176.

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.

Rayner, K. & Pollatsek, A. (1989). The psychology of reading. Englewood Cliffs, NJ: Prentice Hall.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*, 1521-1543.

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