Evidence for Self-Organized Sentence Processing: Local Coherence Effects

Whitney Tabor and Bruno Galantucci
University of Connecticut

Daniel Richardson
Cornell University

Submitted to the *Journal of Memory and Language*

Please address correspondence to: Whitney Tabor
Department of Psychology, U-20
University of Connecticut
Storrs, CT 06269

860 486 4910 (phone)
860 486 2760 (fax)

tabor@uconnvm.uconn.edu

Abstract

A central question for psycholinguistics concerns the role of grammatical constraints in online sentence processing. Many current theories maintain that, at each point in time, the processor constructs an analysis which is consistent with the grammatical information in the input. Several bottom-up, dynamical models make a contrasting prediction: partial parses which are syntactically incompatible with the current input can nevertheless temporarily influence the processor. Three self-paced reading experiments demonstrated the influence of such local structural ambiguities. The first examined the distracting effect of irrelevant Subject-Predicate interpretations of Noun Phrase-Verb Phrase sequences (*The coach chided the player tossed a frisbee*) on reading times. The second revealed an important role for semantics by manipulating semantic coherence in a variant of the first experiment. The third replicated the results in grammaticality judgments instead of reading times. We interpret the results as evidence for a self-organizing parsing mechanism.

Keywords: parsing, sentence processing, garden-path, syntactic ambiguity, dynamical system, self-organization, bottom-up processing

INTRODUCTION

Much has been learned about the nature of sentence processing by studying temporarily ambiguous sentences like (1) – (3):

(1) The mechanic maintained the truck was working beautifully.

(2) The cop arrested by the detective was chagrined.

(3) The cook stirred the soup with the tomatoes.

Each of these sentences has a structural ambiguity which is resolved, either semantically or syntactically, in favor of one structure when the underlined words arrive. Self-paced reading and eye tracking studies show that when such sentences are presented in isolation, readers tend to slow down and/or make regressions in the disambiguating region (underlined), which suggests that they either choose the wrong parse initially or are at least pulled somewhat off course by it.

Most current theories of sentence processing interpret these results by assuming that the processor builds sentence analyses incrementally, fitting each successive word into the current partial-parse. Difficulty arises when several parses are sanctioned by the grammar and the reader either (a) initially makes a wrong choice (serial models) or (b) activates multiple parses in parallel with the wrong one ranked highest and must revise the ranking (parallel models). Theories of ambiguous parsing in this class include the standard Garden Path Theory (Ferreira and Clifton, 1986; Ferreira and Henderson, 1990; Frazier and Rayner, 1982; Frazier, 1988; Frazier and Clifton, 1996), Reanalysis as a Last Resort theories and monotonic parsing theories (Fodor and Frazier, 1980; Frazier and Clifton, 1998; Gorrell, 1995; Phillips, 1998; Schneider and Phillips, 2001), Race-Based Parsing theories (Traxler, Pickering, and C. Clifton, 1998; Gompel, Pickering, and Traxler, 2000, 2001), Syntactic Prediction Locality Theory (Gibson, 1998; Grodner, Gibson, and Tunstall, 2002), parallel parsers which use the input to assign weights to various grammatical options (Gorrell, 1987; Gibson, 1991; Jurafsky, 1996; Hale, 2002).

Here, we consider an alternative possibility: it may be that local syntactic coherence in the input can prompt the construction of syntactic analyses which are not consistent with the input in its entirety.[1] We hypothesize that such errant local parses can have a detectable influence on the time course of processing.

[1] In keeping with our specific interest in syntactic processing, when we say "the input in its entirety", we mean the verbal input since the end of the last sentence, although we acknowledge that comprehensive theories of language processing need to adopt a sense of input that extends beyond mere text and includes the rich prosodic, pragmatic, and sensory context that normally accompanies language use (e.g., Crain and Steedman, 1985; Altmann and Steedman, 1988; Sedivy, 1999; Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy, 1995; Tanenhaus, Spivey-Knowlton, and Hanna, 2000).

Many thanks to Ino Flores d'Arcais, Katherine Finerty, Carol Fowler, Edward Gibson, Sean Hutchins, Leonard Katz, Neal Pearlmutter, David Perkowski, Aaron Schultz, Michael Tanenhaus, and Michael Spivey for helpful contributions and/or comments. This research was supported by NIH Grant 5 T32 MH19389 and by University of Connecticut Research Foundation Grant FRS 444078.

By "local syntactic coherence in the input" we mean sequences of two or more words in the text stream which form a phrase or part of a phrase such that the phrase cannot be grammatically unified with the parse of the preceding words in the sentence. An example is provided in (4).

(4) The coach chided the player tossed a frisbee by the opposing team.

In (4), "the player tossed a frisbee" has a locally coherent interpretation (both syntactically and pragmatically) as an active clause. But, in the context of the sentence, which strongly favors interpretation of "the player" as the direct object of "chided", "tossed a frisbee" can only be grammatically parsed as a reduced relative clause modifier of "player". We wondered if readers would nevertheless be distracted by the active interpretation. The experiments we discuss below were designed to determine whether such merely locally coherent subsequences can have a distracting effect on the parser.

There is a class of parsing models which are well-positioned to predict interference due to local coherences like the one just discussed. These are activation-based syntactic models in which large phrase structure trees are built up via the growth of bonds among open attachment sites on lexically activated tree fragments (e.g., Kempen and Vosse, 1989; Stevenson, 1994; Stevenson and Merlo, 1997; Stevenson, 1997, 1998; Vosse and Kempen, 2000). If all possible bonds are allowed to form initially, with competition among the bonds allowing only the most felicitous ones to grow strong, then interference from merely locally coherent structures is predicted: a piece of tree structure will emerge (at least temporarily) over the locally coherent segment and this will compete with any globally coherent structure, slowing reading. In fact, the models of Stevenson and Vosse and Kempen include special constraints that prevent such local parse formation, but we think that a simpler model, without the special constraints, is preferable, a point that we return to in the CONCLUSION. The present paper examines some distinguishing empirical predictions of this viewpoint.

Models which allow the temporary formation of local parses make a significant divergence from prevailing assumptions about the mechanism of sentence processing. We call these models "self-organizing" because different features of the input announce constraints separately without regard to compatibility with the other pieces, and the resolution of conflicts among these constraints is an important determinant of online processing behavior. In particular, there is no global constraint-checking system (or "central executive") which enforces properties that the whole parse must possess (contrast Frazier, 1990). Instead, it is via the interactions among the syntactic fragments that the selection of a global structure is made.

Motivations

We believe that the self-organization approach can provide unifying insight into a wide range sentence processing phenomena. Vosse and Kempen (2000) show, for example, that their self-organizing model makes appropriate predictions about a number of complexity contrasts (e.g., single versus double embeddings, center embedding versus right branching) as well as about a number of garden path effects. Under their model, both phenomena involve incorrect attachments getting in the way of correct ones. Stevenson (1998) has argued that while many models posit separate mechanisms for handling initial analysis and reanalysis, her Competitive Attachment parser generates both phenomena as a consequence of the interaction of linguistic structures with the bonding mechanism: in essence, bonds that have formed previously because they best satisfied the featural requirements of the bonding elements (initial analysis) will dissolve automatically and new bonds will form (reanalysis) if the weight of later-arriving, disambiguating information is sufficient to overcome the strength of the original attachments. Tabor and Hutchins (2003a) show that a similar model unifies ambiguous region length effects (longer ambiguous regions make for harder disambiguations; e.g. Ferreira and Henderson, 1991, 1993) with an effect of the number of erroneous attachments (more attachments are harder to undo): both cases stem from quantitative summation of constraints which is central to allowing coherent structures to emerge under the self-organization approach.

The cases just mentioned provide an indication of the potential generality of the self-organization approach, but most of these cases also have explanations in other paradigms. The hypothesized local coherence phenomena of the present study constitute a particularly distinctive prediction of the self-organization approach. In this sense, a central goal of the present study is to provide distinguishing empirical evidence for self-organized sentence processing.

Overview

In the next section, we note that the well-known phenomenon of simultaneous initial activation of all the senses of ambiguous words can be thought of as reflecting a kind of self-organization, though one that is readily handled via standard assumptions about the relationship of syntax and lexicon. We then discuss three reading experiments which we conducted to test for self-organization effects in syntax proper. The CONCLUSION reviews a variety of possible explanations for the findings. Because there are several alternative accounts which deserve careful comparison, we do all of the significant discussion in the CONCLUSION.

Related prior work

In a priming study, Swinney (1979) found evidence that when speakers read a sentence like (5) there is a brief period of time during which both the irrelevant sense of "bug" (target = "spy") and the relevant sense (target = "insect") are activated.

(5) The man was not surprised when he found several spiders, roaches, and other bugs...

The activation of the irrelevant sense then dies out under the influence of context. Tanenhaus, Leiman, and Seidenberg (1979) found evidence that a similar phenomenon occurs with syntactic class ambiguity. They presented examples like (6) auditorily and probed with a visual prime that appeared coincidentally with or at various intervals after the last word of the sentence. They found that there was a brief period of time (< 200ms) when both the noun and the verb senses of the final word produced priming (measured via naming latencies), even though the preceding syntactic context is only compatible with one of them.

(6) a. He cut the roll.
    b. They began to roll.

Tanenhaus et al. (1979)'s interpretation of these results is that they reveal a modular relationship between syntax and lexicon: for a brief interval after the perceptual system makes contact with a new word, the lexical module assesses the input independently of the syntactic context. This account is compatible with a restricted form of self-organization which we term "Syntax/Lexicon" self-organization: the syntax (reflecting information in the prior context) and the lexicon (reflecting information in the current word) impose independent constraints on the parse, and conflicts sometimes arise. Coherent parsing involves the resolution of these conflicts.

Syntax/Lexicon self-organization is quite compatible with the many models of sentence processing which make a principled division between syntactic and lexical modules of the grammar. Here, we wish to test a more challenging hypothesis: that even multi-word syntactic structures incompatible with the current syntactic context are constructed if they are sufficiently locally coherent. The experiments reported below were designed to test for "Homogeneous Self-Organization" in this sense.

EXPERIMENTS

Experiment 1: Clauses

Reduced relative clauses have been extensively studied in the case where they occur as modifiers of nouns in subject position in matrix clauses, as in (7) (e.g., Bever, 1970; Ferreira and Clifton, 1986; Frazier, 1978; McRae, Spivey-Knowlton, and Tanenhaus, 1998; Rayner, Carlson, and Frazier, 1983; Trueswell and Tanenhaus, 1994; Trueswell, 1996).

(7) The player tossed a frisbee looked surprised.

The evidence indicates that when it is semantically sensible to interpret the verb following the subject noun as the main verb of the clause, readers have a strong tendency to do so. Consequently, they become confused starting around the word "looked" in a case like (7) because this word disambiguates in favor of the relative clause reading.

Here, we consider cases like (8).[2]

(8) The coach chided the player tossed a frisbee by the opposing team.

Because the verb "chided" requires, in its favored active sense, a direct object, the second verb, "tossed", must be interpreted as a passive verb introducing a reduced relative clause which modifies the noun phrase, "the player". But, as the prior studies have shown, "the player tossed a frisbee" in isolation is prone to be interpreted as a transitive construction with an active verb. Our hypothesis is that readers will be distracted by this active interpretation in a way similar to the way they are distracted by the active interpretation in (7) despite the fact that the active reading is syntactically impossible in the context.

[2]We thank Neal Pearlmutter and Edward Gibson for (independently) suggesting experiments along the lines of Experiment 1.

Method

Participants.

Forty-seven subjects were recruited from classes and through advertisement on the campus of the University of Connecticut. All were native speakers of English and none had participated in a related experiment before. Subjects received either money or course credit for their participation. The paid subjects received $10 per hour. The experiment usually lasted about 30 minutes.

Materials.

Twenty experimental items were created. These are listed in Appendix 1. Each item involved four conditions as in (9):

(9) Example stimuli for Experiment 1 (A = Ambiguous; U = Unambiguous; R = Reduced; UR = Unreduced).

-5	-4	-3	-2	-1
The	coach	chided	the	player...

			0	1	2	3	
a.			tossed	a	frisbee	by...	(AR)
b.	who	was	tossed	a	frisbee	by...	(AU)
c.			thrown	a	frisbee	by ...	(UR)
b.	who	was	thrown	a	frisbee	by ...	(UU)

Each item included a noun phrase in a non-subject position which was modified by a relative clause in passive voice. Two dimensions of contrast in the relative clause gave rise to four conditions for each item. The relative clause was either reduced (R) (9a and c) or unreduced (U) (9b and d); its verb was either ambiguous (A) (i.e. the past participle and past tense forms were homophonous and homographic—9a and b) or unambiguous (U) (9b and d).

We were looking for an effect of Reduction for the Ambiguous case. If this effect obtained and the reduced cases were read faster, starting at Word 0, than the unreduced cases, then the slowing might be caused by interference from the local structural ambiguity. However, it would be premature to take such a result on its own as evidence for the existence of self-organization at the syntactic level. Greater speed of processing might occur at the relative clause verb in (9b) simply because the syntax is more constraining at this point in case (9b) than in case (9a). That is, it is generally the case that processing speed is faster at grammatical events that are more expected (e.g., Hale, 2002; Jurafsky, 1996; McRae et al., 1998; Tabor, Juliano, and Tanenhaus, 1997) and the sequence "who was" makes a following passive participle very expected. Thus, there should be a slowing effect of Reduction in the Unambiguous case as well ((9d) vs. (9c)). For this reason, we employed the more complex 2 x 2 design. We predicted that reduction would slow processing in both cases (9a) and (9c), but would slow it more in (9a) than in (9c).

Four counterbalanced lists of items were constructed. Each list consisted of 20 blocks of items. Each block contained 4 filler items and 1 stimulus item. The stimulus items were never first in a block. The position of the stimulus item within a block was chosen randomly, subject to the condition that the first sentence of a block had to be a filler. The first 10 items of the experiment were fillers. Each subject read six practice trials and then read one of the lists.

Procedure.

The stimuli were presented using the moving-window self-paced-reading method of Just, Carpenter, and Wooley (1982):

Details of Self-Paced Reading. The initial screen for each trial showed an image of the written sentence in which dashes replaced all the letters and punctuation marks, but the spaces between words appeared as spaces. The subject would then push the spacebar on the computer's keyboard, and the first word of the sentence would appear in place of the corresponding dashes. The subject would then push the spacebar again causing the first word to revert to dashes and the second word to replace the dashes in the second word position. This method repeated until the last word was read, whereupon a Yes/No question testing the reader's comprehension of the sentence appeared. The subject answered the question by pushing the f-key for No and the j-key for Yes. If the answer was incorrect, the word "Incorrect" flashed briefly on the screen before the next trial began. If the answer was correct, the next trial began immediately. This presentation method is called "noncumulative" because once a word has been read, it turns back into dashes and subjects do not have the option of reading it again. Subjects were encouraged to read as naturally as possible and to answer the questions according to their first impulse. The program offered them a resting period after every 20 sentences. The program recorded the length of the time interval between successive keypresses as well as the answers the subjects gave to the questions.

The experiment was executed by a PsyScope program (Cohen, MacWhinney, Flatt, and Provost, 1993) written by David Perkowski and Daniel Richardson. The program was run on MacIntosh GE3s with 14 inch monitors. The target and control sentences were designed so that the section of the sentence from the beginning up to the fifth word beyond the ambiguous verb (or the corresponding verb in the control case) fit on one line of the computer screen.

Results

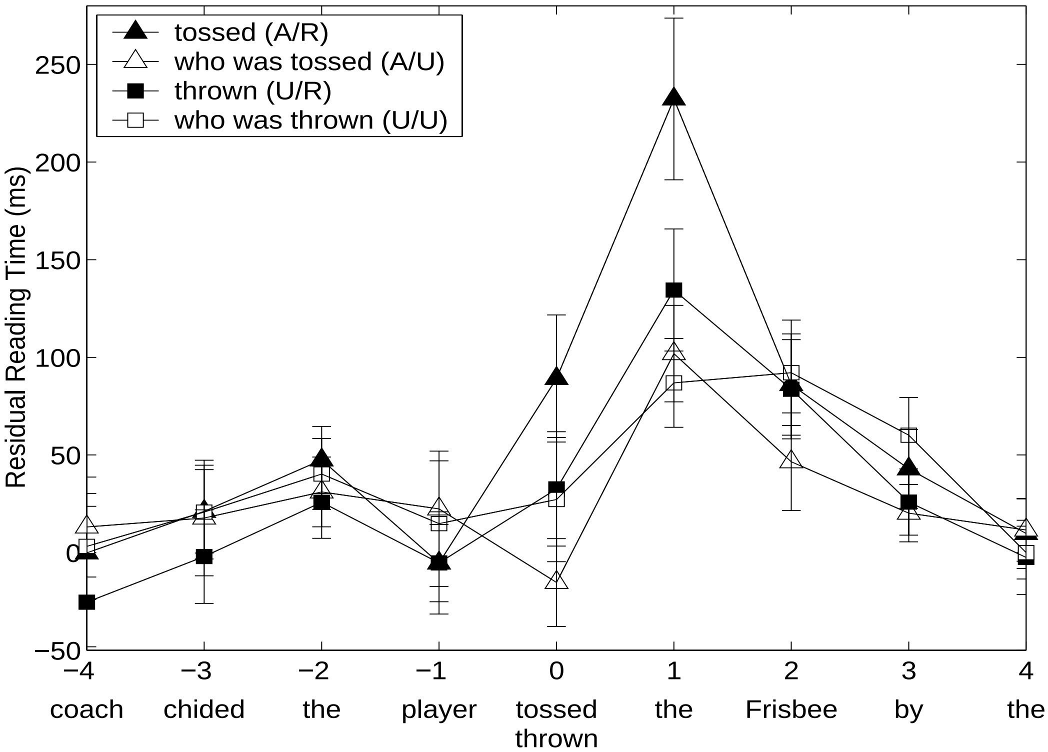
The data from 9 subjects who missed more than 20% of the comprehension questions (stimuli and fillers combined) were removed from the analysis. Due to an implementation error, two of the items (numbers 19 and 20) were never presented to the subjects.

Before analyzing the reading times, we replaced each data point that was more than 4 standard deviations from the global mean reading time (measured over all positions and all subjects) with the value at 4 standard deviations from the mean at the position in question (global mean = 412 ms; 4 s.d.'s = 1037 ms). This adjustment of outliers affected 0.82% of the data. Likewise, for each subject, we performed a linear regression on the adjusted reading times with characters-per-word as independent variable. The variance associated with word length was again small (mean $R^2 = 0.027$), but significant ($p < 0.0001$). The analyses described below were performed on the standardized residuals from this regression analysis.

Figure 1 shows the mean (trimmed, length-corrected) reading times across all positions for the four conditions of Experiment 1.

We analyzed the reading times by dividing each sentence into three regions across

Figure 1. Mean residual reading times from Experiment 1. Error bars show one standard error around each data point.



which we expected contrast. It was at the second verb ("tossed"), that the distracting analysis was first supported by a combination of words, so we refer to this verb as the "critical word". We defined Region 1 to consist of Words -4 through -1. We expected a distracting effect to develop at the critical word or shortly after it, so we defined Region II to consist of Words 0 through 3; The end of the line cut some sentences off as soon as Word 5, so Region III consisted of Word 4.

For each region of interest, subject and item means were subjected to separate analyses of variance (ANOVAs), each with two factors: Ambiguity and Reduction. At Region I, there were no significant main effects or interactions. At Region II, there was a main effect of Reduction (F1(1, 37) = 8.997, MSE = 6044, p = .005 (**); F2(1, 17) = 12.561, MSE = 3291, p = .002 (**)) with slower reading times for the reduced than the unreduced. The predicted interaction between Ambiguity and Reduction was also significant (F1(1, 37) = 18.220, MSE = 2590, p < .001 (***); F2(1, 17) = 6.473, MSE = 2190, p = .021 (*)). At Region III, there were no significant main effects or interactions. We define the *Interaction Strength* as $(\overline{HR} - \overline{HU}) - (\overline{NR} - \overline{NU})$. The average interaction strengths at Regions I, II, and III, were 16, 72, and 1, respectively.

The individual significances by word position are shown in Table 1. The last line of Table 2 gives the word-by-word Interaction Strengths. Because the individual tests showed one significant effect at a position prior to the manipulation (the effect of Reduction at

Position -1 was significant in the Items analysis: F1(1, 37) = 1.373, MSE = 15874, p = .249; F2(1, 17) = 4.502, MSE = 1635, p = .049 (*)), we tested the interaction between Reduction and Region. Indeed, this interaction was significant (F1(2, 74) = 8.297, MSE = 3304, p = .001 (**); F2(2, 34) = 7.113, MSE = 2306, p = .003 (**)). In fact, the reduced conditions were read faster by 23 ms at Position -1 and by 13 ms on average over Region I—the opposite direction from the observed reduction effects in Regions II and III.

Table 1: Word-by-word significances for the Ambiguity x Reduction interaction of Experiment 1. IS = Interaction Strength = $(\overline{HR} - \overline{HU}) - (\overline{NR} - \overline{NU})$.

	-4 coach	- 3 chided	-2 the	-1 player	0 tossed	1 a	2 frisbee	3 by	4 the
F1	< 1	< 1	1.119	< 1	6.725	3.158	2.727	8.137	< 1
MSE	–	–	8096	–	13851	20626	8012	3780	–
p1	–	–	0.297	–	0.014 (*)	0.084 (*)	0.107	0.007 (**)	–
F2	< 1	< 1	1.618	< 1	6.425	2.371	1.578	1.417	1.172
MSE	–	–	1787	–	6647	4763	4931	3896	1781
p2	–	–	0.220	–	0.021 (*)	0.142	0.226	0.250	0.294
IS (ms)	15	26	31	-8	99	83	48	57	1

Discussion

The self-organization view predicts the particularly slow reading times in the Ambiguous/Reduced condition of Experiment 1 because the subsequence "the player tossed a frisbee" forms an active clause which competes with the globally consistent reduced relative interpretation. If readers were not putting these words together into an active clause, but instead were simply laboring under some difficulty associated with processing a reduced relative modifier of a direct object, then they should have found the unambiguous version, "the player thrown the frisbee", equally hard. In the CONCLUSION we compare this account to several other ways of accounting for the critical interaction.

Experiment 2: Role of Semantics

Under the view of self-organized parsing presented here, the rate of growth of bond strength depends on feature match between the bonding nodes. We assume, in keeping with experimental results indicating that semantic well-formedness modulates the speed of processing (e.g., Ford, Bresnan, and Kaplan, 1982; Kim, 2000; MacDonald, Pearlmutter, and Seidenberg, 1994; McRae et al., 1998; Rayner et al., 1983; Trueswell, Tanenhaus, and Garnsey, 1994; Tanenhaus, Carlson, and Trueswell, 1989; Tanenhaus and Trueswell, 1995) that the features on the nodes include both semantic and syntactic features. Thus, we predicted that if an embedded "NP V NP" sequence like that which caused distraction in Experiment 1 did not exhibit semantic coherence, then its distracting effect should be much reduced. Experiment 2 was therefore designed to replicate the main finding of Experiment 1 using a semantic control instead of a morphological one.

Participants.

Eighty-six subjects were recruited from classes on the campus of the University of Connecticut. All were native speakers of English and none had participated in a related experiment before. They received course credit for their participation. The experiment usually lasted about 30 minutes.

Materials.

We considered reusing the Ambiguous stimuli from Experiment 1 and creating controls that lacked semantic coherence in the distractor sequence. However, we found that it was hard to design felicitous sentences with substantial variation in semantic plausibility under this rubric because the constraints on the semantic attributes of a ditransitive beneficiary are quite strict (For example, if our hypothesis about the role of semantics is correct, then "The coach chided the tree tossed a frisbee" is not likely to show elevated reading times due to a distracting active parse of "the tree tossed a frisbee" (because this interpretation is implausible), but it is likely to show elevated reading times due to the anomalousness of "the coach chided the tree" and the "tree (was) tossed a frisbee", so it would be a poor control). Therefore, we created a variant on the theme by replacing the ditransitive ambiguous verbs with transitive verbs. To take advantage of semantic biases which have already been experimentally measured, we used a subset of the noun phrase + transitive verb combinations employed by Trueswell et al. (1994), a study which examined the effect of noun phrase animacy on the processing of standard Main Verb/Reduced Relative ambiguities. We started by creating contrast pairs like (10a-b).

(10) a. The bandit worried about the prisoner transported by the capricious guards.
b. The bandit worried about the gold transported by the capricious guards.

Although we expected a distracting effect of semantic cohesion for a simple sequence like "the prisoner transported" (versus "the gold transported"), we thought an even longer local ambiguity might produce a more pronounced effect, so we added time or path adverbials of the form "the" + Adjective + Noun (e.g., "the first time", "the whole way") after each ambiguous verb. To control for potential reading time differences stemming from lexical contrasts between the manipulated nouns, we included a reduction manipulation like that of Experiment 1. An example item from the resulting 2 x 2 (Animacy x Reduction) design is shown in (11). Appendix 2 lists the stimuli for Experiment 2. Four counterbalanced lists of items were created, as in Experiment 1.

(11) Example stimuli for Experiment 2. (A = Animate; I = Inanimate; R = Reduced; U = Unreduced)

The bandit worried about ...

	-2	-1			0	1	2	3	4	
a.	the	prisoner			transported	the	whole	way	by	(A / R)
a.	the	prisoner	who	was	transported	the	whole	way	by	(A / U)
a.	the	gold			transported	the	whole	way	by	(I / R)
a.	the	gold	that	was	transported	the	whole	way	by	(I / U)

... the capricious guards.

Procedure.

The procedure was the same as in Experiments 1 and 2 except that some of the participants were tested in groups of up to 8 at a time. They sat at separate computer terminals, spaced at least 3 feet apart from one another and wore headphones to block out ambient noise.

Results

The data from fifteen subjects who missed more than 20% of the comprehension questions (stimuli and fillers combined) and from one subject whose computer crashed partway through the experiment were removed from the analysis.[3]

As in Experiment 1, before analyzing the reading times, we replaced each data point that was more than 4 standard deviations from the global mean reading time (measured over all positions and all subjects) with the value at 4 standard deviations from the mean at the position in question (mean = 408 ms; 4 s.d.'s = 1005 ms). This adjustment of outliers affected 0.85% of the data. Likewise, for each subject, we performed a linear regression on the adjusted reading times with characters-per-word as independent variable. The variance associated with word length was again small (mean $R^2 = 0.016$), but significant ($p < 0.0001$). The analyses described below were performed on the standardized residuals from this regression analysis.

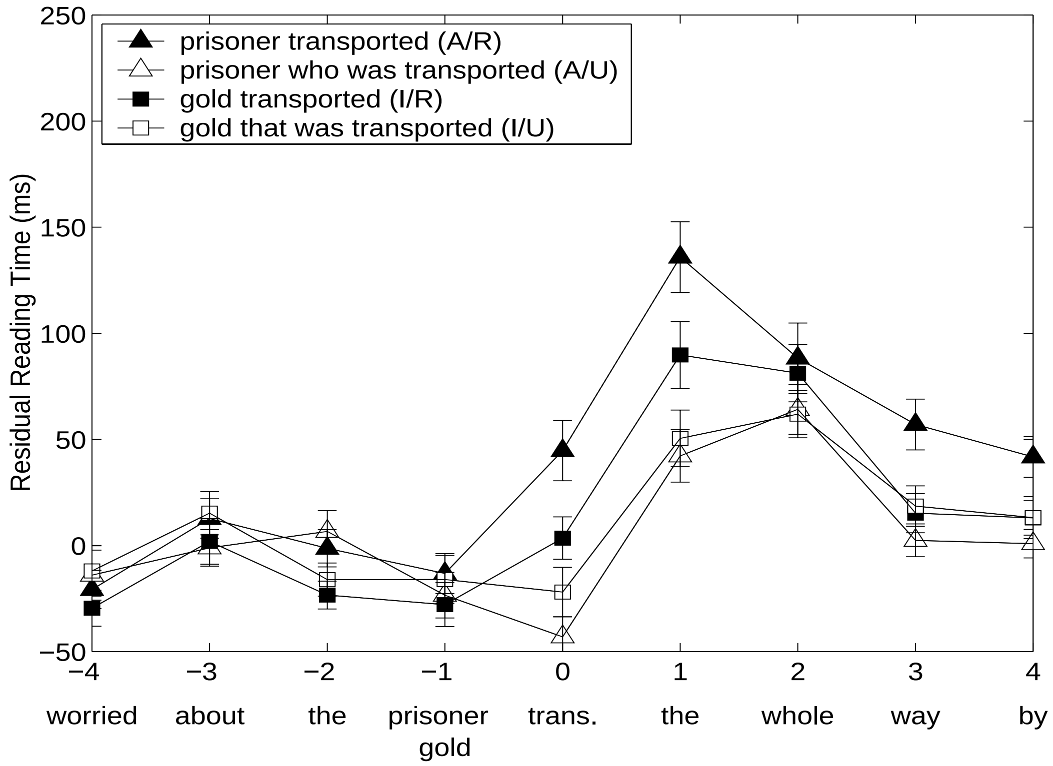
Figure 2 shows the mean (trimmed, length-corrected) reading times across positions for the four conditions of Experiment 2.

As in Experiment 1, we analyzed the reading times by dividing each sentence into three regions across which we expected contrast. Again, it is at the second verb (here, "transported"), that the distracting analysis is first supported by a combination of words, so Word 0 is the critical word. Region I consisted of Words -4 through -1; Region II consisted of Words 0 through 3; Region III consisted of Word 4 (again, the last word on the line was sometimes Word 5).

For each region of interest, subject and item means were subjected to separate analyses of variance (ANOVAs), each with two factors: Animacy and Reduction. At Region I, there were no significant main effects or interactions. At Region II, there was a main effect of Reduction ($F1(1, 69) = 44.336$, MSE = 2865, $p < .001$ (***); $F2(1, 19) = 10.512$, MSE = 2873, $p = .004$ (**)). The predicted interaction between Animacy and Reduction was significant ($F1(1, 69) = 15.688$, MSE = 2245, $p < .001$ (***); $F2(1, 19) = 7.043$, MSE = 2522, $p = .016$ (*)). At Region III, the main effect of Reduction was significant in the Subjects analysis and marginally significant in the Items analysis ($F1(1, 69) = 6.455$, MSE = 4494, $p = .013$ (*); $F2(1, 19) = 4.152$, MSE = 2821, $p = .056$). The Region III interaction between Animacy and Reduction was significant in the Subjects analysis only ($F1(1, 69) = 6.229$, MSE = 4731, $p = .015$ (*); $F2(1, 19) < 1$). The individual significances by word position are shown in Table 2. The average Interaction Strengths by region were: Region I—15 ms/word; Region II—45 ms/word; Region III—41 ms/word. The individual tests showed one small (23 ms) but significant main effect at a position prior to the manipulation: the

[3]Due to an error in the implementation, the distribution of trials over conditions for two items (16 and 17) was strongly asymmetric. Nevertheless, the results of the analysis were not significantly different when these two items were removed. Therefore, they are included in the analyses reported below.

Figure 2. Mean residual reading times from Experiment 2. Error bars show one standard error around each data point.



effect of Animacy at Position -2 (F1(1, 69) = 6.518, MSE = 5372, p = .013 (*); F2(1, 19) = 20.300, MSE = 835, p < .001 (***)).

Discussion.

The effect of interest is the interaction between Animacy and Reduction in Region II. As predicted by the self-organization account, the slowing of reading times under Reduction is greater in the Animate condition than the Inanimate. The slowing is due, on this account, to interference from the active parse of the second NP-V-NP sequence (e.g. "the prisoner transported the whole...."). The fact that the interference is stronger when the noun at Word 0 is animate than when it is inanimate is consistent with the supposition that semantic, as well as syntactic, coherence drives the formation of the irrelevant parse, thus confirming and augmenting the information provided by Experiment 1.

Experiment 3: Grammaticality

Experiment 3 was designed to see if the critical interaction of Experiments 1 and 2 would show up in grammaticality judgments. We also wanted to assess the role of grammaticality in determining our findings, particularly in light of the fact that Experiment 2 involves a fairly obscure construction—Beneficiary extraction from ditransitive verbs—which some speakers may not find grammatical.

Table 2: Word-by-word significances for the Animacy x Reduction interaction of Experiment 2. IS = Interaction Strength = $(\overline{AR} - \overline{AU}) - (\overline{IR} - \overline{IU})$.

	-4 worried	-3 about	-2 the	-1 prisoner / gold	0 transported	1 the	2 whole	3 way	4 by
F1(1, 69)	< 1	2.254	< 1	1.764	13.761	4.777	< 1	10.917	6.229
MSE	–	5721	–	4822	4945	10821	–	5382	4731
p1	–	.138	–	.188	< .001 (***)	.032 (*)	–	.002 (**)	.015 (*)
F2(1, 19)	< 1	1.408	< 1	2.355	5.983	3.847	< 1	9.488	< 1
MSE	–	2385	–	3870	4318	6711	–	5527	–
p2	–	.250	–	.141	0.024 (*)	.065	–	0.006 (**)	–
IS (ms)	11	27	-1	22	62	54	5	58	41

Participants.

Twenty-eight subjects were recruited by advertisement on the campus of the University of Connecticut. All were native speakers of English and none had participated in a related experiment before. They were paid at a rate of $10 per hour for their participation in the experiment. The experiment usually lasted about 30 minutes.

Materials.

Experiment 3 was a composite of two subexperiments, 3.1 and 3.2. The stimuli for these subexperiments were shortened versions of the stimuli from Experiments 1 and 2, respectively. Each stimulus ended after the agentive by-phrase (as indicated in Appendices 1 and 2), because we wanted to minimize the distracting effects of extraneous material. Four lists of items were constructed. Each list included 40 blocks of items, 20 from Experiment 1 and 20 from Experiment 2. Each block contained 2 filler items and 1 stimulus item. The stimulus items were never first in a block. Half of the filler sentences were grammatical and half were ungrammatical. An example of an ungrammatical filler sentence is "The girls consumed the grapefruit which the cat". The position of the stimulus item within a block was chosen randomly, subject to the condition that the first sentence of a block had to be a filler. Each subject read six practice trials and then read one of the ($3 \times 40 =$) 120-trial lists.

We predicted that in each subexperiment, the positive grammaticality judgment rate would interact in the same way as the reading times did in the corresponding reading experiments.

Procedure.

The procedure was the same as in Experiment 2 except that each sentence was followed by the question,"Grammatical? f: Yes j: No" and no feedback was provided.

Results

No subject disagreed with our judgments on more 20% of the filler trials and all of the data were used in the analysis. Table 3 shows the mean values of the grammaticality

Table 3: Mean rates of positive grammaticality judgments for the two subexperiments and the fillers in Experiment 3. The values shown in parentheses are standard errors.

(a) Subexperiment 3.1.

Ambiguous/Reduced	0.22 (0.040)
Ambiguous/Unreduced	0.72 (0.033)
Unambiguous/Reduced	0.39 (0.042)
Unambiguous/Unreduced	0.72 (0.038)

(b) Subexperiment 3.2.

Animate/Reduced	0.58 (0.045)
Animate/Unreduced	0.82 (0.030)
Inanimate/Reduced	0.75 (0.035)
Inanimate/Unreduced	0.84 (0.032)

(c) Fillers

Grammatical	0.93 (0.012)
Ungrammatical	0.08 (0.013)

judgments for each condition and subexperiment.

The two subexperiments were analyzed separately. Subject and item means were subjected to separate analyses of variance (ANOVAs). The strengths of the main effects and interactions are shown in Table 4. As the tables indicate, Reduction has the effect of reducing the rate of positive grammaticality judgment in both subexperiments and it reduces the rate more in the Ambiguous/Animate conditions than the Unambiguous/Inanimate conditions. In particular, the predicted interaction between Ambiguity/Animacy and Reduction was significant in both experiments.

Discussion.

The pattern of results in this experiment conforms to the pattern in the previous two experiments. Moreover, Subexperiment 3.1 suggests that our participants did indeed find the Beneficiary extraction construction ungrammatical in most of the Reduced cases. However, they were significantly more likely to find this construction grammatical in the Unreduced cases. We interpret these findings in the CONCLUSION.

CONCLUSION

We have reported on three experiments which provide evidence that merely locally-coherent subsequences of words in the input can slow procesing. We now identify a number of explanations for these findings. We note that several previously proposed theories have elements of self-organization in them: constraints arrive independently from several sources, sometimes resulting in conflict; the ultimate parse (or failure to parse) is the outcome of an attempt by the system to resolve these conflicts. Under several of the theories, the conflicting information comes from distinct modules (e.g., syntax, lexicon, semantics). We suggest an account ("Homogeneous Self-Organization") that involves competition among

Table 4: Experiment 3 ANOVA results.

(a). Subexperiment 3.1: Ambiguity and Reduction.

	Subjects			Items		
Effect	F(1, 35)	MSE	p	F(1, 19)	MSE	p
Ambiguity	12.096	0.0221	.001 (**)	3.776	0.0687	.067
Reduction	148.078	0.0428	< .001 (***)	97.303	0.0371	< .001 (***)
Ambiguity × Reduction	7.321	0.0341	.01 (*)	5.668	0.0229	.028 (*)

(b). Subexperiment 3.2: Animacy Reduction.

	Subjects			Items		
Effect	F(1, 35)	MSE	p	F(1, 19)	MSE	p
Ambiguity	12.364	0.0252	.001 (**)	7.946	0.0224	.011 (*)
Reduction	29.238	0.0325	< .001 (***)	13.977	0.0322	< .001 (**)
Ambiguity × Reduction	5.852	0.0404	.021 (*)	4.785	0.0218	.041 (*)

structures formed within the same (syntactic-semantic) module. The simplest version of this account requires no central executive or top-down controller—coherent parses arise purely out of the interactions among the local constraints created by the perceptions of words. This proposal represents a significant divergence from prevailing assumptions about how parsing works. It remains to be demonstrated that the approach is computationally feasible. It also remains to be demonstrated that it is compatible with the broad range of sentence processing phenomena on which empirical data are available, although we note that the work of Stevenson and of Vosse and Kempen provides some promising indications in this regard. An advantage of the Homogeneous Self-Organization approach is that it minimizes the number of special mechanisms that psycholinguistic theory needs to posit on top of the structures proposed by linguistic theory. The current studies, by providing evidence for the existence for local coherence effects—a hallmark trait of self-organizing processes—motivate further consideration of the computational and broad coverage questions.

Accounts of the findings

Homogeneous Self-Organization.

We believe that the results of the three experiments reflect the bottom-up, self-organizing nature of the parser. In particular, we assume that if a contiguous subsequence of words can form a felicitous semantic and syntactic structure, then the parser will have a tendency to build that structure, regardless of whether it is consistent with the rest of the parse built thus far. In cases of inconsistency, the competition between the correct structure and the erroneous structure slows reading down. Under the assumption that there is some noise in the self-organizing process, the presence of an erroneous structure may keep a coherent global parse from emerging on some portion of instances, thus producing negative grammaticality judgments. Whether an erroneous local parse will produce detectable interference depends on how much competition the erroneous structure has from alternative formations. The Ambiguous/Animate Reduced conditions of the two experiments reported here were designed so that the distracting local sequence made a felicitous structure, while the correct parse involved a dispreferred (past participle) sense of the ambiguous verb. As a consequence, the correct parse encountered stiff competition from the erroneous alternative, and reading times were particularly slowed in this condition.

Stevenson (1994), Stevenson and Merlo (1997), Stevenson (1997, 1998), Vosse and Kempen (2000) have described implemented models which implement an approach along these lines. In their models, the perception of a word causes the activation of one or more lexically-anchored tree fragments, with open attachment sites. The attachment sites of each fragment attempt to bond with the attachment sites of activated fragments in ways that respect their locally-specified grammatical constraints. Such models have, according to our view, the right kind of mechanism to predict the effects we have observed here. Nevertheless, these particular implemented models include additional mechanisms that explicitly rule out syntactic self-organization of the sort we have just described. For example, Stevenson's model includes a Right Edge Restriction which allows each new word to attach only to the right edge of the main parse tree. Kempen has averred (p.c.) that the (Vosse and Kempen, 2000) model is required to stabilize so thoroughly after each word is read that merely local structures are all but ruled out. But we note that these are stipulations that are added to the central (bond-formation) mechanism of each model and that the central mechanism is well-suited to predict the kinds of effects reported here if the added stipulations are removed.[4] Given our own experience with building models of this sort, and in keeping with the comments of an anonymous reviewer, we think it likely that these stipulations play an important role in keeping the models in these projects from erroneously generating many ill-formed structures. But that does not imply that another self-organizing architecture could not succeed at achieving coherence with a less top-down guided mechanism. In fact, in a separate paper, we report on a self-organizing model which allows local structures to temporarily thrive and which has passed a first round of tests on being able to model several complex reanalysis phenomena (Tabor and Hutchins, 2003b).

Minimal Attachment.

The notion that the parser might choose a particular analysis despite the existence of contradictory information in the input is not new. It is, in fact, a central assumption of the well-established Two-Stage (Garden Path) models of parsing (Kimball, 1973; Frazier and Fodor, 1978; Frazier, 1988; Frazier and Clifton, 1996). Under these accounts, Syntax makes the first parse, guided by certain structural principles (e.g., Minimal Attachment, Late Closure), and semantic information (including verb subcategorization information) is only taken into account later.

In a case like "the coach chided the player tossed a frisbee", the parser would presumably analyze "the player" as the direct object of "chided". It might then reanalyze it as the subject of an embedded clause when "tossed" arrived. This wrong analysis would later be revised when the subcategorization constraints of "chided" were taken into account. But this treatment implies that the subcategorization constraints of "chided" are ignored until much later than other evidence suggests (e.g., McRae et al., 1998; Ni, Fodor, Crain,

[4]Stevenson's model could predict effects like those we observe here either by removing the Right Edge Restriction or by relaxing it to allow a word on the right edge of a tree to play a grammatical role which is inconsistent with its right edge role.

and Shankweiler, 1998; Trueswell and Tanenhaus, 1994). Moreover, in the Experiment 2 stimuli, the account predicts no difference between the Animate and Inanimate Reduced conditions, which have the same lexical category structure.

Syntax vs. Lexicon Autonomy.

Swinney (1979) and Tanenhaus et al. (1979) argue for a certain degree of autonomy of syntax and lexicon. Gibson and Tunstall (1999) argue that the autonomy assumption can handle cases of syntactic ambiguity resolution as well. Applying this idea to the present case, when the previous input is "The coach chided the player" and the current word is "tossed", then the rules of English grammar require that "tossed" be a past participle. Assuming that "tossed" is biased toward its active sense, the resulting conflict slows processing.

One problem with this account is that it predicts that the same slowing should occur when the disambiguating "who was" is present (since the syntactic options for "tossed" are the same in this case). It may be possible to rescue the account by claiming that the syntax is somehow "stronger" when "who was" is present, but some further spelling out of how to measure the relative strength of syntactic and lexical constraints will be needed to make this approach clear.

Semantic Evidence for Local Parse Formation.

A second problem for the Syntax/Lexicon Autonomy account stems from its commitment to the assumption that no local structures are constructed. This assumption predicts that the semantic interpretation of the local subsequence should have no influence on the interpretation of the sentence. By contrast, the self-organization account claims that the parser temporarily entertains the erroneous parse. If we assume that even temporarily entertained structural hypotheses activate semantic interpretations that persist beyond the time during which the sentence is being read, then it may be possible to detect a semantic reflex of the erroneous parsing.

We ran a follow-up self-paced reading study on shortened versions of the the two Animate conditions of Experiment 2 (Unreduced and Reduced) to address this issue. We asked one of two questions after each condition, as shown in (12).

(12) The bandit worried about the prisoner (who was) transported the whole way.
  a. Was the prisoner transported? (prisoner = Theme of transporting)
  b. Did the prisoner transport something? (prisoner = Agent of transporting)

We omitted the material from the Agent phrase onward (e.g., "by the guards...") because we were interested in finding out what people were understanding immediately after reading the segment in which the reading time differences appeared. The filler:stimulus ratio was 3:1. The same procedure was used as in Experiments 1 and 2. Thirty-eight undergraduate native English speakers from the University of Connecticut participated in the study for course credit. We expected an interaction between Reduction and Question type: readers should be more likely to consider the ambiguous noun phrase ("prisoner") to be an Agent and less likely to consider it to be a Theme in the Reduced condition than the Unreduced.

The data from two participants, who made more than 25% incorrect responses on the comprehension questions were removed from the analysis. The results are shown in Table 5.

Table 5: Mean rate of "Yes" responses in the four conditions of the Experiment 2 followup study. "Theme/Agent question" means that the question asked if the ambiguous noun phrase was a Theme/Agent of the subsequent verb, respectively. The numbers in parentheses are standard errors.

	Theme question	Agent question
Unreduced	0.96 (0.013)	0.18 (0.034)
Reduced	0.90 (0.021)	0.27 (0.038)

Subject and item means were subjected to separate analyses of variance (ANOVAs), each with two factors: Reduction and Question Type. There was a main effect of Question Type ($F1(1, 35) = 535.228$, $MSE = 0.0335$, $p < .001$ (***); $F2(1, 19) = 341.093$, $MSE = 0.0292$, $p < .001$ (***)). The predicted interaction between Question Type and Reduction was significant ($F1(1, 35) = 9.820$, $MSE = 0.0188$, $p = .003$ (**); $F2(1, 19) = 6.057$, $MSE = 0.0170$, $p = .024$ (*)), although the effect of Reduction was only significant for the Theme questions in a post hoc test (Fisher's Protected Least Squares, Theme question: $p = .013$ (*); Agent question: $p = .117$). There was no main effect of Reduction. T-tests revealed that the total rate of "Yes" responses (Agent and Theme responses added together) was greater than 100% in both the Reduced and the Unreduced sentences.

Although the effect is not large, the tendency toward a greater ratio of Agent assertions to Theme assertions in the Reduced condition is consistent with the claim of the self-organization account that the erroneous parse becomes temporarily established in the Reduced condition. The Syntax/Lexicon Autonomy account, by not allowing the Agent parse to be formed at all, does not straightforwardly predict this tendency.

The fact that the sum of "Yes" responses for each sentence type is greater than 100% seems puzzling in light of the fact that non-reflexive verbs do not normally allow their arguments to be both Agents and Themes simultaneously.

This result bears a resemblance to the findings of Christianson, Holling s-both questions in our Unreduced condition seems more puzzling at first. Self-organization may offer some insight here as well: structures form, albeit weakly, even when there is stiff competition from a better alternative (e.g., “player tossed” in the active sense, even when “player who was tossed” matches the input). In fact, the self-organizing model of Vosse and Kempen (2000) seems to predict the difficulty of center-embedded structures as a consequence of interference due to this kind of noncontiguous bond formation.

Parsing Breakdown.

Another possible explanation for the effects observed here is that they involve the kind of parsing breakdown that is thought to be associated with certain well-known garden path sentences. Indeed, as an anonymous reviewer pointed out, sentences like (14) give the impression of giving rise to two coherent units, “The horse raced past the barn” and ”the barn fell”, a similar response to that we are claiming occurs with the difficult-to-process stimuli in the current experiments. A standard explanation is that the second coherent unit (”the barn fell”) is perceived because the processor has failed to find a grammatical parse and has resorted to a strategy for salvaging some coherence out of the apparently ill-formed input.

(14) The horse raced past the barn fell.

Perhaps our most difficult conditions involve a similar failure to parse followed by a salvaging strategy.

We understand two versions of this account. Under Version 1, some or all of the stimuli from Experiments 2 and 3 are ungrammatical and for that reason, they fail to be parsed. Under Version 2, the critical cases are not ungrammatical but they are just very hard to parse and parsing fails for this reason.

Version 1 faces two challenges. The first is that Experiment 3 provides empirical evidence against the claim that all of our stimuli are ungrammatical. The Unreduced versions of each sentence were judged grammatical at a particularly high rate and 18 different Unreduced items (9 from Experiment 1 and 9 from Experiment 2) were judged grammatical 100% of the time (9 out of 9 trials). Moreover, three of the Reduced versions (Items 1-Animate, 19-Inanimate, and 20-Inanimate from Experiment 2) were judged grammatical 100% of the time. Given the low rate at which truly ungrammatical constructions were accepted (8%), these results would be highly unlikely if the underlying syntactic constructions were actually ungrammatical. Even on a subject by subject basis, there were no subjects who rated each of the four conditions grammatical less than half the time for the Experiment 2 stimuli and there was only one subject who did so for the Experiment 1 stimuli (5 trials on average per condition per subexperiment). On the other hand, claiming that a relevant subset of our stimuli are ungrammatical requires complicating the theory of grammar in an undesirable way. The simplest grammaticality-based account of the critical interaction would claim that just the hardest condition in each experiment is ungrammatical and all the rest are grammatical. But there is no clear linguistic motivation for treating the Ambiguous (or Animate) Reduced conditions as structurally distinct from the Unambiguous (or Inanimate) conditions. A more plausible grammaticality-based account would draw a line between the Beneficiary extractions of Experiment 1 (ungrammatical) and the Theme extractions of Experiment 2 (grammatical). While such an account could explain the results of Experiment 1 as a consequence of the salvaging operation, it would not distinguish the conditions of Experiment 2. A second challenge to Version 1 of the ungrammaticality approach is that a theory with two mechanisms (one for grammatical processing and one for ungrammatical processing) is more complex, and thus less preferable, than one with only one mechanism. Of course the question remains as to whether an implemented version of the self-organizing model can handle all of the types of cases for which appeal has been made to extra-grammatical strategies, but given the unifying value that the account offers here, we think the question is worth pursuing.

Version 2 claims that all of our stimuli are grammatical, but people fail to discover the grammatical parse in the hard conditions (as they apparently fail in "the horse raced past the barn fell"), and instead salvage a parse of just the last part of the input. This account is quite similar to the self-organizing approach, which is compatible with the notion that parses can sometimes fail to form. For example, if the lexical entry of a verb has a strong bias against a passive interpretation, then the word may need considerable support from the syntax, in the form of function words like "was" and "by", to successfully participate in a passive construction. But Version 2 diverges from the self-organization approach by not allowing the felicity of the salvaged parse to play a role in preventing the fully coherent parse from forming. Instead, an independent cause is required. Perhaps this cause is a strong lexical bias against Beneficiary extraction in Experiment 1, but it does not seem plausible to claim that there is a strong lexical bias against Theme extraction in Experiment 2. In particular, it is noteworthy that the critical interaction in Experiment 2 is already significant at the verb (e.g., "transported"), suggesting that it does not hinge on the employment of the somewhat colloquial NP adverb structure (e.g., "the whole way").

Syntax/Semantics Self-organization.

Rayner et al. (1983) posit a Two Stage account of parsing:

> [U]sing the initial syntactic analysis of an input string to identify (minimal) major phrases, the language processor then calls on real world knowledge and information about the current discourse to compare the pragmatic plausibility of whatever sets of relations between phrases are listed in the thematic structures associated with the head of a phrase. If the thematic selection process turns up a set of relations which is incompatible with the chosen syntactic analysis of the input string, then revision of that analysis will be attempted.

Given the results of Experiment 2, which indicate a role for semantics in causing the difficulty in the Animate/Reduced case, it is natural to ask if the Two Stage account is suited to predict the results at hand. One might assume that interference from the erroneous interpretation of the distracting subsequence in a case like "The bandit worried about the prisoner transported the whole way" occurs not because the syntactic processor computes the local parse, but because the subsequently operating Thematic Selection Module decides that the prisoner transporting something is more plausible than the prisoner being transported and provokes an attempt at reanalysis. Of course, if the sentence is to be successfully interpreted, as the high rate of Theme interpretations in the Experiment 2 followup study indicates, the syntactic module will have to reject the Thematic Module's request, and insist on returning to the original interpretation. Typically, such Two Stage accounts assume that the resolution stage involves the simultaneous use of constraints coming from both modules.

This approach could be called"Syntax/Semantics self-organization" because the two modules initially make independent contributions to the parsing process and it is through their interaction that an interpretation is decided upon. In this regard, it is similar to the Homogeneous Self-Organization account. Nevertheless, we prefer the Homogeneous Self-Organization account because we believe it is simpler. In particular, under the Syntax/Semantics approach, some mechanism is needed for deciding which material should be sent to the Thematic Module and when to send it, for specifying the range of assignments of arguments to roles within the Thematic Module, and for determining how to revise the syntax in a way that will successfully meet the requirements laid down by the Thematic Module. These mechanisms are effectively performing the function of reviewing the possible structures to decide which are simultaneously the most syntactically and semantically felicitous. Under the Homogeneous Self-Organization account, semantic and syntactic information simultaneously constrain the formation of structures from the beginning.[5]

Statistical parsing models.

Those statistical parsing models which assign probabilities to the parses generated by a standard parser (e.g. Jurafsky, 1996; Hale, 2002) will fail to detect locally coherent sequences because the parser fails to detect them. On the other hand, workable statistical approaches often use n-gram approximations (Charniak, 1993; Corley and Crocker, 2000): all contiguous subsequences of length n words are collected from a large corpus. These are used to construct a statistical model of properties that any word is likely to have based on properties of the preceding n-1 words. Such models are well-positioned to pick up on local coherence effects. On the other hand, there is not a linguistically motivated way of integrating n-grams with phrase-structures. It may be possible to establish a relationship between the n-gram approach and self-organization via Variable-Length Markov Modeling (e.g. Guyon and Pereira, 1995), a way of tailoring the size of n to the current context, although questions about reconciling finite- and infinite-state computation remain (see Tabor, 2002).

Related Work: Reanalysis as a Last Resort

Following Fodor and Frazier (1980)'s proposal that the parser prefers to avoid reanalysis whenever other options or possible, Schneider and Phillips (2001) and Sturt, Pickering, Scheepers, and Crocker (2001) set up experiments in which they pitted easy reanalyses against other constraints. Their conclusion was that the parser generally prefers to keep structure that it has already built. It is worth noting that the self-organization approach

[5]On the other hand, when semantic and syntactic information are used together, there is no architectural explanation for differences between semantic and thematic processing which have been observed in other studies (Braze, Shankweiler, Ni, and Palumbo, 2000; McElree and Griffith, 1995, 1998; Ni et al., 1998). We believe that these differences follow from the differences in the structure of syntactic and semantic information. For example, the compatibility of the meaning of "player" with the properties of the subject of "tossed[Active]" is a property of these and a few related nouns and verbs, but the felicity of NP as the subject of VP is a very pervasive property of sentences in the language. As a result, syntactic contrasts have a stark representational reflex while semantic contrasts are more subtle (see Spivey, Fitneva, Tabor, and Ajmani, 2002; Tabor and Hutchins, 2000).

makes a similar, general prediction: since parses are preferred to the extent to which they are supported by the evidence of the input, and since an initial analysis gets a head start on becoming activated compared to a competing analysis that is first signalled by late-arriving disambiguating information, the initial analysis has an advantage (see also Tabor and Hutchins, 2003b). Nevertheless, we are claiming that the current experiments involve cases in which the later-arriving information does have the capacity to at least strongly interfere with the initial analysis. In fact, Schneider and Phillips (2001)'s and (Sturt et al., 2001)'s stimuli (illustrated in (15) and (16) respectively) afford local analyses like ours, and in their cases, these local analyses are even consistent with a grammatical parse of the entire input. In both cases, the researchers found that the "himself" version was dispreferred, indicating that readers attached the ambiguous verb phrases high rather than low. One may thus ask why their readers did not build the local parse of "the funny man wrote…" or "the enemy spy had shot", which would have amounted to a preference for low attachment.

(15) The woman who knows the funny man wrote some comedy sketches himself/herself. (from Schneider and Phillips, 2001).

(16) The troops who found the enemy spy had shot himself/themselves… (from Sturt et al., 2001).

One difference between the experiments is that, in our cases, the grammatical parses involve using the ambiguous verbs in dispreferred senses. As we noted above, Experiment 3 indicated that fronting of the beneficiary argument of ditransitive verbs (vis a vis Experiment 1) is not a reading that readers easily get. Likewise, the results of Trueswell et al. (1994), also mentioned above, indicate that the reduced relative reading is dispreferred with animate subjects in the Experiment 2 stimuli. By contrast, both the high and low attachment interpretations of Schneider and Phillip's study and of Sturt et al.'s study involved the same preferred, active interpretations of their verbs. Another difference is that in our stimuli, the subject of the troublesome verb is an empty element whose existence has to be inferred at the point of reading the verb, while in the Schneider and Phillips and Sturt et al. stimuli, the subject of the critical verb has already been perceived, and indeed, has generated an unsatisfied dependency in the sense of Gibson and Thomas (1996) and Gibson (1998) which the critical verb can satisfy. These differences may have caused readers to prefer building the local structures in our experiments and to disprefer doing so in Phillips and Gibson's and Sturt et al.'s.

Prospect

Without the structural framework provided by Generative Linguistic Theory, the experimental study of sentence processing might never have gotten off the ground. But the aim of Generative Linguistic Theory is to characterize the meanings and structural relationships between the whole, grammatical sentences of a language. It is not designed to model the on-line construction of mental representations. Thus, all explicit models of sentence processing based on Generative Theory have needed to augment the linguistic models with some account of how the structures develop over time. Because the Generative Structures lend themselves rather well to incremental construction of whole structures by stepwise augmentation of partial structures, it is natural to assume that parts of sentences are encoded as partial structures, just as whole sentences are assumed to be encoded as whole structures. The work just reported, however, suggests a different picture. Partial tree structures may be only a convenient approximation. If the model of Vosse and Kempen (2000) is correct, then only a very small subset of states of the mental language processor can be viewed as canonical tree structures. Most of the time, the system is in an unstable state which can be characterized as a position and velocity relative to the rare, stable states that correspond to the tree structures. On this view, the fact that many sentence-processing behaviors *can* be well-modeled by an incremental structure-builder is an artifact of the serial unfolding of language over time—the time that passes after each new word is encountered often allows the system to produce a globally optimal integration of new information with old. It is only in cases like those we have designed here that the self-organizing nature of the process becomes readily apparent; the time between successive word encounters is not sufficient to squash down the atypically vigorous local interpretations embedded in our stimuli.

What are the advantages of understanding sentence processing as a self-organizing process, in addition to providing an explanation for the empirical cases at hand? Self-organizing accounts generally provide simpler explanations than their central-executive counterparts. In a central-executive grammar approach to parsing the researcher specifies many decisions the parser makes about how parsing should be carried out (e.g. what kind of information is attended to first, how far the parser should project structure into the future, how new analyses should be identified when old ones fail, etc.). In a self-organizing account, the goal is to let these choices stem from simpler assumptions at a smaller structural scale: for example, reanalysis is claimed to occur in situations where new information tips the balance against the previous dominant contender, and the scope of the search for alternatives falls out of the energy minimization process which seeks stability under the revised constraints. A second advantage of the self-organizing approach is the simplicity of its mapping to response times: the relaxation of the feedback system is a direct analog of human processing. The approach may also provide a computationally well-grounded way of modeling reanalysis by making a uniform assumption about how tree-restructuring operations consume time (see Stevenson, 1998). The question remains open as to how much work the self-organization assumption will do at explaining the broad range of established sentence-processing phenomena. Vosse and Kempen (2000)'s work provides an encouraging indication that a self-organizing mechanism has some insights to contribute into the set of sentence processing phenomena that are usually thought to be connected with memory-load and not to involve ambiguity. Interestingly, their model seems to succeed at predicting the great difficulties associated with multiple center-embeddings via interference from grammatically inconsistent parses. This account, in effect, treats the parsing of center-embedded structures as a kind of ambiguity resolution problem, an appealing unification if it turns out to be sustainable. On the other hand, Traxler et al. (1998) and Gompel et al. (2001) provide data indicating that multiple attachment possibilities may actually speed processing rather than slow it down in the case of adjunct attachment. These results challenge those constraint-based accounts which claim that the effect of competition is always to slow processing. An insight that self-organization may offer here is a focus on stability, rather than competition. If, by forming equally good bonds at multiple sites simultaneously, an adjunct is more stable than when there is an asymmetry between the sites, then processing will be helped rather than hindered.

In sum, we suggest that the self-organization hypothesis is worth pursuing both because of the simple account it provides of the present empirical data and because it shows promise of providing insight into a wide range of sentence processing phenomena.

References

Altmann, G., and Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238.

Bever, T. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York: Wiley.

Braze, D., Shankweiler, D., Ni, W., and Palumbo, L. C. (2000). *Anomaly detection and recovery in reading: Eye movements.* (Manuscript, Department of Psychology, University of Connecticut)

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

Christianson, K., Hollingworth, A., Halliwell, J. F., and Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368–407.

Cohen, J., MacWhinney, B., Flatt, M., and Provost, J. (1993). Psyscope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*(2), 257–271.

Corley, S., and Crocker, M. W. (2000). The modular statistical hypothesis: exploring lexical category ambiguity. In M. W. Crocker, M. Pickering, and J. Charles Clifton (Eds.), *Architectures and mechanisms for language processing* (pp. 135–160). Cambridge: Cambridge University Press.

Crain, S., and Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. R. Dowty, L. Karttunnen, and A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives.* Cambridge: Cambridge University Press.

Ferreira, F., Christianson, K., and Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research*, *30*(1), 3–20.

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

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

Ferreira, F., and Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, *30*(6), 725–745.

Ferreira, F., and Henderson, J. M. (1993). Reading processes during syntactic analysis and reanalysis. *Canadian Journal of Experimental Psychology*, *47*(2), 247–275.

Fodor, J. D., and Frazier, L. (1980). Is the human sentence parsing mechanism an atn? *Cognition*, *8*, 417–459.

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

Frazier, L. (1978). *On comprehending sentences: syntactic parsing strategies.* (Unpublished doctoral dissertation, University of Connecticut)

Frazier, L. (1988). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance* (p. 559-586). Hillsdale, NJ: Lawrence Earlbaum Associates.

Frazier, L. (1990). Exploring the architecture of the language-processing system. In *Cognitive models of speech processing: Psycholinguistic and computational perspectives.* Cambridge, Massachusetts: MIT Press.

Frazier, L., and Clifton, C. (1996). *Construal.* Cambridge, Massachusetts: MIT Press.

Frazier, L., and Clifton, C. (1998). Sentence reanalysis and visibility. In J. D. Fodor and F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 143–176). Dordrecht: Kluwer.

Frazier, L., and Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*(6), 291-325.

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

Gibson, E. (1991). *A computational theory of human linguistic processing. memory limitions and processing breakdown.* (Ph.D. Dissertation, Carnegie Mellon University)

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

Gibson, E., and Thomas, J. (1996). The processing complexity of english center-embedded and self-embedded structures. In C. Schütze (Ed.), *Proceedings of the NELS 26 workshop on language processing.* Cambridge, MA: MIT occasional papers in linguistics 9.

Gibson, E., and Tunstall, S. (1999). *The parser need not keep track of contingent frequencies: a lexical account of the english determiner/complementizer ambiguity.* (Poster session presented at the 12th Annual CUNY Sentence Processing Conference, New York, New York)

Gompel, R. van, Pickering, M., and Traxler, M. (2001). Reanalysis in sentence processing: evidence against current contraint-based and two-stage models. *Journal of Memory and Language*, *45*, 225–258.

Gompel, R. P. G. van, Pickering, M. J., and Traxler, M. J. (2000). Unrestricted race: A new model of syntactic ambiguity resolution. In A. Kennedy, R. Radach, D. Heller, and J. Pynte (Eds.), *Reading as a perceptual process* (pp. 621–648). Oxford,: Elsevier.

Gorrell, P. (1987). *Studies in human syntactic processing: Ranked-parallel versus serial models.* (Unpublished doctoral dissertation, University of Connecticut)

Gorrell, P. (1995). *Syntax and parsing.* Cambridge: Cambridge University Press.

Grodner, D., Gibson, E., and Tunstall, S. (2002). Syntactic complexity in ambiguity resolution. *Journal of Memory and Language*, *46*, 267–295.

Guyon, I., and Pereira, F. (1995). Design of a linguistic postprocessor using variable memory length markov models. In *International conference on document analysis and recognition.* Montreal: IEEE Computer Society Press. (Available at www.clopinet.com/isabelle/Papers/index.html)

Hale, J. (2002). *The information conveyed by words in sentences.* (Paper presented at the 15th Annual CUNY Conference on Human Sentence Processing, New York, New York)

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

Just, M. A., Carpenter, P. A., and Wooley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*, 228–238.

Kempen, G., and Vosse, T. (1989). Incremental syntactic tree formation in human sentence processing: a cognitive architecture based on activation decay and simulated annealing. *Connection Science*, *1*, 273–290.

Kim, A. E. (2000). *The grammatical aspects of word recognition.* (Ph.D. Dissertation, Department of Psychology, University of Pennsylvania)

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*(2), 15-47.

MacDonald, M. A., Pearlmutter, N. J., and Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.

McElree, B., and Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Language, Memory, and Cognition*, *21*(1), 134–157.

McElree, B., and Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence comprehension: A time-course analysis. *Journal of Experimental Psychology: Language, Memory, and Cognition*, *24*(2), 432–460.

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

Ni, W., Fodor, J. D., Crain, S., and Shankweiler, D. (1998). Anomaly detection: eye movement patterns. *Journal of Psycholinguistic Research*, *5*, 515–539.

Phillips, C. (1998). *Linear order and constituency.* (Manuscript, University of Maryland; revised version to appear in Linguistic Inquiry)

Rayner, K., Carlson, M., and Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*, 358–374.

Schneider, D., and Phillips, C. (2001). Grammatical search and reanalysis. *Journal of Memory and Language*, *45*, 308–336.

Sedivy, J. C. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*, 109–147.

Spivey, M., Fitneva, S., Tabor, W., and Ajmani, S. (2002). The time course of information integration in sentence processing. In S. Stevenson and P. Merlo (Eds.), *Sentence processing and the lexicon: Formal, computational, and experimental perspectives.* (pp. 207–232). Amsterdam: John Benjamins.

Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research*, *23*(4), 295–322.

Stevenson, S. (1997). An integrated symbolic/connectionist parsing archicture. In R. Sun and F. Alexandre (Eds.), *Connectionist-symbolic integration: From unified to hybrid approaches* (pp. 209–206). Mahwah, NJ: Lawrence Erlbaum Assoc.

Stevenson, S. (1998). Parsing as incremental restructuring. In J. D. Fodor and F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 327–363). Dordrecht: Kluwer Academic Publishers.

Stevenson, S., and Merlo, P. (1997). Lexical structure and parsing complexity. *Language and Cognitive Processes*, *12*(2/3), 349–399.

Sturt, P., Pickering, M. J., Scheepers, C., and Crocker, M. W. (2001). The preservation of sturcture in language comprehension: Is reanalysis the last resort? *Journal of Memory and Language*, *45*, 283–307.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*, 645–660.

Tabor, W. (2002). The value of symbolic computation. *Ecological Psychology*, *14*(1/2), 21–52.

Tabor, W., and Hutchins, S. (2000). Mapping the syntax/semantics coastline. In *Proceedings of the 22nd annual meeting of the cognitive science society.* Mahwah, NJ: Lawrence Erlbaum Associates.

Tabor, W., and Hutchins, S. (2003a). *Evidence for self-organized parsing: Digging in effects.* (Paper presented at the 16th Annual CUNY Sentence Processing Conference, Cambridge, Massachusetts)

Tabor, W., and Hutchins, S. (2003b). *Evidence for self-organized sentence processing: Digging in effects.* (Submitted manuscript, University of Connecticut, Department of Psychology: See http://www.sp.uconn.edu/ ps300vc/papers.html)

Tabor, W., Juliano, C., and Tanenhaus, M. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes*, *12*(2/3), 211–271.

Tanenhaus, M., Carlson, G., and Trueswell, J. (1989). The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes*, *4*(3–4), SI211–SI234.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., and Sedivy, J. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science*, *268*, 1632-1634.

Tanenhaus, M. K., Leiman, J. M., and Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, *18*, 427–440.

Tanenhaus, M. K., Spivey-Knowlton, M. J., and Hanna, J. E. (2000). Modeling thematic and discourse context effects with a multiple constraints approach: Implications for the architecture of the language comprehension system. In M. W. Crocker, M. Pickering, and J. Charles Clifton (Eds.), *Architectures and mechanisms for language processing* (pp. 90–118). Cambridge: Cambridge University Press.

Tanenhaus, M. K., and Trueswell, J. C. (1995). Sentence comprehension. In J. Miller and P. Eimas (Eds.), *Handbook of perception and cognition: Volume 11* (p. 217-262). San Diego: Academic Press.

Traxler, M. J., Pickering, M. J., and C. Clifton, J. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, *39*(558–592).

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

Trueswell, J. C., and Tanenhaus, M. K. (1994). Toward a constraint-based lexicalist approach to syntactic ambiguity resolution. In . K. R. C. Clifton, L. Frazier (Ed.), *Perspectives on sentence processing.* Hillsdale, NJ: Erlbaum.

Trueswell, J. C., Tanenhaus, M. K., and 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.

Vosse, T., and Kempen, G. (2000). Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, *75*, 105–143.

Appendix 1. Stimuli for Experiments 1 and 3

The vertical slash (|) indicates the location of the end of the sentence in Experiment 3.

1. The kindergartners liked the little girl (who was) brought/chosen a toy by her parents | on the first day of Chanukah.
2. We saw a movie about an artist (who was) painted/drawn a picture by her father | while he was on his deathbed.
3. The health officials pounced on a restaurant (which was) sent/flown a shipment of salmon by a company | that had failed to comply with refrigeration laws.
4. At the dinner party, I met a man (who was) allowed/forbidden the pleasure of eating sweets by his doctor. |
5. One expects a man (who is) told/forgiven his sins by his own god to have tolerance for weaknesses in others.
(Experiment 3: One should respect a man told/forgiven his sins by his own god.)
6. An elderly gentleman addressed the woman (who was) offered/given a beer by the hostess. |
7. Balthazar praised the professor (who was) taught/given Swahili lessons by a graduate student. |
8. The manager watched a waiter (who was) served/given pea soup by a trainee. |
9. James entertained the children (who were) dyed/hidden Easter eggs by their teachers. |
10. The foreman yelled at a carpenter (who was) cut/sawn a board by his buddy. |
11. The preschool teacher congratulated the little boy (who was) knitted/sewn a hat by his grandmother. |
12. The janitor chatted with the young man (who was) rented/shown an apartment by his uncle. |
13. The nurse admonished a student (who was) nabbed/stolen a muffin by her friends | from the dining hall.
14. The play centered around an innkeeper (who was) recited/sung a verse by a travelling monk. |
15. The hotel owner questioned a guest (who was) brought/taken a drink by the bellboy. |
16. The coach smiled at the player (who was) tossed/thrown a frisbee by the opposing team. |
17. The anthropologist interviewed a woman (who was) knitted/woven a shawl by her mother. |
18. The FBI questioned a congressman (who was) mailed/written a letter by the activist. |
19. The deliveryman teased the accountant (who was) saved/given a coupon by her boss. |
20. The prophet spoke of a man (who was) planted/grown a tree by his daughter. |

Appendix 2. Stimuli for Experiments 2 and 3

1. The man reminded us of the defendant/evidence (who/that was) examined the previous day by the lawyer | from the other company.
2. The bandit talked remorsefully of the prisoner/gold (who/that was) transported the whole way by the guards | with horse whips and curry powder.

3. The parents admired the teacher/textbook (who/that was) loved the very best by the class | which came from the ghetto.

4. The foreman worried about the workers/bricks (who/that were) lifted the whole way by the crane | with the faltering engine.

5. The dean asked about the student/paper (who/that was) graded the very lowest by the professor | with a history of discriminating on the basis of nationality.

6. The master of ceremonies discussed the contestant/recipe (who/that was) selected the third time by the judges | who favored originality.

7. The doctor spoke highly of the specialist/equipment (who/that was) requested the second time by the hospital | with the new scanning machine.

8. The police took pictures of the thief/jewelry (who/that was) identified the first night by the victim | from the highrise when a cricket was chirping so loudly in the stairwell.

9. The general thought about the soldier/valley (who/that was) captured the previous day by the troops | with pistols and bayonets.

10. The governor sent reinforcements to the troops/power plant (who were/that was) attacked the previous week by the terrorists | from the higher elevations.

11. The curator criticized the artist/painting (who/that was) studied the previous term by the students | from the university.

12. The governess identified the boy/necklace (who/that was) described the first time by the lady | with the turquoise earrings.

13. Joseph forgot about the mailman/package (who/that was) expected the next day by the secretary | in charge of public relations.

14. The children were worried about the woman/sofa (who/that was) scratched the first day by the cat | with the tawny fur.

15. The agent photographed the man/van (who/that was) recognized the previous day by the spy | from the Russian intelligence bureau.

16. The manager was concerned about the client/account (who/that was) observed the previous month by the agents | from the government.

17. The activist admired the speaker/solution (who/that was) proposed the first time by the group | from the IGC.

18. The congresswoman was referring to the lawyer/package (who/that was) sent the first time by the governor | of Mississippi.

19. The board was concerned about the students/awards (who/that were) accepted the first year by the school | for people with Downs Syndrome.

20. The poet was looking for the woman/portrait (who/that was) sketched the most endearingly by the artist | with the wooden shoes.

Appendix 3

Experiment 1: Means of raw reading times by position (ms).

	Ambiguous		Unambiguous	
Position	Reduced	Unreduced	Reduced	Unreduced
-4	422	438	405	475
-3	462	491	447	476
-2	402	378	390	419
-1	397	440	427	466
0	542	397	475	468
1	632	451	520	471
2	481	455	514	549
3	392	370	393	441
4	367	377	385	397

Experiment 2: Means of raw reading times by position (ms).

	Animate		Inanimate	
Position	Reduced	Unreduced	Reduced	Unreduced
-4	416	426	395	430
-3	434	425	417	428
-2	384	399	357	365
-1	421	405	390	407
0	506	400	449	422
1	553	457	503	443
2	516	480	492	480
3	467	397	433	414
4	414	375	397	396