Journal of Experimental Psychology:
Learning, Memory, and Cognition
2004, Vol. 30, No. 2, 431–450

0278-7393/04/$12.00 DOI: 10.1037/0278-7393.30.2.431

Evidence for Self-Organized Sentence Processing: Digging-In Effects

Whitney Tabor and Sean Hutchins
University of Connecticut

Dynamical, self-organizing models of sentence processing predict "digging-in" effects: The more committed the parser becomes to a wrong syntactic choice, the harder it is to reanalyze. Experiment 1 replicates previous grammaticality judgment studies (F. Ferreira & J. M. Henderson, 1991b, 1993), revealing a deleterious effect of lengthening the ambiguous region of a garden-path sentence. The authors interpret this result as a digging-in effect. Experiment 2 finds a corresponding effect on reading times. Experiment 3 finds that making 2 wrong attachments is worse than making 1. Non-self-organizing models require multiple stipulations to predict both kinds of effects. The authors show that, under an appropriately formulated self-organizing account, both results stem from self-reinforcement of node and link activations, a feature that is needed independently. An implemented model is given.

A widespread view of sentence processing holds that comprehenders of natural-language sentences sometimes commit (or at least partially commit) to an analysis that later-arriving information reveals to be wrong. Long processing times associated with the arrival of the disambiguating information are assumed to stem from the extra time the processor uses to revise its parse or reanalyze. For many years, most of the work on such phenomena focused on the factors that lead the parser to make the wrong choice initially and not on the reanalysis process itself. More recently, several accounts of reanalysis have been proposed. One of the ways these accounts distinguish themselves is by predicting differences in the difficulty of reanalysis across different constructions.

According to one widely held view, which we refer to as the top-down mechanism selection view (TDMS), there is a mechanism for building parses, and there is at least one other mechanism for repairing them when they are discovered to be incorrect. Differences in the degree of difficulty of reanalysis arise because there are some errors that the repair mechanism cannot fix or that require different kinds of repair mechanisms (e.g., Fodor & Inoue, 1994, 1998, 2001; Frazier & Clifton, 1996, 1998; Gorrell, 1995; Inoue & Fodor, 1995; Sturt & Crocker, 1996, 1997, 1998; Sturt, Pickering, & Crocker, 1999).

A contrasting view holds that the parser is a (continuous-time) dynamical self-organizing (DSO) system (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Stevenson, 1994, 1997, 1998; Stevenson & Merlo, 1997; Vosse & Kempen, 2000). DSO models claim that attachments have continuous-valued strengths and that these increase or decrease as a function of how well the evidence of the perceived speech sustains them. A particular attachment will grow stronger if it outcompetes its competitors and will grow weaker if it is outcompeted or decays for lack of support. As Stevenson (1998) has observed, there is only one mechanism in a DSO system for both analysis and reanalysis (see also Grodner, Gibson, Argaman, & Babyonyshev, 2003), and this distinguishes DSO systems from TDMS systems. The DSO mechanism is self-organizing in that each event of perceiving a word independently gives rise to a set of attachment sites that interact with the attachment sites generated by the perceptions of other words. As a result of the interactions among the attachment sites, it is sometimes appropriate to say the system is "constructing a parse"; at other times, it is appropriate to say that it is "dismantling (and simultaneously reconstructing) a parse." However, there is no overseeing mechanism as there is in TDMS accounts, which decides which activity the parser should engage in at a particular time. The use of continuous-valued attachment strengths is critical in a DSO system (and not particularly relevant in a TDMS system) because it allows the system to relax into a state in which the constraints are satisfied in an optimal way. In this sense, the continuous-valued activations are a hallmark trait of DSO models. An empirical consequence of this feature of the models is the existence of what we call "digging-in effects": The more stable a parse becomes, the harder it is to undo.

In fact, some dynamical constraint-satisfaction models predict two kinds of digging-in effects. First, single-attachment digging-in effects involve an increase in the strength of a single erroneous bond as time passes. This account is one explanation for what have often been called "length effects" (e.g., Ferreira & Henderson, 1991b; Frazier & Clifton, 1998; Warner & Glass, 1987): Reanalysis is more difficult after a long ambiguous region than after a short one. Second, multiple-attachment digging-in effects involve a succession of wrong attachments. The prediction is that, other factors being equal, undoing multiple wrong attachments is more difficult than undoing fewer.

Whitney Tabor and Sean Hutchins, Department of Psychology, University of Connecticut.

Thanks to Fernanda Ferreira for providing many insightful comments as well as materials that formed the basis of the stimuli, Leonard Katz, Maryellen MacDonald, and Michael K. Tanenhaus for insightful comments, and Naomi Hyun, David Perkowski, Daniel Richardson, and Aaron Schultz for technical assistance, and Garvin Boudle and Edward D'Agata for supporting the computers. This work was supported in part by University of Connecticut Grant # HD 40353, Research Foundation Grant # FRS 444078, and by the National Institute of Child Health and Human Development.

Correspondence concerning this article should be addressed to Whitney Tabor, Department of Psychology, U-20, University of Connecticut, Storrs, Connecticut 06269. E-mail: tabor@uconnvm.uconn.edu

Relatively little work has been done to test the differences between TDMS and DSO approaches to reanalysis. The main goal of this article is to fill part of that empirical gap by providing some evidence in support of the claim that online digging-in effects exist. Before describing the experiments, we review prior work on related phenomena and outline an implemented DSO model.

Prior Work

Offline Studies

Ferreira and Henderson (1991b, pp. 729), in their follow-up of work by Warner and Glass (1987), found that when people were asked to read sentences like Sentences 1a and 1b in rapid serial visual presentation (RSVP) and make grammaticality judgments about them, they tended to judge Sentence 1a versions grammatical less often than Sentence 1b versions.[1] One of the accounts that Ferreira and Henderson suggested for this contrast is that readers erroneously parse “the town” as the direct object of “invaded” in both examples (i.e., they are “garden pathed”), but they become more committed to this wrong analysis in Sentence 1b because of the relatively late arrival of disambiguating information (see also Ferreira & Henderson, 1991a).

After the Martians invaded the
town that the city bordered was evacuated. [18%] (1a)

After the Martians invaded the town
was evacuated. [69%] (1b)

Ferreira and Henderson (1991b) argued that the contrast was not due simply to the larger number of words in Sentence 1a than 1b, because a similar length contrast in a pair of sentences that did not encourage misanalysis of the matrix subject (Sentence 2a vs. 2b, p. 729) produced a smaller difference in rate of positive grammaticality judgments.

After the Martians invaded the town that the
city bordered the people were evacuated. [64%] (2a)

After the Martians invaded
the town the people were evacuated. [82%] (2b)

Moreover, Ferreira and Henderson (1991b, 1993, 1995) showed that the effect obtained when the words were presented in cumulative, segment-by-segment self-paced reading, and with syntactically different kinds of lengtheners (e.g., “the town that is small,” “the town with narrow streets”), but not when the head of the manipulated noun phrase (NP) was adjacent to the main clause verb (e.g., “the small and friendly town”). Ferreira and Henderson (1998) attribute this head-position effect to the fact that the nonadjacent head condition involves an additional embedded clause, which puts an extra load on processing. However, Sturt et al. (1999) failed to find such a head-position effect in a self-paced reading study with similar materials.

In a related study, Bailey and Ferreira (2003, p. 187) found that spoken versions of sentences like Sentence 3b, which contained a disfluency between an ambiguous nominal head and its subsequent governing verb, elicited a lower rate of positive grammaticality judgments than corresponding cases in which the disfluency came before the ambiguous noun (Sentence 3a).

Sandra bumped into the busboy and the
uh uh waiter told her to be careful. [93%] (3a)

Sandra bumped into the busboy and the
waiter uh uh told her to be careful. [78%] (3b)

As Bailey and Ferreira (2003) have noted, these results are explained by the hypothesis that incorrect assignments of nominal heads to thematic roles grow stronger during the pronunciation of a disfluency. Another explanation for the contrast in Sentence 3 is that the disfluency is interpreted as a cue for a clause boundary. Bailey and Ferreira found indications that both causes may be involved: People had more trouble recognizing grammaticality when disfluencies were unhelpfully placed (in support of the cue hypothesis), but they also showed a similar contrast to that of Sentence 3 when the disfluencies were replaced by random noises from the environment (e.g., doors slamming, dogs barking). The latter finding may stem from a digging-in effect. It might also be that listeners ignore the environmental noises and perceive pauses in their places, interpreting these as cues. Thus, more research is needed to establish with certainty whether there are effects of pure time, as Bailey and Ferreira and the DSO approach claim.

Online Studies

Because the DSO approach assigns such a central role to time, its most obvious predictions are those about online-processing times. Several empirical studies have probed for online length effects, but the results have been mixed. Frazier and Rayner (1982, p. 208) used an eye-tracker to study both early- and late-closure ambiguities like Sentence 1 and minimal attachment ambiguities like Sentence 4.[2]

Tom heard the gossip (about the new neighbors)
wasn’t true. [nonminimal]

I wonder if Tom heard the gossip
(about the new neighbors). [minimal] (4)

Indeed, in the early-closure and nonminimally attached conditions, the sentences with long ambiguous regions had longer reading times and more regressions, but no such contrast was observed in the corresponding late-closure and minimally attached controls. However, as Frazier and Clifton (1998) have noted, the stimuli in the 1982 study were not optimally designed (e.g., quite a few different syntactic forms were used, the head position in the ambiguous NP of the closure ambiguities varied across items, and the disambiguating information consisted only of the sentence-

[1] The percentages judged grammatical are shown in parentheses after each sentence.

[2] In the short conditions the parenthetical material in Sentence 4 was absent, whereas in the long conditions it was included without the parentheses.

final period in the short conditions of the minimal attachment study). Ferreira and Henderson (1993) studied eye movements in a more systematically controlled set of closure ambiguities, but they found an opposite effect: Total reading times on the first word of the disambiguating region (the matrix verb in Sentence example 1) were longer with short ambiguous regions than with long regions. Although the numerical interaction between closure type and ambiguous region length went (nonsignificantly) in the direction predicted by our DSO model for their total reading times, the comparison between the early- and late-closure conditions (Sentence 1 vs. Sentence 2) is suspect because the subject noun of the main clause changed across closure type. Thus, for example, the difference between the time spent on "was" in Sentence 1 and the time spent on "were" in Sentence 2 may stem from properties of the preceding NPs ("the town" vs. "the people") or from a lexical difference between "was" and "were." The comparisons with unambiguous control sentences are critical for discerning digging-in effects in an online study because there is an independently motivated explanation for why long ambiguous regions produce longer reading times: The long ambiguous region requires the processor to perform a long-distance integration of dependent constituents (Gibson, 1998, 2000; e.g., "was" needs to be integrated with its subject NP, "the town," precisely at the point at which digging-in effects are expected to appear). Thus, it is possible that a more systematic control condition would produce different results.

SOPARSE: A DSO Model

Our DSO model is called SOPARSE. (A formal synopsis is provided in Appendix A.) It consists of a collection of lexically anchored tree fragments, which are activated when the words corresponding to their lexical anchors are perceived. The root and terminal nodes of these fragments (here called head and foot nodes, respectively) consist of feature vectors, which specify the combination preferences of the fragments. Both syntactic and semantic features are represented. For example, one fragment associated with the word "rode" has a head node whose features specify that it needs to attach to a site that licenses an inflectional phrase (IP); it has one foot node that specifies that it needs to combine with a preceding (subject) NP describing a rider and a second foot node that specifies that it can combine with a subsequent (object) NP describing a rideable object. When a word is perceived, its tree fragment becomes activated, and its peripheral nodes attempt to form links with the peripheral nodes of other activated fragments. A link between two nodes grows in strength in proportion to its current strength (link self-feedback), the activations of its supporting fragments (fragment ↔ link feedback), and the feature match between the foot and head nodes that it connects. A fragment's activation, in turn, grows with the strengths of its links. Reading time is modeled as fragment stabilization time. The two feedback mechanisms implement a "rich get richer" principle. This principle is needed to ensure that good parses outcompete poor parses. The link self-feedback has the added consequence of predicting ambiguous region length effects (tested in Experiments 1 and 2): As time passes after an erroneous link is formed, the erroneous link becomes stronger and hence harder to undo. Because of the links' dependence on fragment-link feedback, a tree fragment whose possible attachment sites have been taken up by the fragments associated with preceding words will not easily succeed in forming links. This leads to multiple-attachment digging-in predictions, the topic of Experiment 3.

SOPARSE is similar in concept to the frameworks of Stevenson (1994, 1997, 1998; Stevenson & Merlo, 1997) and of Vosse and Kempen (2000). We use the term *DSO models* to refer to the class of models exemplified by Stevenson's work, Vosse and Kempen's work, and SOPARSE. We use the term *online DSO models* to refer to models like Stevenson's and ours, which predict online reaction time data.

SOPARSE is closely related to lexical dependency grammar (e.g., Sleator & Temperley, 1993) and (lexicalized) tree adjoining grammar (LTAG) (e.g., Joshi & Schabes, 1996; Joshi & Srinivas, 1994), which are grammatical formalisms. We focus here on the similarity to LTAG in recognition of the prior work on sentence processing in this framework (Kim, 2000; Kim, Srinivas, & Trueswell, 2002). The elements of LTAG are lexically anchored fragments of phrase structure trees (called LTAGs), which specify their grammatical (and sometimes semantic) combination possibilities. As in SOPARSE, the peripheral nodes of these fragments can bond together if they have compatible features; thus, sentence parses are built up out of the fragments. Prominent differences between the frameworks are (a) LTAG establishes attachments instantaneously, whereas SOPARSE establishes them gradually; (b) LTAG forms legal structures (if possible), whereas SOPARSE forms structures that are locally optimal with respect to its dynamics; (c) LTAG enforces a noncrossing branch constraint explicitly, whereas SOPARSE merely tends to form noncrossing branches because the most available attachments usually involve adjacent constituents; (d) thus far, the assignment of LTAGs to words has been accomplished by statistical optimization with respect to an *n*-gram model (e.g., Joshi & Srinivas, 1994) or by a neural network trained on LTAG prediction (e.g., Kim, 2000); SOPARSE assigns tree fragments to words by interactive activation. Nevertheless, the fundamental units of LTAG are very similar to those of SOPARSE, and the worked-out analyses of LTAG may be useful for scaling SOPARSE up to wider coverage.

SOPARSE also has much in common with dynamical models that integrate a wide variety of constraints in a fixed-size connectionist architecture but do not make use of linguistic structural entities (e.g., McRae, Spivey-Knowlton, & Tanenhaus, 1998; Tanenhaus, Spivey-Knowlton, & Hanna, 2000). We focus here on the more restrictive, tree-building models to be able to study the consequences of letting the linguistic structures interact in a dynamical setting. We also refrain from asking how the linguistic structures themselves might arise in a self-organized fashion by learning, noting that there are some indications that they may be able to do so (e.g., Elman, 1991; Tabor, Juliano, & Tanenhaus, 1997; Rohde, 2002).

Overview

Experiment 1 replicates Ferreira and Henderson's (1991b) grammaticality judgment results using noncumulative self-paced reading and a uniform matrix subject across all conditions. Experiment 2 studies the same stimuli in an online environment in which the response measure is time, the fundamental dependent variable about which DSO models make predictions. Experiment 3 investigates multiple-attachment digging-in effects by comparing a case in which readers make one wrong attachment with one in which they make two.

A number of plausible accounts of reanalysis have been proposed. Many of them predict at least some of the results we report here. To discuss them clearly and concisely, we refrain from assessing their predictions after describing each experiment and review them in series, along with our own model, in the General Discussion. Our overall conclusion is that TDMS models have pointed to many of the right mechanisms, but that it is possible to provide a simpler, more unified account by self-organization.

Experiment 1: Effect of Length on Grammaticality Judgments

Method

Participants. Twenty University of Connecticut undergraduates participated in the study for course credit. All were native speakers of English. None had participated in a related experiment before.

Materials. Experiment 1 used a 2 × 2 design with transitivity of the subordinate verb and length of the ambiguous NP as factors. An example item is shown in Sentence 5. Under the most plausible global analysis of each sentence, the subordinate clause verb was either transitive or intransitive and the subject NP of the main clause consisted of a determiner and noun alone (short condition) or of a determiner, noun, and modifying gerund phrase (long condition).

As the author wrote the book grew. (intransitive/short) (5a)

As the author wrote the book describing Babylon grew. (intransitive/long) (5b)

As the author wrote the essay the book grew. (transitive/short) (5c)

As the author wrote the essay the book describing Babylon grew. (transitive/long) (5d)

Although several other studies have used relative clause modification to lengthen the matrix subject NP (e.g., "the book that described Babylon"), we used the gerund because the memory load associated with processing a gerund is plausibly lower (fewer words used to express comparable meaning). For the gerund to serve as a lengthener without altering other syntactic properties of the sentence, it must be parsed as a modifier of the noun preceding it and not as a modifier of the subordinate clause verb phrase (VP) or subject (cf. "The author wrote the book living in Belfast/with a fountain pen"). To make sure our gerunds had an object NP attachment bias, we ran a norming study (word-by-word, self-paced reading, noncumulative display; see description in *Procedure* section) in which the sentence-final verb of the original intransitive/long condition was replaced by a full clause (Sentence 6a). Participants were asked a comprehension question that probed their attachment choice (Sentence 6b).

As the author wrote the book describing Babylon it grew. (6a)

Was it the author or the book that described Babylon? (6b)

The mean rate of attachment to the preceding NP object in the norming study was 48% (*SD* = 37%), indicating that the norming study was able to detect a significant range of attachment biases. The stimuli of Experiment 1 were, on average, highly biased toward local attachment (mean rate = 82%, *SD* = 38%).

The transitive conditions of Experiment 1 were designed so that readers would not misparse the subject NP of the matrix clause as the subject of an object relative clause modifying the preceding noun (as in "As the author wrote the essay [that] the book later critiqued . . ."). Phillips and Gibson (1997) analyzed reading times on sentences with similar structures and found that when the connectives were nontemporal (e.g., "because," "if") and the third NP of the sentence was a pronoun (e.g., "it" in place of "the book" in Sentence 5c), then readers tended to interpret the third NP as the subject of an object relative clause, but in the opposite condition (temporal connective, full NP for the third NP), the opposite tendency obtained. Twenty-nine of 36 of our Experiment 1 items used temporal connectives, and all used a full NP for the third NP of the control condition. We provide evidence in the *Results* section of Experiment 1 that our stimuli did not tend to induce the object relative interpretation.

We also incorporated semantic biases against interpreting the "-ing" words in our intransitive/long conditions as head nouns. Such a reading would occur, for example, if "book describing" in Sentence 5b were taken as analogous to "mountain climbing" in the (standard) interpretation in which "climbing" is the head of the compound. Such a reading is only plausible in stimuli structured like ours if the verb preceding the potential compound selects for an event or activity and if the putative compound itself is semantically felicitous. We designed our items to avoid these interpretations. We consider the impact of a few borderline cases in the *Results* section.

One other feature of our stimuli distinguishes them from those used in previous studies of similar phenomena: The disambiguating region consists of a single word, the last word of the sentence. The motivation for using a one-word, sentence-final disambiguating region was to concentrate the effects of garden-path recovery in this region. We hypothesized that at least part of the difficulty that prior researchers have had with detecting length effects in reading times has stemmed from the subtlety of the effects. We expected shortening the disambiguating region to enhance them. It is true that there is a cost incurred by seeking effects on the last word: Last-word reading times tend to have more noise, so there was a danger that the noise would obscure the effects of interest. We hypothesized that the concentration effect would be strong enough to overcome the noise. The use of the last word as a critical region also raises the possibility that any results we might get are due to sentence wrap-up effects, not to normal online processing. We return to this point in the Experiment 1 *Results* section, noting that our theory predicts the reading times we observe later, while we know of no theory of sentence wrap-up effects that does so. The stimuli for Experiment 1 are listed in Appendix B.

Four counterbalanced lists of items were constructed. Each list included 36 blocks of items. Each block contained two filler items and one stimulus item. Half of the filler sentences were grammatical and half were ungrammatical. The ungrammatical filler sentences were sentences like "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 participant read six practice trials and then read 1 of the 108-trial (3 × 36) lists.

Procedure. The sentences were presented on a computer monitor using noncumulative, word-by-word, self-paced reading (Just, Carpenter, & Wooley, 1982). Each trial started with an image of the sentence in which dashes replaced all the printed characters. Participants pressed the space bar to reveal each new word, causing the preceding word to revert to dashes. At the end of the sentence, participants answered the question, "Grammatical?" by pushing the *F* key for "yes" and the *J* key for "no." Participants were encouraged to read as naturally as possible and to make the grammaticality judgments according to their first impulse. The program recorded the grammaticality judgments.

The experiment was executed by a PsyScope program (Cohen, MacWhinney, Flatt, & Provost, 1993) written by David Perkowski and Daniel Richardson. The program was run on MacIntosh GE3s with 14-in. (35.56-

cm) monitors. The program was configured so that each stimulus sentence was displayed as a single line of text in a fixed-width font.

Predictions. We predicted, in keeping with Ferreira and Henderson's (1991b) results, an interaction between transitivity and length: Increased length should be associated with a greater decrease in positive grammaticality judgments in the intransitive conditions than in the transitive conditions.

Results

All participants scored at least 75% correct on the grammaticality judgments on the filler sentences, and all the data were used in the analysis.

Figure 1 plots the percentage of sentences judged grammatical in each of the four conditions. The rates of positive judgments were subjected to an analysis of variance (ANOVA) with two factors: transitivity and length. There was a main effect of transitivity, $F_1(1, 19) = 22.80$, $p < .01$, $F_2(1, 35) = 18.89$, $p < .01$. There was a main effect of length, $F_1(1, 19) = 8.72$, $p = .01$, $F_2(1, 35) = 8.23$, $p = .01$. The predicted interaction between transitivity and length occurred, $F_1(1, 19) = 10.52$, $p = .01$, $F_2(1, 35) = 4.79$, $p = .04$.

Experiment 1 thus replicates the results of Ferreira and Henderson (1991b) and Ferreira and Henderson (1993) with a control for memory load effects that is not confounded by variation in the semantics of the matrix subject.

To make sure that the outcome of the experiment was not dependent on the seven cases that lacked temporal connectives, we ran the analysis without these seven stimuli. The means had very similar values, and the same comparisons were significant. In fact, if the subject NPs of our control stimuli were parsed as subjects of object relative clauses, as Phillips and Gibson (1997) found in their nontemporal connective with pronominal NP condition, then readers ought to have judged the control sentences ungrammatical much more often than they did.

Some of the items in the Experiment 1 norming study (see Sentence 6) produced fairly low object-attachment ratings. The minimum rating was 39%. Some of the low ratings may have stemmed from ambiguity resulting from implicature (e.g., in response to Sentence 6b, it would not be unreasonable to answer "the author"). However, if any of the gerunds were frequently attached as verb phrase (VP) modifiers, then the long conditions might have been rated ungrammatical more often than the short conditions because they produced a different kind of garden path. To make sure that our effects were not dependent on the items with low object-attachment ratings, we ran an analysis on just the 29 items with an object attachment bias of at least 70%. Again, the means of the four conditions had very similar values and the same comparisons were significant, suggesting that the length effect did not stem from VP attachment of the gerund.

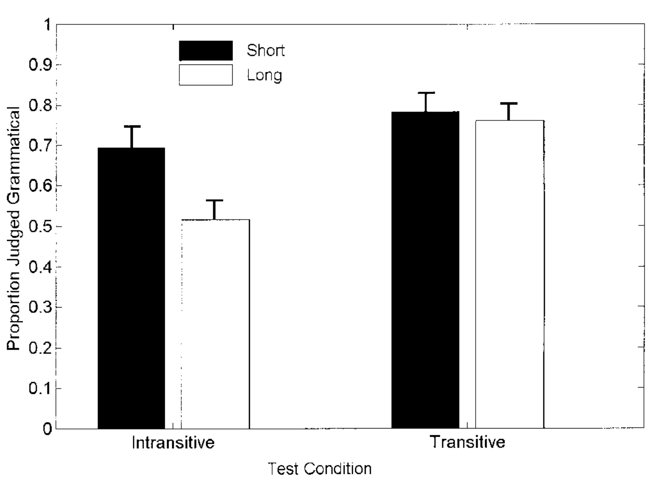


Figure 1. Proportions of sentences judged grammatical (+*SE*) in Experiment 1 as a function of transitivity and length.

As noted, the results would also be confounded if readers interpreted the "-ing" word in sequences like "book describing" in Sentence 5b as a nominal head (as in "mountain climbing"). This interpretation works if the gerund (a) is a plausible object of the subordinate verb and (b) forms a felicitous compound with the preceding noun. We judged that both of these conditions were met in 6 of our stimuli (Items 9, 10, 22, 25, 26, and 35). To make sure our effect was not due to the presence of these 6 items, we ran the analysis on the remaining 30 items. Again, the means of the four conditions had very similar values, and the same comparisons were significant.

Because time plays an explicit role in dynamical models, DSO accounts are well positioned to predict timing data. Therefore, Experiment 2 assesses whether the reading times themselves show the Transitivity × Length interaction.

Experiment 2: Effect of Length on Reading Times

Method

Participants. Forty-eight undergraduates from the University of Connecticut participated in the experiment for course credit. All were native speakers of English. None had participated in a related experiment before.

Materials. The stimuli for Experiment 2 were the same as those for Experiment 1, except that each sentence was followed by a comprehension question instead of a grammaticality judgment. Moreover, the ungrammatical filler sentences were replaced with grammatical sentences so that a meaningful comprehension question could be asked about each.

Procedure. The procedure was the same as in Experiment 1 except that reading times were recorded during the self-paced reading, and negative feedback was given when a comprehension question was incorrectly answered: The word "incorrect" appeared on the screen for 3 s.

Predictions. The predictions for Experiment 2 parallel the predictions for Experiment 1. We hypothesized that readers would initially attach the second NP in each condition as a direct object of the subordinate verb. This tendency is caused by lexical bias in our DSO model. Direct object attachment necessitates a reanalysis in the intransitive conditions. The reanalysis delays the stabilization of the link between the final verb and its subject, resulting in high reading times on the final verb. The model predicts a larger delay in the intransitive/long than the intransitive/short condition because the wrong attachment grows in strength during the reading of the modifier and takes correspondingly long to undo. Thus, the interaction between transitivity and length should manifest itself in the reading times on the final verb.

Results

All of the participants scored at least 75% correct on the comprehension questions (experimental items and fillers combined), and all the data were used in the analysis. For the purpose of

Table 1
Experiment 2 Region Division

Condition	1	2	3	4	5	6
A	As the author	wrote		the book		grew. (intrans/short)
B	As the author	wrote		the book	describing Babylon	grew. (intrans/long)
C	As the author	wrote	the essay	the book		grew. (trans/short)
D	As the author	wrote	the essay	the book	describing Babylon	grew. (trans/long)

Note. Intrans = intransitive; trans = transitive.

analyzing reading times by region, we divided each sentence into the regions shown in Table 1.

Before analyzing the reading times, we removed three individual trials with reading times greater than 10,000 ms. We then performed, for each participant, a linear regression on the reading times with characters-per-word as an independent variable. The regression analysis was a method of factoring out effects of word length, which are irrelevant to the purposes of the current study (see Trueswell, Tanenhaus, & Garnsey, 1994). The variance associated with word length was small (mean $R^2 = .0151$) but significant ($p < .01$). Figure 2 graphs the residual mean reading times by condition and region for Experiment 2.

Subject and item means were subjected to an ANOVA, with residual reading time as dependent variable and transitivity, length, and region as independent factors. We ran the analysis across the four regions for which all conditions contained data (Regions 1, 2, 4, and 6). The three-way interaction among transitivity, length, and region was significant in both the subjects and items analyses, $F_1(3, 141) = 7.42$, $MSE = 7950$, $p < .01$, $F_2(3, 105) = 4.83$, $MSE = 13622$, $p = .01$. This significant result supported examining the interactions on a region-by-region basis. For each region of interest, we ran separate ANOVAs on subject and item means, each with two factors: transitivity and length.

The predicted Region 6 Transitivity × Length interaction was significant, $F_1(1, 47) = 9.20$, $p = .01$, $F_2(1, 35) = 4.61$, $p = .04$. In the same region, there was also a main effect of length, $F_1(1, 47) = 29.70$, $MSE = 24920$, $p < .01$, $F_2(1, 35) = 13.14$, $MSE = 60658$, $p = .01$. The main effect of transitivity in this region was significant in the subjects analysis and marginally significant in the items analysis, $F_1(1, 47) = 12.31$, $MSE = 29214$, $p = .01$, $F_2(1, 35) = 3.92$, $MSE = 94914$, $p = .06$. There were no other significant main effects or interactions. Although the mean reading time in the transitive long condition at Region 6 was higher than that in the transitive short condition, this difference was not significant in a post hoc test. Nor was there any significant difference between the transitive and intransitive sentences in the short condition.

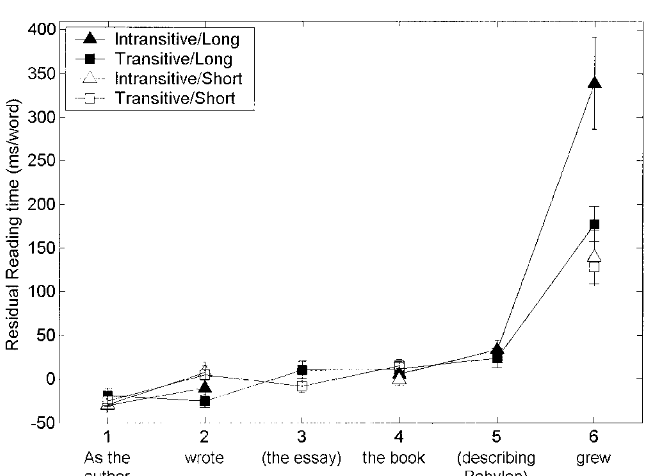


Figure 2. Residual reading times per word from Experiment 2 as a function of transitivity and length. Points represent means of residuals from the length regression. Vertical lines depict standard errors of the means.

We included each trial in which the comprehension question was incorrectly answered as well as each trial in which it was correctly answered. This choice was motivated by the assumption that processing times can reflect the natural tendencies of the parser even in cases in which the sentence is not perfectly understood. Rerunning the analysis with the incorrect trials held out produced the same pattern of significant contrasts except that the Transitivity × Length interaction was only marginally significant in the items analysis.

We also performed a full factorial ANOVA on the correctness of participants' answers to the comprehension questions. The pattern of the correctness means paralleled the reading time pattern on Region 6 in the sense that participants made the most errors on the intransitive/long condition, and all the other conditions had approximately the same rates of correctness: intransitive/short, 0.89; intransitive/long, 0.83; transitive/short, 0.89; transitive/long, 0.88. The effect of transitivity was significant in the subjects analysis, $F_1(1, 47) = 4.22$, $MSE = 0.0091$, $p = .05$, $F_2(1, 35) = 2.52$, $MSE = 0.0119$, $p = .12$. The effect of length was marginally significant in the subjects analysis and significant in the items analysis, $F_1(1, 47) = 3.86$, $MSE = 0.0174$, $p = .06$, $F_2(1, 35) = 11.47$, $MSE = 0.0083$, $p = .01$. The Transitivity × Length interaction was marginally significant in the subjects analysis, $F_1(1, 47) = 3.41$, $p = .07$. No other effects or interactions approached significance. A parallel analysis of the question-answering times produced no significant effects or interactions.

As in Experiment 1, to ensure that the outcome of the experiment was not dependent on the seven cases that lacked temporal connectives, we ran the analysis without these seven stimuli. The means had very similar values. The same comparisons were significant except that the predicted interaction between transitivity and length was only marginally significant in the items analysis.

Also, as in Experiment 1, to make sure that the effects were not dependent on the items with low object-attachment ratings, we ran an analysis on just the 29 items with an object-attachment bias of at least 70%. The means again had very similar values. The same comparisons were significant at Region 6 with the exceptions that the main effective of transitivity and the predicted Transitivity × Length interaction were only marginally significant in the items analysis.

Finally, as in Experiment 1, to ensure the observed effects were not due to interpretation of the gerund in the intransitive/long condition as the head of the direct object of the subordinate verb ("book describing" like "mountain climbing"), we ran the analysis on the 30 items that were clearly biased against this reading. Again, at Region 6, the means of the four conditions had very similar values and, except for the items analysis of the main effect of transitivity and the Transitivity × Length interaction, which were only marginally significant, the same comparisons were significant.

Discussion

These results provide evidence that ambiguous region length interacts with the presence versus absence of ambiguity to produce an effect on reading time data. Because this phenomenon is a fundamental prediction of DSO models that has not been clearly detected before, it is important to establish its existence.

The results conform to the predictions of the DSO models with one exception. The absence of a transitivity effect in the short condition alone at Region 6 (the disambiguating verb) does not bear out the prediction of the model of a small garden-path effect here. It may be that the design was not sensitive enough to detect this small, predicted effect.

The fact that the Transitivity × Length interaction is significant on the last word of the sentence, in which high variance often washes out subtle effects, is encouraging because it suggests that the effect is a robust one. On the other hand, variation on the last word may stem from interpretive processes that are different from the processes that govern midsentence reading. Thus, as an anonymous reviewer pointed out, a last-word effect might not be an online effect. The current experiment does not rule out this possibility. However, the accounts of which we are aware along these lines do not seem preferable to the DSO account. Longer processing times on the last word of a sentence may be due to an effort by readers to take stock of what they have just read. However, it is not clear why stock taking should take longer in some conditions than others. Alternatively, it may be that readers pause, perhaps to rest, when they have just processed a garden-path sentence. In that case, readers should rest equally in the short and the long conditions. On the other hand, it might be that they need to rest more in the long condition because it involves a difficult recovery. In that case, their resting behavior would be a direct reflex of their parsing behavior. Positing a postreading interpretive process adds unmotivated complexity to the account.

The advantage of the DSO framework is that it predicts the data as a consequence of mechanisms that are needed for handling normal parsing. It is noteworthy that in one of their five experiments, Ferreira and Henderson (1991b) found that the Transitivity × Length interaction manifested itself in the response times associated with grammaticality judgments as well as in the rates of positive judgments. With a minimal extension, online DSO models predict these data as well. In particular, we assume that there is a limit on how long each word can be read, as proposed in Ferreira and Henderson (1990; but see McRae et al., 1998; Rayner, Sereno, & Raney, 1996; Vitu & O'Regan, 1995). When this time limit has been exceeded after reading the last word, the model moves on to the grammaticality judgment task. Before assessing grammaticality, it waits until the whole parse has stabilized. If we set the move-on threshold to be shorter than the stabilization time of hard cases, the model has additional stabilizing to accomplish in the intransitive/long condition, and the judgment in that condition takes a relatively long time.

Experiment 3: Effect of Revision Size on Reading Times

If the self-organization account is correct in claiming that parses are created through the competitive modification of the strengths of bonds, then a further prediction about the relative difficulty of reanalyses obtains: A reanalysis that involves severing more links should be more difficult than one involving severing fewer links, other factors being equal. Our DSO model makes this prediction because of its link/fragment feedback: Fighting battles on several fronts makes the fragment weaker than fighting on fewer fronts. The fragment thus sends less support to the links, and these, in turn, have more difficulty getting established. Experiment 3 was designed to test this prediction.

Method

Participants. Twenty-seven undergraduates from the University of Connecticut participated in the experiment for course credit. All were native speakers of English. None had participated in a related experiment before.

Materials. The three conditions shown in Table 2 were compared.

Condition A in Table 2 is a control case that is similar to the long transitive condition of Experiment 2. SOPARSE predicts that bonds will not be broken once formed, in this case, because the optimal interpretation early in the reading of the sentence (the transitive reading of "wrote") remains optimal as the remainder of the sentence is perceived. Consequently, when the matrix VP "grew rapidly in her mind" is encountered, the model predicts that it will be read quickly. In keeping with the terminology of Sturt and Crocker (1996), we refer to this as the monotonic condition; parse construction is predicted to proceed without backtracking.

Table 2
Experiment 3 Condition Comparison

	Stimuli									
Condition	1	2	3	4	5	6	7	8	9	10
A	As the author	wrote	the essay	. . . the book	that	she envisioned	grew	rapidly	in	her mind. (monotonic)
B	As the author	wrote		. . . the book	that	she envisioned	grew	rapidly	in	her mind. (one revision)
C	As the author	wrote		. . . the book		she envisioned	grew	rapidly	in	her mind. (two revisions)

Condition B in Table 2, in contrast, is predicted to be more difficult. In this case, the favored transitive parse of "wrote" leads to attachment of "the book that she envisioned" as a direct object, but the matrix VP forces reattachment of this NP as the subject of the matrix clause (just as in the intransitive conditions of Experiment 2). In this case, one bond—that between "wrote" and its direct object—must be severed. We refer to this as the one-revision condition. Figure 3 shows the tree structure that the model builds just before and just after encountering the matrix verb phrase in Condition B. Condition C of Table 2, based on examples from Warner and Glass (1987)'s study of length effects, was designed to induce the reader to make two wrong attachments before the disambiguating information arrived. In particular, it was expected that, just as in Condition B, "the book" would be attached as the direct object of "wrote." The following pronoun, "she," can then be attached as either the subject of an embedded relative clause or the subject of the main clause. We hypothesized that the use of the temporal conjunction "as" results in a bias in favor of main clause attachment (see later discussion). The following word, "envisioned," should reinforce this interpretation. However, when "grew rapidly" arrives, the parse must be radically revised. Not only must the link between "the book" and "wrote" be severed, but the link between "she envisioned" and the IP subcategorized by "as" must as well. Because "grew rapidly" must compete with both "wrote" and "as" in Condition C but only with "wrote" in Condition B, reading the first few words of the main clause VP is predicted to take longest in Condition C. Figure 4 shows the tree structures before and after the matrix clause revision for this "two-revisions" case.

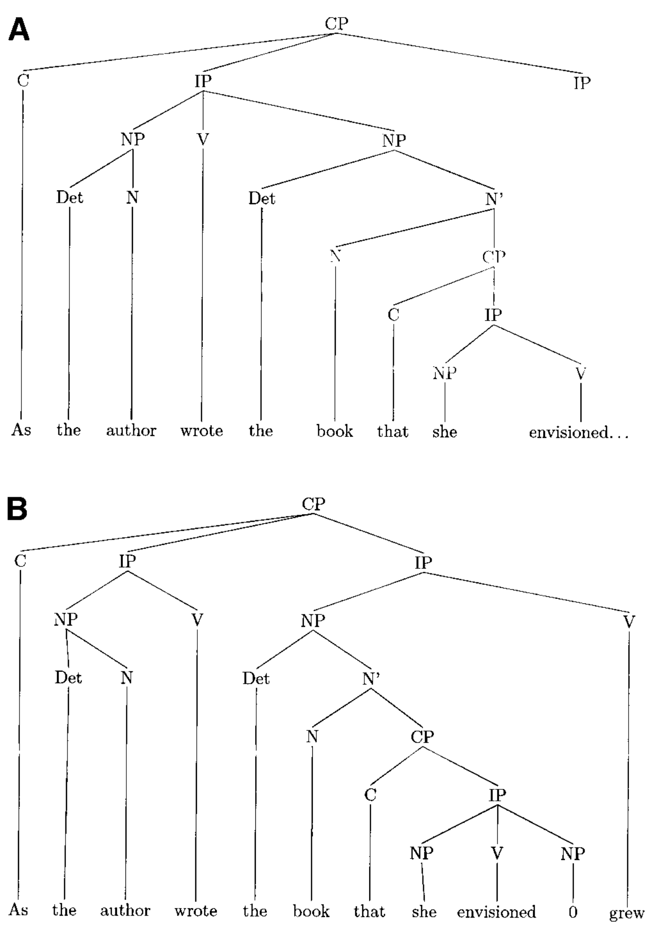


Figure 3. (A) Initial and (B) final analyses of the sentence "As the author wrote the book that she envisioned grew" (Condition B of Table 2) before and after reading the word "grew." CP = complementizer phrase; C = complementizer; IP = inflectional phrase; N′ = intermediate nominal projection; NP = noun phrase; V = verb; Det = determiner.

As noted in the discussion of Experiment 1, Phillips and Gibson (1997) used word-by-word self-paced reading to study sentences with similar structure. Their sentences (Sentence 9) were temporarily ambiguous in the underlined region: The underlined clause could be attached either as a reduced relative modifier of the preceding verb or as the matrix clause. A subsequent prepositional phrase disambiguated in favor of subordinate attachment. Phillips and Gibson's stimuli were divided into two major groups[3]: those that had a temporal complementizer (e.g., "while," "as," "when") and a full NP subject in the ambiguous region (Sentence 9a, p. 343) versus those that had a nontemporal complementizer (e.g., "because," "although") and a pronominal subject in the ambiguous region (Sentence 9b, p. 330). They also had conditions that were disambiguated in favor of subordinate clause attachment of the underlined region. All stimuli were compared with unambiguous controls. The finding was that readers tended to choose matrix attachment when the complementizer was +temporal/−pronominal, whereas they tended to choose subordinate attachment when the complementizer was −temporal/+pronominal. That is, they slowed down during the four words after disambiguation, relative to unambiguous controls, when the just-mentioned preferences were violated.

While I talked with the lawyer John was watching at the party

I became rather nervous. [Subordinate, temporal, full NP] (9a)

Because Rose praised the recipe I made for her birthday it was

worth all the effort. [Subordinate, nontemporal, pronoun] (9b)

Phillips and Gibson's (1997) stimuli had two highly correlated features: They used temporal connectives almost exclusively with overt NPs in their focused ambiguous region ("While I talked with the lawyer John was watching . . .") and used nontemporal connectives almost exclusively with pronominal NPs ("Because Rose praised the recipe I made . . ."). Thus, it is not clear whether the bias in their stimuli stemmed from the type of connective (temporal vs. nontemporal) or from the nature of the subject of the ambiguous clause (pronoun vs. full NP). We hypothesized that using pronouns that matched the subordinate subject in number and gender in combination with temporal complementizers would enhance the bias toward matrix attachment because it is natural for the agent of the first clause to be the agent of the second clause when a situation of coincident timing is being discussed. This was the condition of which Phillips and Gibson had only one example, so their results did not provide a clear indication about the bias in such cases.

To objectively establish a set of stimuli with a bias toward matrix attachment, we conducted a norming study using Phillips and Gibson's (1997) 10 subordinate bias cases (P&G stimuli) and our 28 new cases (T&H stimuli), most of which had temporal complementizers and all of which had pronoun subjects of the ambiguous clause. Partial sentences of the form shown in Sentence 10 were interspersed with incomplete filler sentences and presented to participants on pencil-and-paper questionnaires. The participants were asked to complete each sentence in the first way that came to mind.

Because Rose praised the recipe I made . . .

(P&G, −temporal/+pronominal) (10a)

As the author wrote the book she envisioned . . .

(T&H, +temporal/+pronominal) (10b)

The completions on the critical items were scored as to whether they involved subordinate attachment, matrix attachment, ambiguous attach-

[3] They also had one +temporal, +pronominal stimulus and two -temporal, -pronominal stimuli, but, as they note, these were so few in number that it is not clear what influence they had on the results.

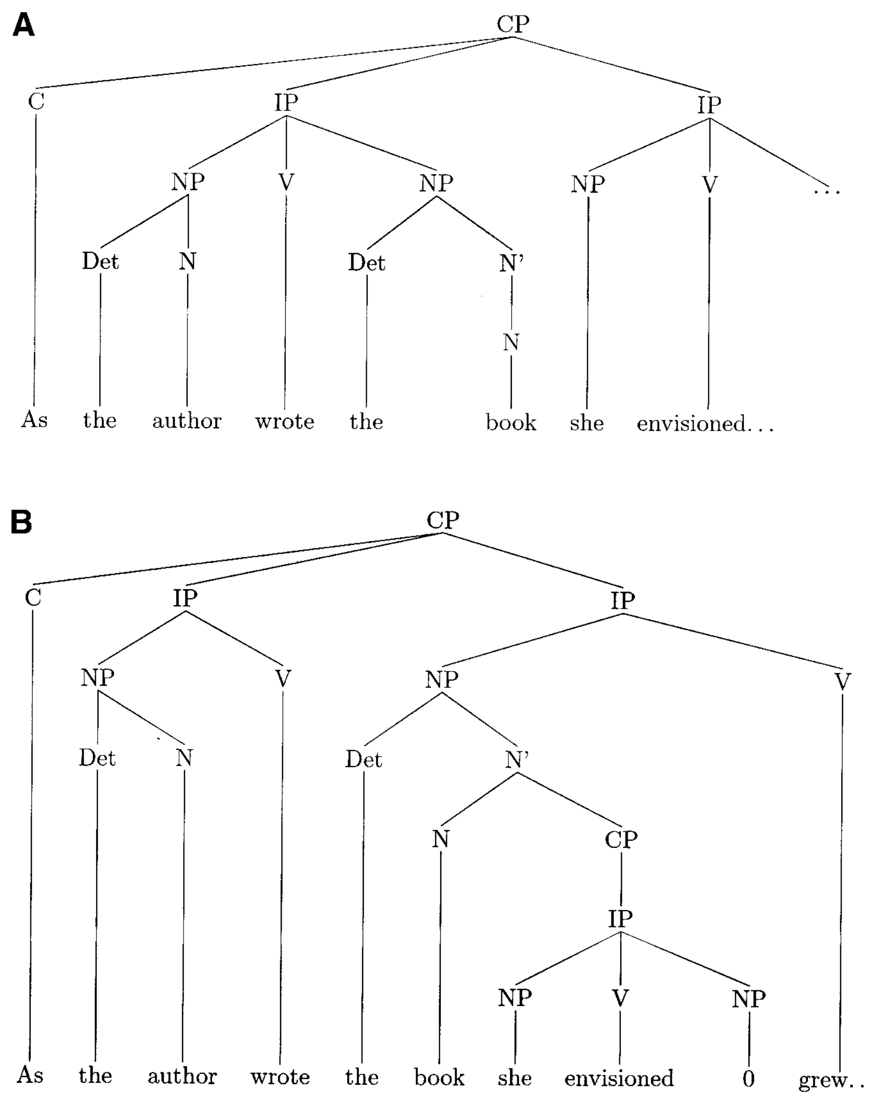


Figure 4. (A) Initial and (B) final analyses of the sentence "As the author wrote the book she envisioned grew" (Condition C of Table 2) before and after reading the word "grew." CP = complementizer phrase; C = complementizer; IP = inflectional phrase; N′ = intermediate nominal projection; NP = noun phrase; V = verb; Det = determiner.

ment, or none. The ambiguous attachment and none classes comprised 2% and 1% of the responses, respectively, and were removed from further analysis. The mean rate of matrix attachment (vs. subordinate attachment) for the P&G stimuli was 0.41 (SD = 0.30, N = 10), whereas the rate of matrix attachment for the T&H stimuli was 0.73 (SD = 0.17, N = 28).

A paired (by subjects) t test confirmed that the difference between the T&H items and the P&G items was significant ($p < .01$). These results suggest that the T&H items were biased in favor of matrix attachment.

Among the 28 T&H items, 21 were attached to the main clause more than 70% of the time in the norming study and thus could be said to have a strong bias toward matrix attachment. These 21 items were used in the main experiment.

Three lists were constructed. Each list included 21 blocks of items, with seven items in each of the three conditions. Each block contained four filler items and one stimulus item. 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. To counterbalance the two-revisions stimuli, which always exhibited subordinate attachment of the pronoun and following verb, seven structurally analogous sentences exhibiting main clause attachment of a pronoun and following verb (see Sentence 11) were included among the fillers. These examples were distinct from the stimuli in that they used distinct verbs and described distinct scenarios.

Whenever Clementine tried out a shoe

she found that it pinched her foot in several places. (11)

Each list started with a block of five filler items. Each participant read six practice trials and then read one of the 110-trial (5 × 21 + 5) lists. The lists were counterbalanced across participants.

Procedure. The procedure was the same as in Experiment 2.

Results

All participants scored at least 75% correct on the comprehension questions (stimuli and filler questions combined), and all the data were used in the analysis.

Before analyzing the reading times, we removed two individual trials with reading times greater than 10,000 ms. We then performed, for each participant, a linear regression on the reading times with characters-per-word as an independent variable. The regression analysis allowed us to factor out effects of word length, which are irrelevant to the purposes of the current study (see Trueswell et al., 1994). The variance associated with word length was small (mean $R^2 = 0.009$) but significant ($p < .01$). Figure 5 shows the means of the residuals from this regression, split by condition for Experiment 3.

Residual reading times were subjected to a full factorial ANOVA with tree change (monotonic vs. one revision vs. two revisions) and region as independent factors. The analysis was run across the eight regions for which all conditions contained data (Regions 1, 2, 4, 6, 7, 8, 9, and 10). There was a main effect of tree change, $F_1(2, 52) = 22.61$, $p < .01$, $F_2(2, 40) = 6.57$, $p = .01$. There was a main effect of region, $F_1(7, 182) = 11.66$, $p < .01$, $F_2(7, 140) = 25.42$, $p < .01$. The Tree Change × Region interaction was significant in both the subjects and items analyses, $F_1(14, 364) = 5.98$, $p < .01$, $F_2(14, 280) = 7.45$, $p < .01$.

To find out where the means differed significantly, we ran post hoc tests on a region-by-region basis using Fisher's protected least significant difference test. No means differed except in Regions 7, 8, and 9. The one-revision condition was read significantly faster than the two-revisions condition in Region 8. Likewise, the monotonic condition was read significantly faster than the one-revision condition in Region 9 and marginally faster in Regions 7 and 8.

Again, we included trials in which the comprehension question was answered incorrectly as well as trials in which the comprehension question was answered correctly. Removing the incorrect trials produced the same pattern of results.

Analyses of the correctness of participants' answers to the comprehension questions and of the question-answering times produced no significant effects or interactions.

Discussion

The results suggest that readers suffered from a garden-path effect in both the one-revision and the two-revision conditions. Their reading times were slowed within the two words after the disambiguating verb. The fact that the effects did not occur strongly on the verb itself, but were spread over the verb and the

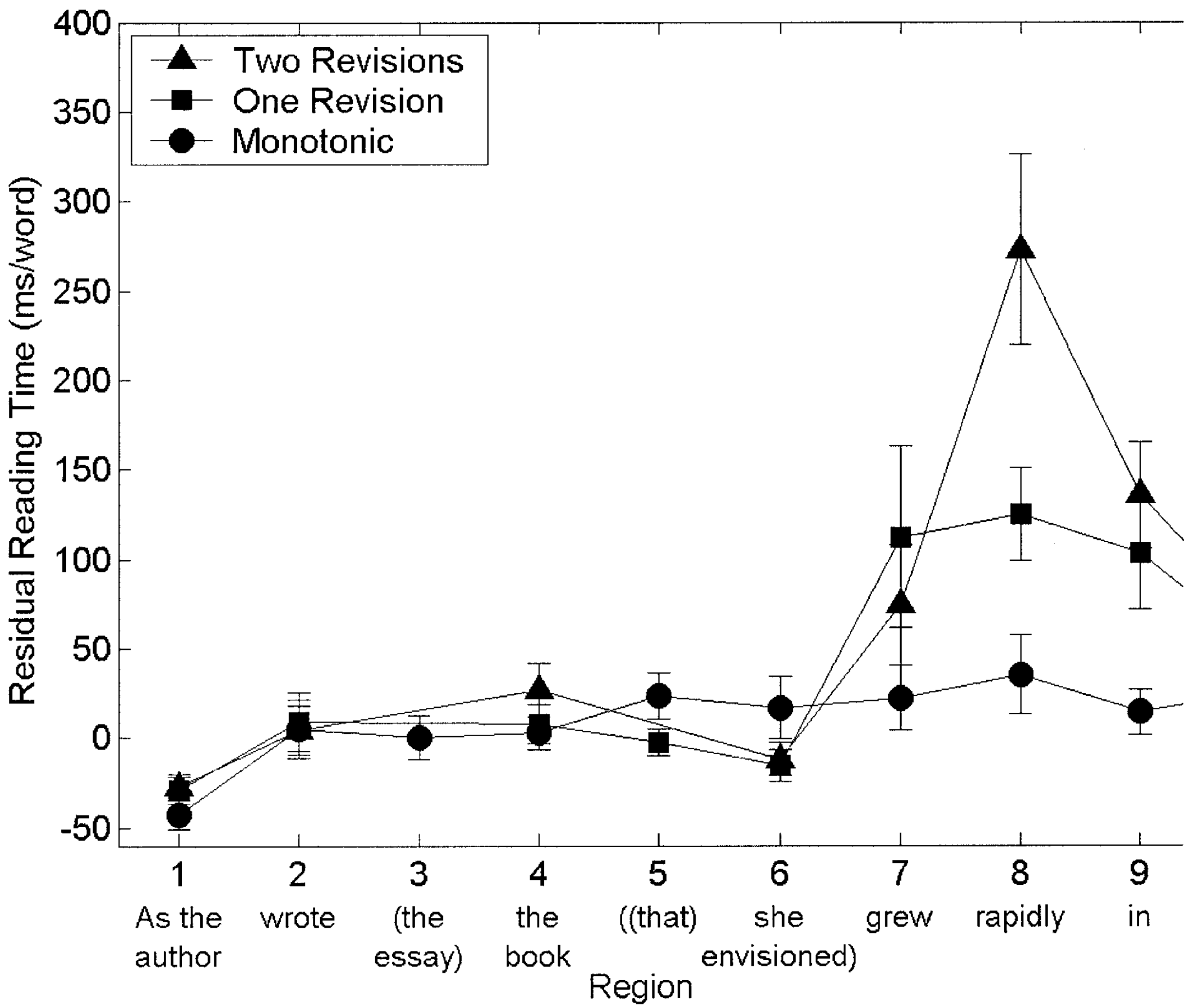


Figure 5. Residual reading times per word from Experiment 3, split by tree change. Points represent means of residuals from the length regression. Vertical lines depict standard errors of the means.

following two words, is consistent with the assumption underlying the design of Experiment 2: that having a one-word disambiguating region intensifies garden-path effects by concentrating them. The difference between the one-revision condition and the monotonic condition, in combination with the difference between the two-revision condition and the one-revision condition lend support to the claim that processing difficulty increases with the amount of revision that needs to be made to the current analysis, in keeping with the predictions of SOPARSE.

Ferreira and Henderson (1991b) included sentences that were similar in structure to those of the one-revision and two-revisions conditions in RSVP format and self-paced segment-by-segment reading, with rate of grammaticality judgment as the dependent measure. They reported a slight numerical difference between the two cases that went in the opposite direction of the result here (the one-revision condition was incorrectly judged to be ungrammatical more often than the two-revisions condition). They did not indicate that this difference was significant. They also found extremely low rates of correct judgments on these sentences (roughly 22% on average), and the subjects of the relative clauses were NPs of the form "the" + noun rather than pronouns as used here. The difference between Ferreira and Henderson's (1991b) results and those of Experiment 3 may stem from the difference in task: The online measure may pick up contrasts in the magnitudes of people's reactions at the point at which the sentence becomes confusing, but both kinds of sentences are so difficult that participants tend to judge them ungrammatical equally often by the time the end of the sentence has been read. The difference may also stem, in part, from the use of a nominative pronoun subject of the relative clause in the current experiment. As noted, readers may favor identifying the matrix subject with the subordinate subject for pragmatic reasons. The nominative pronoun in our two-revisions condition permits them to do this, whereas a full NP does not.

General Discussion

The results of all three experiments support the notion, suggested by many researchers, that there are degrees of difficulty of reanalysis. We now compare our account of the reanalysis process with several others. Our self-organizing account predicts the data from a common core of principles, whereas other accounts fail to predict the results or require several disjoint stipulations to do so. In fact, analogues of several of the mechanisms posited by other accounts arise naturally under the self-organization approach. Thus, we suggest that the self-organization approach is helpfully unifying.

SOPARSE: A DSO Parser

As we noted early in this article, SOPARSE is organized around a competition among links. At the center of the theory is a feedback mechanism that reinforces links contributing to successful fragments and reinforces fragments that have strong links. Fragments get activated by the input.

This "rich get richer" mechanism predicts a Transitivity × Length interaction in Experiments 1 and 2 because of self-reinforcement of links. In an example like Sentence 5b, the link between the noun "book" and the direct object foot of the verb "wrote" becomes stronger during the reading of the gerund clause "describing Babylon." Therefore, when the disambiguating verb ("grew") arrives, it takes a long time to undo the erroneous attachment. (See also the more detailed description in Appendix A.)

A plausible additional assumption allows the model to predict grammaticality judgments. Following Vosse and Kempen (2000), we assume that a small amount of noise is continuously added to the link activations. The model judges a sentence ungrammatical if it fails to parse it. Because the wrong interpretation has a lead initially in the intransitive conditions of Experiment 1, the noise will sometimes bump it so high that recovery is impossible. Length interacts with transitivity because, in the intransitive/long condition, there is a longer window of opportunity for the noise to bump the wrong analysis past the point of no return.

Returning to reading time predictions, the model also predicts a small main effect of length that is independent of the Transitivity × Length interaction just discussed: Response times on "grew" are elevated in the transitive conditions relative to the intransitive conditions because of interference from the material in the modifying phrase ("describing Babylon"). In particular, the bond from "Babylon" to the subject foot of "grew" provides weak competition for the correct link from "book" to "grew." This slows the attachment of "grew" to "book" by a small amount. We note that SOPARSE's behavior here parallels the claim of storage cost accounts like Gibson's (1998) that a cost is associated with parsing a head that is separated from a dependent, but it attributes the cost to a specific cause: interference. Van Dyke and Lewis (2003) and Vasishth (2002) present evidence supporting the view that interference can play a significant role in such cases. Nevertheless, the effect of length in the transitive condition alone was not significant in our experiment, so it is appropriate that SOPARSE predicts a weak effect here.

SOPARSE treats the monotonic and one-revision conditions of Experiment 3 similarly to the transitive/long and intransitive/long conditions of Experiment 2. In the two-revisions condition of Experiment 3, the link from "book" to the direct object foot of "wrote" and the link from "envisioned" to the second foot of "as" are both well established by the time the disambiguating word "grew" arrives. Thus, "grew" must fight a battle on several fronts to get established and does not succeed in doing so. Consequently, the feedback that the "grew" fragment obtains from its links is so weak that the fragment is not an effective competitor and proper parsing fails. (See Appendix A.)

We now consider other possible accounts of our findings, discussing other DSO and TDMS accounts.

Other DSO Accounts

Unification Space

In Vosse and Kempen's (2000) unification space (U-Space) parser, when words are perceived, they enter the "unification space" and attempt to form links with open nodes on other words, obeying linear order constraints and syntactic category constraints. The U-Space parser is similar to SOPARSE and was a major inspiration for it. As in the noisy version of SOPARSE, noise in the link activations tends to make the parse fail in difficult examples, thus predicting grammaticality judgments. On the other hand, the U-Space parser is not intended as a model of word-by-word processing times and does not make meaningful predictions about them. Also, the U-Space parser assumes that the activations of tree fragments decay with time and length effects stem from decay. The U-Space parser thus does not seem to have a way of predicting the multiple-attachment digging-in effects of Experiment 3.

Competitive Attachment

Stevenson's competitive attachment parser (Stevenson, 1994, 1997, 1998; Stevenson & Merlo, 1997) claims that the language processor maintains a single parse tree for each sentence and that alternative ways of attaching new material to this parse tree compete with one another in an activation settling framework. With some adjustments, Stevenson's model appears to be able to handle the results of both Experiments 2 and 3.

Stevenson (Stevenson, 1998; Stevenson & Merlo, 1997) posits a right-edge restriction: Only attachment sites along the right edge of the parse tree are available for attaching new material. This mechanism predicts that the two-revisions case of Experiment 3 will be impossible to reanalyze: As Figure 4 indicates, the attachment of "she envisioned" as subject and verb of the matrix clause takes "the book" off the right edge. Because it is "the book" that needs to be reanalyzed as matrix subject to achieve the correct parse, the parse fails.

Stevenson (1998) shows that the competitive attachment mechanism also fails to parse garden-path sentences of the type that occur in the intransitive conditions of Experiments 1 and 2 and in the one-revision condition of Experiment 3. As in the SOPARSE account, Stevenson assumes that the intransitive uses of optionally transitive verbs (like "wrote") are instantiated by attachment of the verb's direct object site (or, in our terms, "foot") to a null node. Stevenson assumes that null nodes are generally weaker competitors than their overt counterparts and that once they have been usurped by an overt NP, they disappear. Therefore, in the cases at hand, the parser will attach "the book" as the direct object of "wrote," lose the null node, and then have no way of recovering when "grew" arrives. Stevenson (1998) notes, however, that sentences like "When Kiva left the room got quiet" (p. 355) are not as difficult as other garden paths for which the right-edge constraint is intended, and she remarks that one could include a special mechanism that would allow resurrection of defunct null nodes. If it is assumed that a moderate cost is associated with invoking this mechanism, then competitive attachment will predict the three-way contrast of Experiment 3. With only these assumptions, the framework does not predict the ambiguous region length effects of Experiments 1 and 2, but Stevenson (1998) also posits a decay mechanism, which could plausibly handle these results. We prefer the SOPARSE approach to the competitive attachment model because it handles the data with fewer assumptions.

TDMS Accounts

Monotonicity

Weinberg (1993, 1995), Gorrell (1995), Sturt and Crocker (1996, 1997, 1998), and Sturt et al. (1999) develop an account of processing difficulty based on the notion that some new information can be accommodated by building more structure upon structure previously built (monotonic), whereas other new information requires revising previous assumptions (nonmonotonic). Their claim is that the nonmonotonic cases will be more difficult than the monotonic cases. In earlier formulations, the focus was on constituent structure changes, whereas several later formulations (Sturt & Crocker, 1997, 1998) argue that it is revision of thematic structure that matters.

This class of models predicts the main effect of intransitivity in Experiments 2 and 3 (i.e., the familiar garden-path effect) by postulating that there is a structural preference for transitive interpretation of the first verb, so only the cases in which the verb turns out to be intransitive require revision of previously built configurational (or thematic) structure. These models do not predict the main effect of length in Experiment 2 because the long and short versions do not differ in monotonicity. For the same reason, they also do not predict any interaction between transitivity and length. Although the articles cited do not consider what happens when multiple revisions must be made, it is natural to assume that more violations of monotonicity are more costly than fewer. In this sense, the framework is well positioned to predict the tree change effects of Experiment 3. Indeed, the central insight of monotonicity—that bigger structural changes are more difficult than smaller ones—is fundamental to our self-organizing account.

Visibility

Frazier and Clifton (1998) propose that the notion of visibility captures a number of insights about sentence-processing difficulty. A node in a phrase structure tree is more visible if (a) it was postulated more recently or (b) it is part of the same "perceptually-given package" as the current node. An example of a perceptually given package is the kind of phonological phrase that is marked by phrase-final lengthening. The authors hypothesize that the parser has a preference to attach an incoming constituent to the most visible site with which it is grammatically compatible. Processing slows down whenever some or all of a postulated tree structure must be revised (i.e., whenever there is reanalysis). Moreover, the difficulty of any attachment is inversely proportional to the visibility of the attachment site.

The visibility account predicts the main effect of length in Experiments 1 and 2 with its recency clause: Because the VP node that needs to be revised in the intransitive conditions has been more recently postulated in the short condition than the long condition, the parser has an easier time finding it in the short condition and the recovery is quicker. In the transitive conditions, the visibility account also predicts an effect of length because the IP node to which the main verb ("grew") needs to attach has been more recently postulated in the short condition than in the long condition. However, because the visibility principle applies in both first analysis and reanalysis, the visibility part of Frazier and Clifton's (1998) account does not predict an interaction between transitivity and length.

Visibility provides helpful insight into the complementizer effect of Experiment 3 by its "perceptually given package" concept. In the one-revision case, it seems reasonable to assume that, at the point of reading the matrix verb ("grew"), the current package is still the subordinate clause (see Figure 3A). However, at the analogous point in the two-revisions case, the current package is unequivocally the matrix clause package (see Figure 4A). Therefore, the bond between the subordinate VP and its direct object NP, which must be broken, is less visible in the two-revisions case than in the one-revision case.

Visibility and digging-in are similar conceptions: Under our DSO account, as one analysis becomes more dug in, competing analyses become weaker and thus harder to revive; in a sense, they become less visible. We prefer our version over Frazier and Clifton's (1998) because its treatment is less disjunctive: Both single-attachment and multiple-attachment digging-in effects stem from the "rich get richer" principle.

Attach Anyway

Fodor and Inoue (1998; see also Fodor & Inoue, 1994, 2001; Inoue & Fodor, 1995) propose that the parser always attaches a perceived word to the current parse tree, guided by the grammar and a simplicity principle (minimal everything) that subsumes minimal attachment, late closure, and other principles. If there is no felicitous attachment site, it attaches the word at the node, *n*, where it "least severely violates the grammar" (Fodor & Inoue, 1998, p. 105) and least severely violates the parser's preference for simpler structures. In cases in which the grammar is violated, the grammatical dependency principle (GDP) mandates that the parser modify a part of the tree on which *n* is grammatically dependent in order to fix the problem. The GDP causes the modification mechanism, called adjust, to apply iteratively until the parse is grammatical or no change can be made. Under these assumptions, the account predicts the observed transitivity effects because the transitive cases will be parsed on the first pass, with no need for revision, whereas the intransitive cases will fail to be parsed. In particular, in Sentences 5a and b, "grew" projects IP and initially attaches where it needs to as the main verb of the matrix clause, but it lacks a subject. Because the NP that needs to be its subject ("the book") is embedded in the VP of the subordinate clause, a node that is not grammatically dependent on the complementizer phrase (CP) node of the matrix clause where "grew" has attached, the adjust operation fails to capture a subject for "grew." This does not mean, however, that cases like Sentence 5a and b are never parsed. There is a mechanism, called theft, that searches progressively back through the word string in such cases to find an element that would fit the needs of the unsatisfied node. In the case of Sentences 5a and b, searching back identifies "book" as a subject for "grew," whereupon "book" and its determiner, "the," are installed as the subject of "grew" and the lexicon is checked to see whether "wrote" can be intransitive. Because it can, the parse goes through after all. However, the likelihood of theft's success is assumed to decrease as a function of how far back in the string the word search must proceed. Consequently, Fodor and Inoue (1998) predict not only the main effect of transitivity but also the interaction between transitivity and length because the long condition involves a longer string search.

If it had no additional mechanisms, the attach anyway framework would not predict the complementizer effect of Experiment 3. In particular, assuming that "she envisioned" in Condition C in Table 2 were treated as the beginning of the matrix clause in the initial parse, then "grew" would be attached as a matrix verb. Adjust would then attempt to fix the problem of having two nonconjoined matrix verbs, but it would fail to discover the correct solution because "the book," which needs to be the matrix subject, is embedded in the subordinate VP and thus not grammatically dependent on any matrix nodes. At this point, the word search associated with theft could commence, and "the book" might be discovered as a reasonable subject of "grew," as occurred in the intransitive conditions of Experiment 2. However, the length of the search would be slightly shorter in the two-revisions condition, so the account would make the opposite prediction from what was observed. On the other hand, if a mechanism for undoing the incorrect attachment of "she" and "envisioned" is used, then this process will add time to the two-revisions, but not the one-revision, condition, thus correctly distinguishing those.

There are several commonalities with the self-organization approach: an information-driven parse modification mechanism, a willingness to build grammatically unsanctioned structures, an apparently quantifiable measure of severity of violation, and the notion, in capture and theft, of a kind of competition among verbs for arguments. Our modeling results suggest that some of the complexity of the attach anyway account may be removed if these notions are used in an activation-settling framework.

Modular Syntax With Thematic Activation

Ferreira and Henderson (1991a) have proposed that a modular syntactic parser guided by the principles of minimal attachment and late closure sends its output to a thematic processor, which controls the activations of verb argument structures.

In Sentences 5a and b, for example, the syntactic parser initially parses "the book" as the direct object of "wrote" and sends its output to the thematic processor. When the thematic processor detects the word "wrote," it initially activates both the intransitive and transitive frames of that verb, but the information coming from the syntactic processor depresses the activation of the intransitive frame in favor of the transitive. Then, when the disambiguating information arrives ("grew"), the deactivated thematic structure must be reactivated. If "grew" arrives immediately after the head ("book") of the ambiguous noun phrase, as in Sentence 5a, then reactivation is easy, but if modification separates the head from the disambiguating information, as in Sentence 5b, then reactivation is hard, and readers are expected to be more likely to judge the sentence ungrammatical and to read more slowly. No such contrast occurs in the transitive conditions because the syntactic processor sticks with its initial parse throughout. In this way, the account predicts the Transitivity $\times$ Length interaction of Experiments 1 and 2.

This account is very similar to ours: Both involve the waxing and waning of verb subcategorization frames under the influence of (at least) structural cues. However, the two-revisions case of Experiment 3 does not involve any more reactivation of argument frames than the one-revision condition; in fact, it involves slightly less because it lacks the extra word, "that."

Summary

DSO models rely fundamentally on the assumption that constraint strengths add together and compete with one another. Under the particular parameterization of DSO principles that we have adopted in SOPARSE, this assumption forms the basis of the ganging effects, which permit recovery from garden paths in cases in which the system has not become too entrenched in an incorrect analysis. The competition also provides the basis for the "rich get richer" principle, which produces the two kinds of digging-in phenomena reported here. As we note, SOPARSE exhibits behaviors that closely resemble mechanisms posited by several TDMS approaches. Interference resulting from attachment competition predicts memory load effects similar to those that Gibson (1998) handles by positing a limited capacity store. Digging-in is analogous to Frazier and Clifton (1998)'s visibility. Activation-based competition between verbs for arguments is similar to the operations of capture and theft in Fodor and Inoue's (1998) attach anyway framework. The use of semantic information to guide the formation of bonds between lexical heads is similar to Ferreira and Henderson's (1991a) use of thematic activation. However, in SOPARSE, unlike in TDMS approaches, these behaviors are by-products of the model's central constraint-resolution mechanism, the mechanism by which it accomplishes normal parsing. Thus, an advantage of the self-organization approach is its simplicity: A range of kinds of processing effects sometimes thought of as distinct can be captured by the interaction of independently motivated dynamical and linguistic constraints.

References

Bailey, K. G. D., & Ferreira, F. (2003). Disfluencies influence syntactic parsing. *Journal of Memory and Language, 49,* 183–200.

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

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning, 7,* 195–225.

Ferreira, F., & 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., & Henderson, J. M. (1991a). How is verb information used during syntactic parsing. In G. B. Simpson (Ed.), *Understanding word and sentence* (pp. 305–330). Amsterdam: North-Holland.

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

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

Ferreira, F., & Henderson, J. M. (1995). Reading processes during syntactic analysis and reanalysis. In J. M. Henderson & M. Singer (Eds.), *Reading and language processing* (pp. 119–147). Hillsdale, NJ: Erlbaum.

Ferreira, F., & Henderson, J. M. (1998). Syntactic reanalysis, thematic processing, and sentence comprehension. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 73–100). Dordrecht, The Netherlands: Kluwer Academic.

Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research, 23,* 407–434.

Fodor, J. D., & Inoue, A. (1998). Attach anyway. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 101–141). Dordrecht, The Netherlands: Kluwer Academic.

Fodor, J. D., & Inoue, A. (2001). Garden path re-analysis: Attach (anyway) and revision as last resort. In M. D. Vincenzi & V. Lombardo (Eds.), *Cross-linguistic perspectives on sentence processing* (p. 21–61). Dordrecht, The Netherlands: Kluwer Academic.

Frazier, L., & Clifton, C. (1996). *Construal.* Cambridge, MA: MIT Press.

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

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.

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

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, A. Marantz, & W. O'Neill (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.

Gorrell, P. (1995). Japanese trees and the garden path. In R. M. N. Nagai (Ed.), *Japanese sentence processing* (pp. 331–350). Hillsdale, NJ: Erlbaum.

Grodner, D., Gibson, E., Argaman, V., & Babyonyshev, M. (2003). Against repair-based reanalysis in sentence comprehension. *Journal of Psycholinguistic Research, 32,* 141–166.

Inoue, A., & Fodor, J. D. (1995). Information-paced parsing of Japanese. In R. M. N. Nagai (Ed.), *Japanese sentence processing* (pp. 9–63). Hillsdale, NJ: Erlbaum.

Joshi, A. K., & Schabes, Y. (1996). Tree-adjoining grammars. In G. Rozenberg & A. Salomaa (Eds.), *Handbook of formal languages* (Vol. 3, pp. 69–123). New York: Springer-Verlag.

Joshi, A. K., & Srinivas, B. (1994). Disambiguation of super parts of speech (or supertags): Almost parsing. In *Proceedings of the 15th International Conference on Computational Linguistics* (COLING '94). Kyoto, Japan.

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

Kim, A. E. (2000). *The grammatical aspects of word recognition.* Unpublished doctoral dissertation, University of Pennsylvania.

Kim, A. E., Srinivas, B., & Trueswell, J. C. (2002). A computational model of the grammatical aspects of word recognition as supertagging. In P. Merlo & S. Stevenson (Eds.), *The lexical basis of sentence processing: Formal, computational, and experimental issues* (pp. 109–136). Amsterdam: John Benjamins.

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

McRae, K., Spivey-Knowlton, M. J., & 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.

Phillips, C., & Gibson, E. (1997). On the strength of the local attachment preference. *Journal of Psycholinguistic Research, 26,* 323–346.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1188–1200.

Rohde, D. (2002). *A connectionist model of sentence comprehension and production.* Unpublished doctoral Dissertation, Carnegie Mellon University.

Sleator, D. D., & Temperley, D. (1993, August). *Parsing English with a link grammar.* Paper presented at the Third International Workshop on Parsing Technologies, Tilburg, The Netherlands.

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

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

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

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

Sturt, P., & Crocker, M. W. (1996). Montonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes, 11,* 449–494.

Sturt, P., & Crocker, M. W. (1997). Thematic montonicity. *Journal of Psycholinguistic Research, 26,* 297–322.

Sturt, P., & Crocker, M. W. (1998). Generalized monotonicity for reanalysis models. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 365–400). Dordrecht, The Netherlands: Kluwer Academic.

Sturt, P., Pickering, M. J., & Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language, 40,* 136–150.

Tabor, W., Juliano, C., & 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,* 211–271.

Tanenhaus, M. K., Spivey-Knowlton, M. J., & 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, & J. Charles Clifton (Eds.), *Architectures and mechanisms for language processing* (pp. 90–118). Cambridge, England: Cambridge University Press.

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.

Van Dyke, J. A., & Lewis, R. L. (2003). *Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. Journal of Memory and Language, 49*(3), 285–316.

Vasishth, S. (2002). *Working memory in sentence comprehension: Processing Hindi center embeddings.* Unpublished doctoral dissertation, Ohio State University.

Vitu, F., & O'Regan, J. K. (1995). A challenge to current theories of eye movements in reading. In J. M. Findlay, R. Walker, & R. W. Kentridge (Eds.), *Oculomotor control and cognitive processes: Normal and pathological aspects* (pp. 381–393). Amsterdam: North-Holland.

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

Warner, J., & Glass, A. L. (1987). Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticality judgments of garden path sentences. *Journal of Memory and Language, 26,* 714–738.

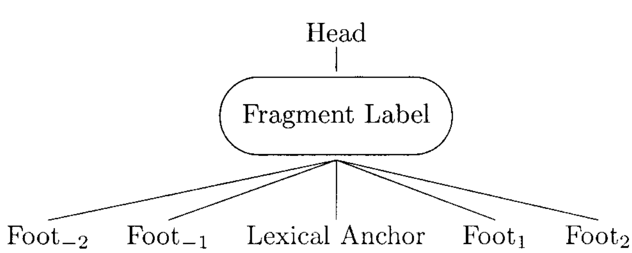
Weinberg, A. (1993). Parameters in the theory of sentence processing: Minimal commitment theory goes east. *Journal of Psycholinguistic Research, 22,* 339–364.

Weinberg, A. (1995). Licensing constraints and the theory of language processing. In R. Mazuka & N. Nagai (Eds.), *Japanese sentence processing* (pp. 235–255). Hillsdale, NJ: Erlbaum.

Appendix A

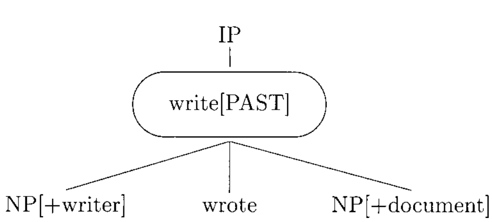
SOPARSE: A Self-Organizing Parser

An atom of SOPARSE is a lexically anchored tree fragment, as shown in A1.



(A1)

Each fragment consists of a head node, a lexical anchor, and zero or more foot nodes, which are partially ordered with respect to one another and the lexical anchor. The head and foot nodes are annotated with features that specify syntactic and semantic combination preferences, illustrated in A2.



(A2)

When a word is perceived, its tree fragments become activated. Activated tree fragments form links with other activated tree fragments. Heads link only with feet and feet only with heads, respecting linear precedence constraints (indicated by left-to-right order in A1 and A2), and the constraint that a fragment cannot form a link with itself. Each link, *i*, is associated with a link strength, q_i, which is initialized to 0, and is restricted to the interval (0, 1). Each link has a set of competitors. The competitors of a link *i* are all the activated links that have the same foot or the same head as *i*. A link's activation waxes and wanes as a function of how well it does at inhibiting its competitors. Equation A3 specifies how link activations change in time:

$$q(t+1) = q_i(t) + \eta \cdot dq_i \cdot q_i(t) \cdot (1 - q_i(t)) \qquad \text{(A3)}$$

where *t* denotes time, which is counted in discrete ticks for the purpose of implementation; η is a small positive constant that specifies the global link growth rate; and dq_i, defined below and restricted to (0, 1), is the central link modulation term. Note that $q_i(t) \cdot (1 - q_i(t))$ specifies squashed exponential growth; dq_i is the squashed, weighted sum of two terms, cdq_i and fdq_i:

$$dq_i = \tanh(\alpha \cdot cdq_i + \beta \cdot fdq_i), \qquad \text{(A4)}$$

where fdq_i encourages across-the-board growth of links when the link activations are small (see later discussion); cdq_i defines the competition between links:

$$cdq_i = \text{rating}_i - \max_{j \in \text{Competitors}_i} \text{rating}_j \qquad \text{(A5)}$$

$$\text{rating}_i = \text{viability}_i - \sum_{j \in \text{Competitors}_i} \text{viability}_j, \qquad \text{(A6)}$$

(*Appendixes continue*)

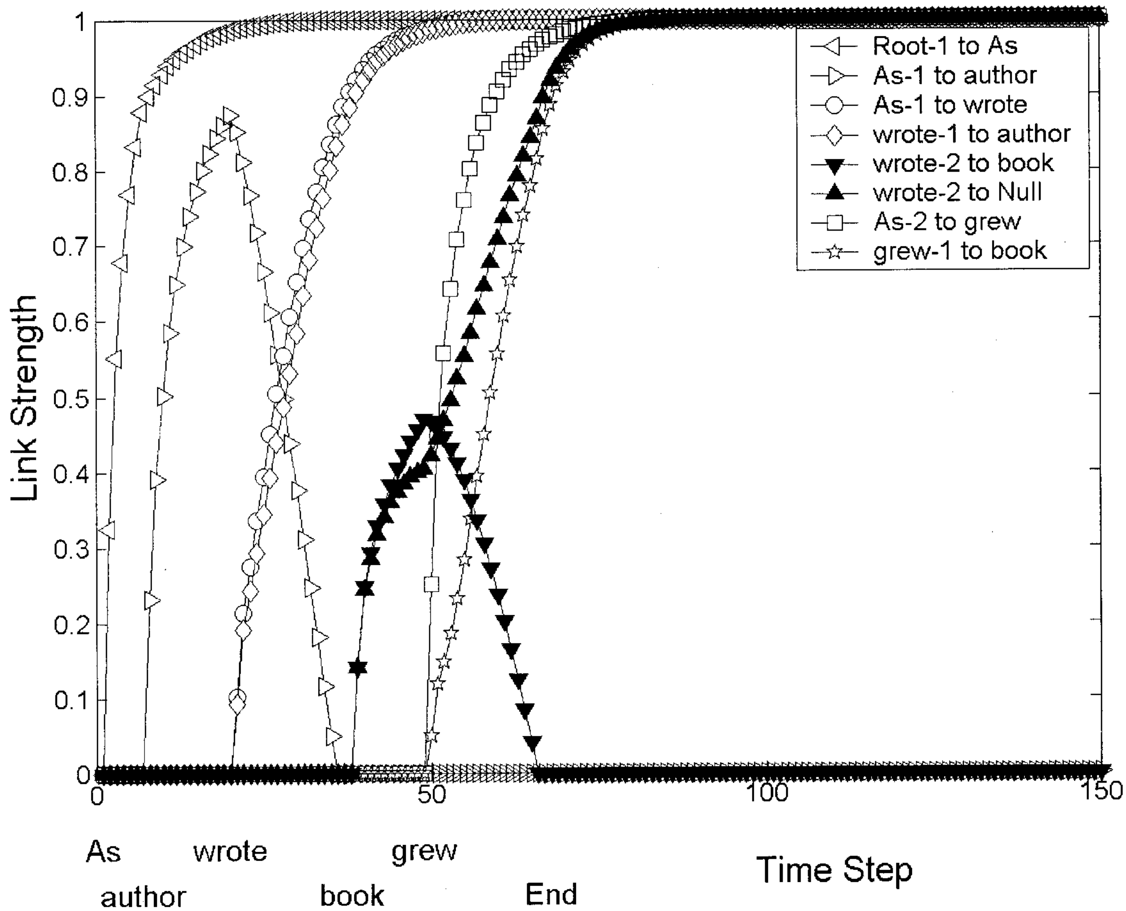


Figure A1. Growth and decline of link strengths during the processing of the intransitive/short condition of Experiment 2. $\langle \text{lexeme}_i \rangle \langle x \rangle$ to $\langle \text{lexeme}_j \rangle$ refers to the link from the xth foot of lexeme_i to the head of lexeme_j. Only links that became more than 10% active are shown.

where cdq_i is ultimately based on viability, which is a measure of how well a link is doing in absolute terms (see later discussion); rating_i compares a link's viability to the summed viabilities of its competitors. The rating Equation A6 gives rise to ganging effects, wherein a group of weak links can gang up on and overthrow a strong link. The *cdq* Equation A5 implies that, although a link's rating depends on its standing among all of its competitors, at any moment in time a link competes only with its current best competitor. A single strong link will thus keep all of its competitors at bay when they all belong to the same competitor set (e.g., other factors being equal, the model will prefer the best feature match in a case of [possibly multiple] lexical ambiguity). However, a strong link may, in principle, be overthrown if it has to engage in separate competitions with several viable competitors. This can happen if one link with similar head feature requirements has the same head as *i* but not the same foot, whereas another link with similar foot feature requirements has the same foot as *i* but not the same head. Such ganging effects play a central role in the model's recovery from garden paths (see later discussion). This is the mechanism by which self-organization gives rise to a preference for global coherence despite the absence of a supervisory mechanism.

The viability of link *i* increases with its strength (q_i), the degree of featural match between its daughter node and its mother node (match_i), and the activations of its daughter and mother fragments ($\text{act}(k)$).

$$\text{viability}_i = \left(q_i \cdot \text{match}_i \cdot \frac{\text{act}(\text{daughter}_i) + \Gamma}{1 + \Gamma} \cdot \frac{\text{act}(\text{mother}_i) + \Gamma}{1 + \Gamma} \right)^{\rho}, \quad \text{(A7)}$$

where ρ, generally set to a positive value near 0, is a constant that modulates the disparity among competitors and Γ is a constant that is used to put a positive lower bound on the contribution of fragment activation to link viability. The activation of a tree fragment, k, is a weighted sum of the activations of all the links formed by the fragment. The weighting is in proportion to the activations of the links rather than a straight average to make the function smoother:

$$\text{act}(k) = \frac{\overrightarrow{\max}\, q[k] \cdot \overrightarrow{\max}\, q[k]}{\sum_{j \in \text{Nodes}_k} \max q_j[k]}. \quad \text{(A8)}$$

Here max $q[k]$ denotes the set of highest activated links to daughter and mother nodes of fragment k; fdq_i specifies link i's portion of the "unused activation" among the pool of links with which i competes.

$$fdq_i = \frac{q_i}{\sum_{j \in \text{Competitors}_i \cup \{i\}} q_j} \cdot \left(1 - \sum_{j \in \text{Competitors}_i \cup \{i\}} q_j \right) \quad \text{(A9)}$$

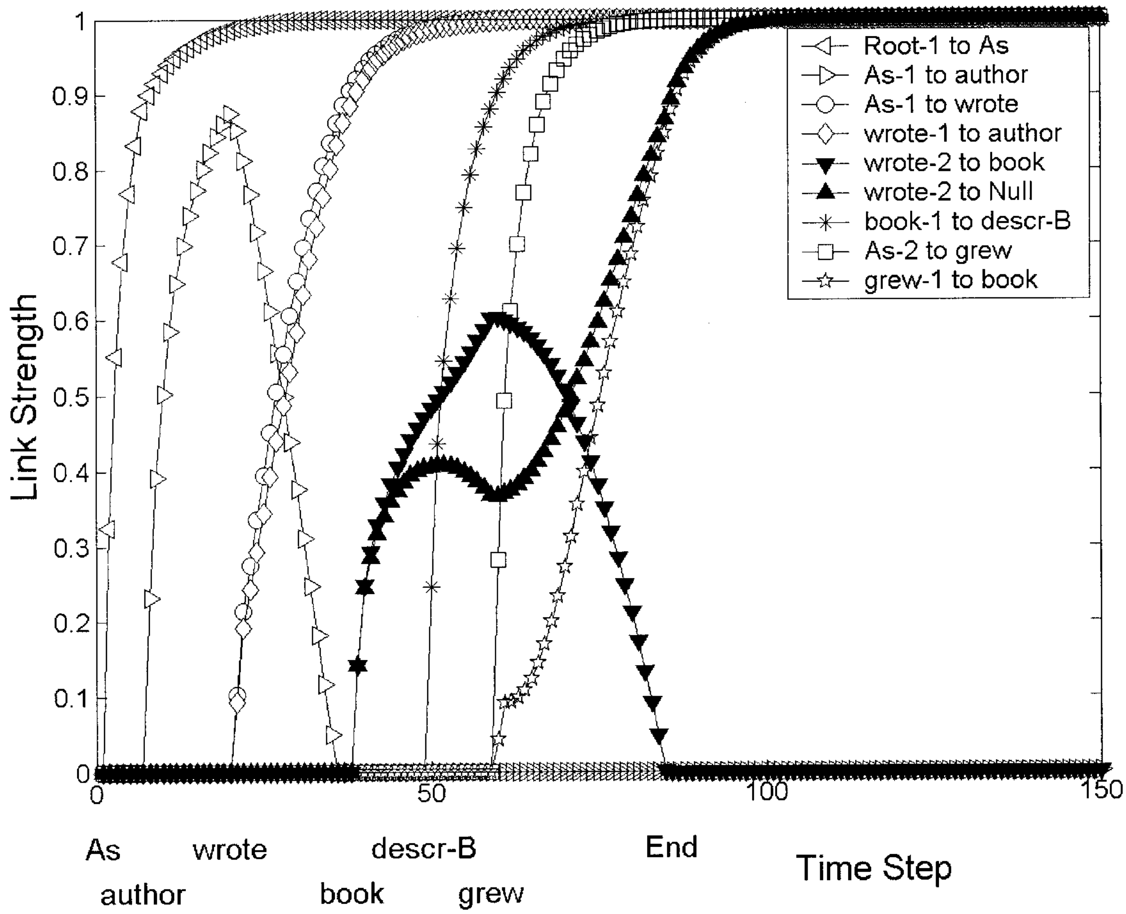


Figure A2. Growth and decline of link strengths during the processing of the intransitive/long condition of Experiment 2. ⟨lexeme$_i$⟩⟨x⟩ to ⟨lexeme$_j$⟩ refers to the link from the xth foot of lexeme$_i$ to the head of lexeme$_j$. Only links that became more than 10% active are shown.

where ∪ denotes set union. As noted, the function of the fdq_i term in Equation A4 is to spur general link activation growth when all the link activations in a competitor set are low. A new word is read when all the foot and head nodes of the previous word have become sufficiently strongly bonded (threshold θ) or a maximal number of time steps have passed (maxtime). Saturation can be accomplished either by forming a strong link with another fragment or by forming several weak links with other fragments. The feet of a fragment are opened in succession, once the previous feet have passed the bonding threshold. The parameter settings for the reported simulations are given in Equation A10.

$$[\alpha, \beta, \rho, \eta, \theta, \text{maxtime}] = [0.6, 1.5, 0.2, 0.3, 0.87, 40] \quad \text{(A10)}$$

We describe in detail how SOPARSE handles the intransitive conditions of Experiment 2.

Figure A1 provides an illustration of SOPARSE's link activations during the processing of a sentence of the type intransitive/short from Experiment 2. Determiners have been left out of the model for simplicity's sake. First, the word "as" is perceived and the link connecting "as" to a special root fragment (this instantiates the context of the sentence) grows quickly to maximum strength. When this link has stabilized, the first argument of "as" (an IP node) is opened, and the next word ("author") is simultaneously read. For lack of any better attachment, "author" forms a link to the first argument of "as," and at stabilization the next word ("wrote") is read. Because "wrote" needs an NP subject with features matching those of "author," the link between the first argument of "as" and "author" (a bad match) quickly dies off and is replaced by a link from the subject argument of "wrote" to "author" (a good match). Meanwhile, because the head type of "wrote" is IP, the first argument of "as" combines quickly with the head of "wrote." At stabilization, "wrote" opens an argument slot for a direct object, and simultaneously, the next word ("book") is read. The match between the direct object of "wrote" and the head of "book" is strong, so the link between them grows quickly. However, because "wrote" also selects for a null node (its intransitive usage), a link between "wrote" and a null node dedicated to being the direct object of "wrote" also rises in strength. At first, the attachment of the direct object of "wrote" to "book" outcompetes the null node because of a lexical bias in favor of transitive usage; however, when "grew" arrives, "grew" lays claim to "book" as well. This puts the link from the direct object of "wrote" to the head of "book" in the position of holding two separate competitions: Its foot node is in competition with the null node; its head node is in competition with "grew"; but the null node and "grew" do not compete with one another. Because the link from the direct object of "wrote" to the head of "book" has not yet become sufficiently strongly established, the competition overwhelms it and it

(*Appendixes continue*)

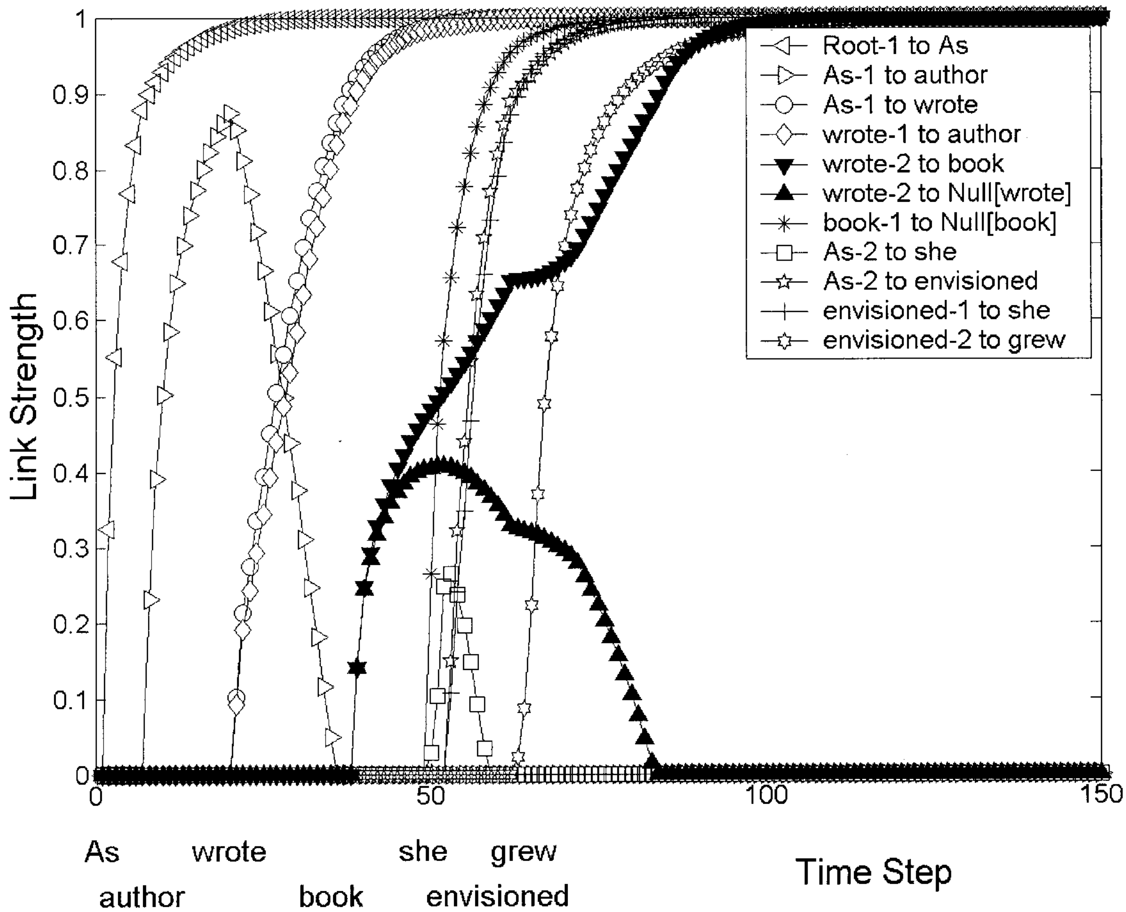


Figure A3. Growth and decline of link strengths during the processing of the two-revisions condition of Experiment 3. $\langle \text{lexeme}_i \rangle \langle x \rangle$ to $\langle \text{lexeme}_j \rangle$ refers to the link from the *x*th foot of lexeme_i to the head of lexeme_j. Only links that became more than 10% active are shown.

falls, leaving the null node to attach as the direct object argument of "wrote" and "grew" to take "book" as its subject. Thus, the model exhibits weak garden pathing (analysis and reanalysis) in the intransitive/short condition.

By contrast, in the intransitive/long condition, the link between "wrote" and "book" grows much stronger before the arrival of the disambiguating verb because of the presence of the intervening material ("describing Babylon"). As a consequence, although the model eventually establishes the correct parse, the parsing of "grew" takes considerably longer. Figure A2 shows the link strength trajectories in this case. In the two transitive conditions, by contrast, the noun "essay" fills the direct object foot of "wrote" so "book" is available to become the subject of "grew" and easily does so, leading to quicker convergence in both these conditions. Thus, SOPARSE predicts the Experiment 2 interaction.

SOPARSE responds to the two-revisions condition of Experiment 3 as follows (Figure A3). After "book" attaches as the direct object of "wrote," "she" and "envisioned" arrive[4]; "envisioned" attaches as the matrix verb, and "she" becomes its subject. In the current simulation, we implement the preference for attachment as matrix elements over attachment as elements of an object relative clause as a lexical bias of "envisioned."[5] Thus, when "grew" is perceived, the only good matches for its open-head node (an IP) and its open-subject node are taken. Because a fragment needs to form attachments in order to get activated, and it needs activation to allow its links to compete with strong competitors, "grew" fails to attach in the way that would produce a successful parse. In fact, in this case, "grew" ends up attaching as the direct object of "envisioned" (a very poor match), and its subject foot never gets filled. The result is that the criterion for reading the next word (satisfying all required attachment sites) is never met, and the model only moves on when the number of time steps has reached the bound (maxtime). In this way, SOPARSE predicts the three levels of difficulty observed in Experiment 3.

[4] We made the assumption that the function words "she" and "that" are processed much more quickly than content words. Without this assumption, the model failed to recover from the garden path in the one-revision condition.

[5] We are working on a version that lets the featural properties of the sentence-initial connective influence the choice (as indicated by the results of Phillips & Gibson, 1997, and of our Experiment 3 norming study).

Appendix B

Stimulus Items

Experiments 1 and 2

1. While the boy scratched (himself) the dog (sleeping peacefully) yawned.
2. If the clerk forgets (something) the customer (waiting below) complains.
3. As the doctor lectured (the nurse) the student (taking notes) listened.
4. While the girls raced (the boys) the kids (attending camp) watched.
5. When the worker moved (the pails) the boxes (lacking tape) fell.
6. As the author wrote (the essay) the book (describing Babylon) grew.
7. As the people watched (the screen) the show (documenting crimes) stopped.
8. Whenever the children leave (the house) the dog (barking loudly) sleeps.
9. When the aunts visit (the family) the cousins (attending prep-school) talk.
10. When the maid packs (for vacation) the bag (containing socks) breaks.
11. If the locksmith turns (the key) the knob (needing grease) sticks.
12. Because the kids tripped (the principal) the student (delivering papers) yelled.
13. When the trappers meet (the wolf) the coyote (prowling mischievously) hunts.
14. As the lawyers studied (the trial) the case (involving fraud) languished.
15. While the punks were stealing (the bike) the car (sporting tailfins) stalled.
16. While the divers searched (the ship) the river (roaring incessantly) rose.
17. As the actors rehearsed (the critical scene) the play (starring Brando) improved.
18. If the heir refuses (the fortune) the money (earning interest) doubles.
19. When the people notice (his nametag) the actor (playing Clinton) departs.
20. When the magician tries to juggle (five objects) the torches (burning kerosene) tumble.
21. When the band left (the town) the hotel (advertising entertainment) closed.
22. When the martians invaded (earth) the town (mining uranium) disappeared.
23. While the guests ate (the pie) the cake (containing walnuts) collapsed.
24. When the men hunt (the deer) the birds (singing mating-songs) scatter.
25. Wherever the officers lead (the dogs) the troops (learning wilderness skills) follow.
26. If the cheerleaders practice (their parts) the cheers (involving jumps) improve.
27. When the chauffeur parks (behind the hotel) the car (leaking oil) stalls.
28. As the driver shifted (out of second) the gears (lacking lubrication) broke.
29. Whenever the children ride (the donkey) the horse (taking medication) bucks.
30. Because the students observed (the newcomer) the instructor (demonstrating balance) paused.
31. Whenever the crew films (the river) the bear (catching salmon) leaves.
32. When the boys struck (the door) the dog (guarding the house) leapt.

Experiment 3

1. While the boy scratched (himself) the dog (that) he watched yawned loudly and rolled over.
2. When the boys were climbing (the rock wall) the rope (that) they found frayed badly and suddenly broke.
3. As the doctor lectured (the nurse) the intern (that) he advised listened carefully to each word.
4. When the worker moved (the pails) the boxes (that) he discovered fell suddenly from the shelf.
5. As the author wrote (the essay) the book (that) she envisioned grew rapidly in her mind.
6. While the artist was weaving (the basket) the rug (that) he nearly completed unraveled slowly at the other end.
7. As the people watched (the screen) the show (that) they cheered stopped abruptly in the middle.
8. Whenever the children leave (the house) the dog (that) they pester sleeps peacefully under the bench.

(*Appendixes continue*)

9. When the aunts visit (the family) the cousins (that) they support talk openly about the scandal.

10. While the maid packs (for vacation) the bag (that) she stuffs breaks apart and the clothes spill out.

11. Whenever the crew films (the river) the bear (that) they seek runs quickly into the brush.

12. Because the kids tripped (the principal) the student (that) they worried yelled loudly to summon help.

13. When the trappers meet (the wolf) the coyote (that) they fear hunts boldly on the open tundra.

14. As the lawyers studied (the trial) the case (that) they labored over languished further and got no press.

15. While the divers searched (the ship) the river (that) they sandbagged rose again and overflowed its banks.

16. As the actors rehearsed (the critical scene) the play (that) they wrote improved markedly and was soon ready to perform.

17. When the magician tries to juggle (five objects) the torches (that) he throws tumble down and scorch his feet.

18. When the band left (town) the hotel (that) they visited closed up for major renovations.

19. When the martians invaded (earth) the town (that) they discovered disappeared mysteriously and Superman found himself imprisoned in a kryptonite cave.

20. While the apprentices ate (the appetizers) the souffle (that) they discussed collapsed slowly and the ice cream melted.

21. When the men hunt (the deer) the birds (that) they like scatter quickly into the bushes and trees.

Received September 30, 2002
Revision received August 18, 2003
Accepted August 20, 2003 ■

Instructions to Authors

For Instructions to Authors, please consult the January 2004 issue of the volume or visit www.apa.org/journals/xlm and click on Submission Guidelines.