

Syntactic Ambiguity Resolution after Initial Misanalysis: The Role of Recency

Patrick Sturt

Human Communication Research Centre, Department of Psychology, University of Glasgow, Glasgow, Scotland

Christoph Scheepers

Department of Computational Linguistics, Universität des Saarlandes, Saarbrücken, Germany

and

Martin Pickering

Department of Psychology, University of Edinburgh, Edinburgh, Scotland

Although a great deal of research has investigated the factors affecting initial syntactic processing, little is known about the factors that affect processing during reanalysis. To address this question, we report a self-paced reading and an eye-tracking experiment which tested sentences in which there is initially more than one way for reanalysis to proceed, but where this choice is eventually disambiguated by a gender-marked reflexive (like *The photographers found the countess who heard the choirboy had really enjoyed himself/herself at the concert in the town hall*). The experiments provide evidence that the human sentence processing mechanism obeys a recency preference in reanalysis. This suggests that at least some of the factors guiding reanalysis are similar to those guiding initial analysis. © 2001 Elsevier Science (USA)

Key Words: sentence processing; ambiguity resolution; reanalysis; recency.

A great deal of research in sentence processing has been directed at the issues surrounding initial ambiguity resolution; for example, whether there is an initial stage in which purely structural information guides choice of analysis (e.g., Frazier, 1987) or whether all relevant information can be employed (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994). However, on

We acknowledge Oliver Garrod for his help in preparing the eye-tracking data for statistical analysis and Simon Garrod and other members of the Glasgow Psychology Department for their comments and advice. We also thank Keith Rayner, Colin Phillips, and Janet Fodor for detailed and helpful comments on a previous draft of this paper. This research was supported by ESRC Grant R000222286, a British Academy Postdoctoral Fellowship, and a British Academy research grant. The work was presented as posters at the Thirteenth Annual CUNY Conference on Human Sentence Processing, San Diego 2000, and the AMLaP conference, Leiden, Holland, 2000.

Address correspondence and reprint requests to Patrick Sturt, Human Communication Research Centre, Department of Psychology, 58 Hillhead Street, Glasgow G12 8QB, Scotland, United Kingdom. Fax: +44 141 330 4606. E-mail: sturt@cogsci.ed.ac.uk.

its own, an account of initial processing can never provide us with a complete account of syntactic processing. Initial decisions are often incorrect, and so we also need to consider the principles that govern *reanalysis* (the recovery of the correct analysis). Reanalysis has recently become an important focus of research, with attention being paid to issues such as the relative difficulty of different types of reanalysis (Ferreira & Henderson, 1991; Gorrell, 1995; Pritchett, 1992; Sturt, Pickering, & Crocker, 1999), the effect of prosodic structure on reanalysis processes (Bader, 1998; Hirose, 1999), the connections between reanalysis and the perception of ungrammaticality (Meng & Bader, 2000), and the detailed steps of grammatical inference involved in the reanalysis process itself (Fodor & Inoue, 1998). The papers included in Fodor and Ferreira (1998) provide an overview of recent work.

The present paper examines the disambiguation preferences which apply when there is more than one way for reanalysis to proceed.

We will consider the effect of one strategy on ambiguity resolution during reanalysis: recency. A recency principle predicts that, other things being equal, newly encountered words or phrases in the input are attached to recently processed material in the syntactic representation. This idea has been captured in a number of different parsing principles, such as right association (Kimball, 1973), late closure (Frazier, 1978), and more recently, in terms of the decay of activation in connectionist and hybrid networks (Altmann, van Nice, Garnham, & Hirst, 1998; Stevenson, 1994; Vosse & Kempen, 2000). As an example, consider (1), discussed by Fodor and Frazier (1980):

- (1) John bought the book that I had been trying to obtain for Susan.

The preferred interpretation involves attaching the phrase *for Susan* as a dependant of the more recently processed verb *obtain* rather than of the less recently processed verbs *trying* or *bought*. This intuition has been confirmed in experiments by Kamide (1999) (in English) and in Igoa, Carreiras, and Meseguer (1998) (in Spanish). The findings are compatible with a number of results showing evidence for a recency preference in other constructions (Altmann et al., 1998; Frazier & Rayner, 1982, Phillips & Gibson, 1997). Note that these studies involve ambiguities in which the alternative attachment sites are separated by a clause boundary, and the preferred alternative is to attach the ambiguous phrase into the most recent clause. The recency preference does not always apply in cases where the two possible attachment sites are in the same local clause. For example, recency is violated in relative clause attachment in Spanish and Dutch (Cuetos & Mitchell, 1988, Brysbaert & Mitchell, 1996) as well as in many other languages, when the alternative attachment sites are in the same local clause. In the ambiguity which we will examine in this paper, the alternative attachment sites are in separate clauses.

Most discussions of recency and other disambiguation principles have been concerned with how these principles relate to initial attachment. In (1), the sentence comes to an end without any reason to change the initial attachment of *for*

Susan to the most recent verb phrase. But initial attachments often turn out to be wrong and need to be revised (or, equivalently, *reanalyzed*). As we shall see, it is quite possible for there to be more than one way for reanalysis to proceed, in which case we need to consider how people disambiguate. There has been very little experimental work looking at how people resolve reanalysis ambiguities. The main exception to this has been work in Japanese, where the rigidly head-final phrase structure means that such ambiguities are very pervasive. For example, Hirose and Inoue (1998) and Hirose (1999) considered Japanese sentences in which a previously built main clause can be restructured as a relative clause in more than one way. Their experimental results suggest that Japanese readers resolve the ambiguity by employing systematic strategies based on the prosodic packaging of the preceding input and the semantic plausibility of alternative thematic role assignments. In the work presented in this paper, we extend this pioneering line of investigation to English and consider the role played by recency. As mentioned above, this is a well-known structural preference in disambiguating initial attachment ambiguities.

Consider (2).

- (2) The countess heard the choirboy had really enjoyed himself at the concert.

Most theories of first pass parsing predict that *the choirboy* is initially misanalyzed as the object of the preceding verb. This could be because this is the simplest attachment (Frazier, 1987; Gorrell, 1995; Pritchett, 1992) or because this is the most frequent analysis for the verb *heard* (Macdonald et al., 1994). A large body of experimental data supports this conclusion (Ferreira & Henderson, 1990; Frazier & Rayner, 1982; Garnsey, Pearlmutter, Myers, & Lotocky, 1997, Pickering & Traxler, 1998; Pickering, Traxler, & Crocker, 2000; Sturt et al., 1999). When the verb *had* is encountered, this analysis can no longer be maintained, and reanalysis is necessary, in order to reinterpret *the choirboy* as the subject of the sentence complement. What kind of parsing operation is necessary to achieve this reanalysis?

One recent claim is that at least some types of reanalysis are achieved by mechanisms which are very similar to those used in initial attachment. *Snip* (Lewis, 1993), *Attach Anyway* (Fodor & Inoue, 1988), and *Tree Lowering* (Sturt & Crocker, 1996) are all parsing operations which are used for reanalysis, but they can also be seen as attachment operations; they all involve attaching a new word in the input to an incrementally expanding syntactic representation. For example, in Sturt and Crocker (1996), the *tree-lowering* operation is used to attach the word *had* in (2) to the tree. This involves *the choirboy* being reattached as the subject of *had*, and the sentence whose auxiliary verb is *had* is attached as the complement of *heard*. It can be seen that, although it is used for reanalysis, tree lowering is actually a parsing operation that is used to attach a word, in this case *had*, and it is comparable to the types of parsing operation that are used in initial attachment, although it is more complex. This suggests that such operations should be subject to the same attachment principles that govern the resolution of initial attachment ambiguities. See Fodor and Inoue (2000) for more discussion on these issues.

Now consider (3):

- (3) The photographers found the countess who heard the choirboy had really enjoyed himself/herself at the concert.

Again, we assume that *the countess . . .* is initially attached as the object of the preceding verb *found*, either because *found* most frequently takes a direct object argument or because this analysis is syntactically simplest. As in (2) reanalysis is also forced on the input of *had*, but unlike in (2) there are two ways in which reanalysis can proceed: Either the recent noun phrase (*the choirboy*) or the distant noun phrase (*the countess . . .*) can be reanalyzed as the subject of *had*. The recent (or *low*) reanalysis is compatible with the *himself* continuation, while the distant (or *high*) reanalysis is compatible with the *herself* continuation. If initial attachment preferences apply to ambiguity resolution during reanalysis, then the recent (low) attachment should be preferred. This is the prediction of Lewis (1993), Fodor and Inoue

(1998), and Sturt and Crocker (1996), because in all these models, the reanalysis necessary in (3) is accomplished by an operation which resembles initial attachment, and initial attachment is subject to recency.

We can contrast (3) with (4), a similar sentence type examined in previous studies conducted by Sturt, Pickering, Scheepers, and Crocker (2001) and Schneider and Phillips (2001):

- (4) The countess who heard the choirboy had really enjoyed himself/herself at the concert (and) organized a charity event afterward.

In (4), *had* can be attached either high or low, as in (3). High attachment is compatible with *herself* and the conjunction *and*, while low attachment is compatible with *himself* and the absence of *and*. The crucial difference between (3) and (4) is that in (4), reanalysis is optional at the point where *had* is reached, while in (3), reanalysis is forced. In (4), the high attachment of *had* allows the processor to avoid reanalysis. This is because *the countess* is not initially attached as a dependant of any verb in the syntactic representation, so no reanalysis is necessary if this noun phrase is attached as the subject of *had*. On the other hand, the recent noun phrase *the choirboy* is initially attached to the verb *heard*, so reanalysis would be necessary if *had* took this recent noun phrase as its subject. Sturt et al. (2001) and Schneider and Phillips (2001) show evidence for a high attachment preference in sentences like (4), demonstrating that recency does not apply in such cases. These results are compatible with the claim that the processor follows a reanalysis avoidance principle, such as *Revision as Last Resort* (Fodor & Frazier, 1980). One implementation of this strategy is given in Sturt and Crocker (1996), where the processor attempts to apply simple attachment operations which do not involve reanalysis before more complex attachment operations that do.

In contrast to the high attachment preference found for sentences like (4), the model predicts that there should be a low attachment preference for (3), because reanalysis is unavoidable, and the parsing operation that is required is subject to recency. A questionnaire experiment reported

in Sturt (1997) shows that there is a recency preference for globally ambiguous versions of sentences like (3). However, there is currently no evidence for such a preference being applied on-line. The experiments reported in this paper are designed to test the attachment preference for sentences like (3), specifically in order to determine whether recency applies to reanalysis in on-line sentence processing.

EXPERIMENT 1

Experiment 1 was a self-paced reading experiment and employed sentences similar to (3) above, using reflexive pronouns to disambiguate toward either high (nonrecent) or low (recent) attachment. High and low ambiguous conditions were compared with control conditions which used the complementizer *that*:

- (5) a. High ambiguous: The photographers found the countess who heard the choirboy had really enjoyed herself at the concert in the town hall.
 b. High control: The photographers found that the countess who heard the choirboy had really enjoyed herself at the concert in the town hall.

- c. Low ambiguous: The photographers found the countess who heard the choirboy had really enjoyed himself at the concert in the town hall.
 d. Low control: The photographers found the countess who heard that the choirboy had really enjoyed himself at the concert in the town hall.

Before we consider the patterns of reanalysis that may occur with these sentences, we will outline our assumptions for the first-pass attachments up to the point where *the choirboy* is attached. These are schematically represented in Fig. 1.

The first verb (in this case, *found*) will be referred to as the *high verb*, and the second verb (in this case, *heard*) will be referred to as the *low verb*. The noun phrases immediately following these verbs will be referred to as the high and low noun phrases, respectively. In the two ambiguous conditions, we assume that the high and low noun phrases are initially attached as objects of their respective verbs. This initial object attachment will be crucial, because we are interested in how people choose which of these initial attachments to break during subsequent

Ambiguous conditions:

The photographers found [_{NP} the countess who heard [_{NP} the choirboy.....

**Attachment as
object (assumed)**

**Attachment as
object (assumed)**

Low control:

The photographers found [_{NP} the countess who heard that [_S [_{NP} the choirboy.....

**Attachment as
object (assumed)**

**Attachment into
complement clause
(grammatically forced)**

High control:

The photographers found that [_S [_{NP} the countess who heard [_{NP} the choirboy.....

**Attachment into
complement clause
(grammatically forced)**

**Attachment as
object (assumed)**

FIG. 1. Assumed and grammatically forced first-pass attachments for the ambiguous conditions and the two control conditions.

reanalysis. As can be seen below, we attempted to encourage this object attachment by using verbs which were statistically biased toward the noun phrase object reading. In the control conditions, the complementizer *that* forces a sentential complement analysis of the high verb in the high condition and the low verb in the low condition, thus preventing the attachment of the following noun phrases as objects of each of these respective verbs. As in the ambiguous conditions, we assume that the object attachment is made where no complementizer appears, so that the high noun phrase is attached as an object in the low control condition and vice versa in the high control condition.

We now consider the processing patterns that we expect when the words *had really enjoyed* are processed. Given the assumptions outlined above about initial attachment, reanalysis will be necessary in the ambiguous conditions, because *had* will need to take either the high or the low noun phrase as its subject, but both of these have already been attached as objects of their respective verbs. In the control conditions, however, no reanalysis is necessary, as in each case, the complementizer has prevented an initial misattachment from being made. Note that in the high control condition, it is also grammatically possible for reanalysis to occur here. In other words, it is possible for *had* to be attached low, with *the choirboy* being reanalyzed as its subject (as in *The photographers found that the countess who heard the choirboy had really enjoyed himself was a very unpleasant woman*). We assume that this low attachment involving reanalysis is either not attempted or is at least strongly dispreferred and attempted only rarely. This claim is supported by the experiments reported in Sturt et al. (2001) and Schneider and Phillips (2001), which show consistent evidence for a high attachment preference in circumstances where the alternative low attachment would involve reanalysis, for a similar ambiguity (see example (4) discussed above). In summary, at *had really enjoyed*, we do not expect reanalysis to occur in either of the control conditions, but we do expect it to occur in both of the ambiguous conditions. This should lead to longer reading times in the ambiguous condi-

tions than the control conditions around this point in the sentence.

Reading times for the reflexive pronoun (and possibly the following words) should indicate whether high or low reanalysis was performed at the preceding region in the ambiguous conditions. If there is a preference for recency in reanalysis, then the low ambiguous condition should be relatively easy to process, because the gender of the reflexive would be consistent with this analysis. In the high ambiguous condition, on the other hand, the gender marking would not be consistent with the low reanalysis, and the processor would have to revise its analysis again (i.e., it would have to perform re-reanalysis). Therefore, the high ambiguous condition should cause more processing difficulty than the low ambiguous condition around this point in the sentence. On the other hand, both control conditions should be relatively easy to process here, leading to an interaction.

The self-paced reading methodology allows the presentation of long lines, unlike the eye-tracking technique which we use in our laboratory (see Experiment 2). In Experiment 1, we were able to position the line break after the critical reflexive pronoun, thus avoiding any influence of the line break position on the resolution of the ambiguity being tested (Kennedy, Murray, Jennings, & Reid, 1989). Self-paced reading also makes it possible to compare the results of this experiment directly with those in Sturt et al. (2001) and Schneider and Phillips (2001), which used a similar technique. Of course, self-paced reading also has some disadvantages in comparison with eye tracking, and for this reason, we also conducted an eye-tracking experiment (see Experiment 2).

Method

Participants

Thirty-two native English speakers from the Glasgow University community were paid to take part. The data for one further participant were discarded because of a low comprehension score (less than 70% correct for comprehension questions following experimental items and fillers).

Items

Twenty-eight items like (5) were selected from a larger set on the basis of a plausibility pretest (see below). The selected items are given in the Appendix. Each item included two critical verbs, both of which could take either a noun phrase or a tensed clause as a complement. For example, in (5) the critical verbs are *found* in the high position and *heard* in the low position. Fourteen critical verbs were used in the experiment. They were arranged in pairs so that for each item with high verb *X* and low verb *Y*, there was a corresponding item with high verb *Y* and low verb *X*. This procedure prevents any bias due to differences in lexical preferences of the verbs. No verb pair appeared in more than four items. The verbs were all biased toward the noun phrase direct object analysis. These biases were determined on the basis of the corpus study described in Sturt et al. (1999) and Sturt et al. (2001), and the criterion for calculating the biases was the same. Random samples of sentences containing the critical verbs were obtained from the British National Corpus, and sentences were classified as NP if the critical verb took a noun phrase direct object and as S if it took a tensed clause without a complementizer. The noun phrase bias for each verb was simply the proportion of sentences classified as NP in the sample, relative to the total of NP and S sentences in that sample. This proportion was .65 or greater for all verbs used in the experiment. All the items were disambiguated with either *himself* or *herself*. The gender of the reflexive pronoun was manipulated to create the high and the low readings. In the 28 items selected for the experiment, *himself* was used in 13 items to disambiguate high and 15 items to disambiguate low (and *vice versa* for *herself*).¹

Plausibility pretest. Four sentences were constructed for each item. Two of these sentences compared the plausibility of the (initial) noun phrase analysis for the low and high verbs. The other two sentences compared the plausibility of the high and low attachments. For (5), the four sentences were:

Low noun phrase analysis: The countess heard the choirboy.

High noun phrase analysis: The photographers found the countess.

Low attachment: The choirboy had really enjoyed himself at the concert.

High attachment: The countess had really enjoyed herself at the concert.

We constructed booklets which each contained the four sentences for 32 candidate items randomly interspersed with 50 filler sentences of varying plausibility, using four fixed random orders. Seventeen participants from the University of Glasgow community were paid to rate the plausibility of each sentence on a scale from 1 (least plausible) to 7 (most plausible). We then selected 28 items which maintained the above-mentioned counterbalancing of verb positions, but which also combined the highest overall plausibility with the smallest pairwise differences in plausibility between the Low and High noun phrase analysis sentences and between the Low and High attachment sentences. We analyzed the ratings using the Wilcoxon matched-pairs signed-ranks test, aggregating by participants (Z_1) and items (Z_2).² For the 28 selected sentences, there were no significant differences between the two NP analysis conditions (high, 6.28; low, 6.37; $|Z_1|, |Z_2| < 1.5$, ns), nor between the two attachment conditions (high, 5.85; low, 5.92; $|Z_1|, |Z_2| < 1.1$, ns).

Procedure

As in Sturt et al. (2001), the experiment employed the noncumulative segment-by-segment self-paced reading technique, implemented using the Psyscope experimental package (Cohen, MacWhinney, Flatt, & Provost, 1993). The experiment was run on a Macintosh-compatible computer, which was connected to a button box. Each sentence was divided into nine segments as indicated below, where we use a

¹ It was not possible to achieve a completely even balance in the gender of disambiguation, given that we also needed to balance verb position and plausibility.

² Parametric tests (*t* tests) yielded statistically identical results.

single slash to indicate a segment boundary. A double slash indicates where there was also a line break. The segmentation was similar to that used by Sturt et al. (2001).

The photographers / found (that) / the countess / who heard (that) / the choirboy / had really enjoyed / himself / at the concert // in the town hall.

We defined the regions as follows:

1. The initial noun phrase.
2. The first verb plus the complementizer *that* in the high control condition.
3. The first part of high noun phrase, from the initial determiner to the head noun, inclusive.
4. The relative pronoun and the second verb plus the complementizer *that* in the low control condition.
5. The noun phrase following the second verb (*precritical*).
6. The first three words of the final verb phrase (*first critical region*).
7. The disambiguating reflexive pronoun (*reflexive region*).
8. A (usually prepositional) phrase (*spillover region*).
9. Another (usually prepositional) phrase (*final region*).

The items were divided into four lists, so that each list included exactly one version of each item, and there were equal numbers of items per condition in each list. Each experimental list was combined with 84 fillers of varying syntactic constructions, and comparable segmentation, some of which included the word *that* and reflexive pronouns.

The participant sat in front of the computer monitor and was instructed to press the middle button of the Pyscope button box with his or her dominant hand, in order to see each segment of the sentence. Each button press revealed the next segment, and the viewing time between each button press was recorded. All segments

except the segment currently under view appeared as a series of underscore characters, with spaces corresponding to the actual spaces in the text. The trial began with an asterisk on the left of the screen marking the position of the first character of the sentence. The participant pressed the button to see the series of underscore characters for the entire sentence, and after the button was pressed again, the first segment appeared. A comprehension question (e.g., *Who had a good time at the concert?*) followed each trial. The questions were followed on the next line by two possible answers (e.g., *the choirboy*, *the countess*), displayed on the left and the right of the screen, and the participant had to select the correct answer using either the left or the right button (with the correct answer counterbalanced across trials). For the experimental sentences, half of these questions directly probed the attachment under investigation; in the other half, the question could be answered without the attachment having been made successfully. In this way, we attempted to maintain participants' attention to the (rather complex) sentences, without inducing strategies by over-highlighting the critical ambiguity. The order of the trials was randomized separately for each participant, with the constraint that no two experimental items appeared adjacent to each other. Five practice trials were presented initially to familiarize the participants with the experimental procedure, and the experiment proper began with at least 10 filler sentences.

Results and Discussion

All analyses were conducted on residual reading times, calculated by performing a simple linear regression predicting reading time from region length (in terms of number of characters) for each participant. The predicted reading time was subtracted from the actual reading time of each data point in the experiment (Ferreira & Clifton, 1986; Trueswell, Tanenhaus, & Garnsey, 1994). Thus a positive value indicates a reading time which is longer than what would be predicted by its region length, and a negative value indicates a reading time which is shorter than what would be predicted by its length.

TABLE 1
Trimmed Residual (and Untrimmed Raw) Reading Times (ms) for Experiment 1

Region	Precritical The choirboy	First critical had really enjoyed	Reflexive himself	Spillover at the concert	Final in the town hall
High/ambiguous	87 (729)	147 (878)	63 (617)	109 (698)	64 (769)
High/control	77 (717)	91 (820)	16 (540)	36 (585)	67 (753)
Low/ambiguous	83 (700)	156 (905)	68 (596)	29 (565)	42 (756)
Low/control	120 (736)	90 (820)	40 (569)	14 (539)	10 (696)

For the purposes of calculating the regression equations, we did not include data points that were likely to include large variance due to garden pathing or wrap-up effects. Thus we excluded times from the first region, the last region, and the reflexive region in the experimental items. We also excluded from the regression equations any raw reading time greater than 2000 or less than 150 ms (see Sturt et al., 2001).

As in Sturt et al. (2001), we trimmed the residual data according to the definition of extreme values in the SPSS function *explore*; for each region \times condition cell, we calculated the upper and lower quartiles of the distribution of residual reading times. Cutoff values were set at $3 \times$ the interquartile range above the upper quartile and $3 \times$ the interquartile range below

the lower quartile, and any data points above or below these cutoff values were replaced by the relevant cutoff value. This procedure affected 3.5% of the data.

The resulting data for each region were submitted to analyses of variance, treating ambiguity (ambiguous vs. control) and attachment site (high vs. low) as within-participants and within-items factors. Analyses were performed on aggregated data for both participants (F_1) and items (F_2). We report analyses from the precritical region, which immediately precedes the first theoretically interesting region.

Table 1 gives reading times for all five analysis regions, and Fig. 2 shows this information graphically from the first critical to spillover regions.

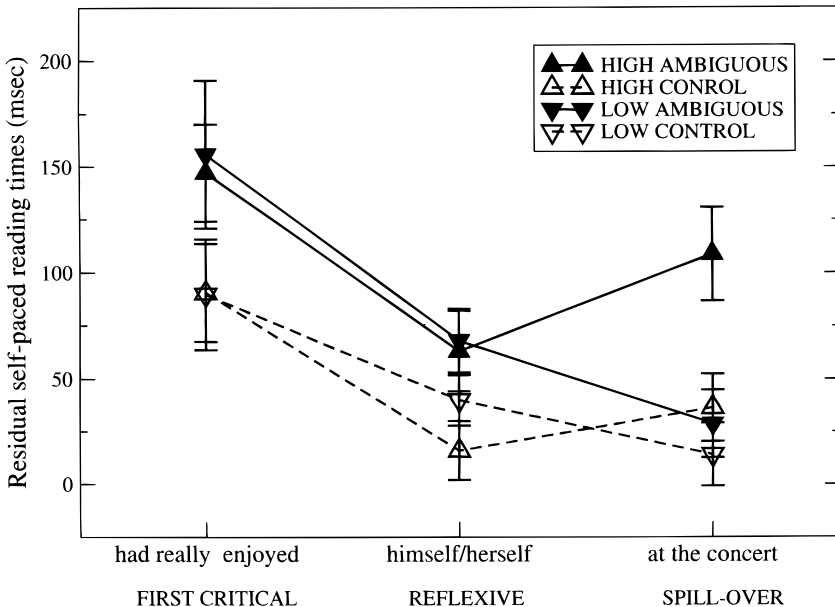


FIG. 2. Trimmed residual reading times with standard errors for Experiment 1: critical to spillover regions.

In the precritical region, no effects approached significance (all p 's > .1). In the first critical region, a main effect of ambiguity confirmed the expected garden path effect in this region, with ambiguous sentences being read more slowly than their controls ($F_1(1,31) = 8.99, p < .01; F_2(1,27) = 8.29, p < .01$). There was no effect of attachment or interaction between the two factors (F 's < 1). The same pattern showed up at the reflexive region (ambiguity: $F_1(1,31) = 6.75, p < .05; F_2(1,27) = 6.22, p < .05$; all other effects, F 's < 1). Thus there was no evidence for an interaction between ambiguity and attachment at the disambiguating reflexive. However, the expected interaction occurred in the spillover region, though marginal in the items analysis ($F_1(1,27) = 5.11, p < .05; F_2(1,31) = 3.85, p = .06$). The interaction was accompanied by main effects of both ambiguity ($F_1(1,31) = 7.88, p < .01; F_2(1,27) = 9.40, p < .01$) and attachment ($F_1(1,31) = 9.78, p < .01; F_2(1,27) = 6.01, p < .05$). Contrast analyses demonstrated that the high attached sentences were read more slowly than the low attached sentences in the ambiguous conditions ($F_1(1,31) = 9.25, p < .05; F_2(1,27) = 6.78, p < .05$) but not in the control conditions (both p 's > .09). The high ambiguous condition was read more slowly than its control ($F_1(1,31) = 10.62, p < .01; F_2(1,27) = 9.62, p < .01$) but the low conditions were read equally quickly (both F 's < 1). In the final region, there was a marginal trend toward a main effect of attachment, with high attached sentences being read more slowly than low attached sentences ($F_1(1,31) = 3.00, p < .1; F_2(1,27) = 3.98, p < .06$). There were no other significant effects in this region (all F 's < 1.1).

To summarize, the garden path effect at the first critical region confirmed that an initial misanalysis had been made in the ambiguous conditions, and therefore reanalysis was necessary. The interaction effect at the spillover region following the reflexive indicates that there was a preference to perform this reanalysis at the low attachment site. Hence the results demonstrate a recency effect in reanalysis ambiguity resolution. The presence of the interaction at the spillover region rather than the critical reflexive

region is fairly typical of self-paced reading data and is consistent with earlier results (Sturt et al.; 2001, Experiment 3) which show that such delays systematically occur when a disambiguating reflexive is embedded in a long sentence. Note that, although the interaction appears only in the spillover region, the main effect of ambiguity at the reflexive region still needs to be explained. There are at least two possible accounts. On one account, the gender information associated with the reflexive does not affect reading times until the following region, in which case, the ambiguity effect at the reflexive could itself be a spillover effect from the first critical region, where a similar ambiguity effect was found. On the other hand, the results are also compatible with an immediate disambiguation model, if we assume that disambiguation preferences in reanalysis are probabilistic. On this account, low reanalysis occurs in a majority of trials, but high reanalysis also occurs occasionally. The trials on which high reanalysis occurs would elevate reading times for the low ambiguous condition at the reflexive, weakening any interaction effects and contributing to the main effect of ambiguity. From the results of Experiment 1, it is impossible to determine which of these two accounts is correct. We return to this issue in the discussion of Experiment 2.

Comprehension Accuracy

Comprehension accuracy for the experimental trials which included questions was as follows: high ambiguous, 82%; high control, 89%; low ambiguous, 91%; low control, 95%. ANOVAs confirmed a main effect of ambiguity ($F_1(1,31) = 7.90, p < .01; F_2(1,27) = 6.04, p < .05$) and attachment ($F_1(1,31) = 19.08, p < .001; F_2(1,27) = 8.86, p < .01$). The tendency was for comprehension accuracy to be reduced in the high attached conditions and the ambiguous conditions, in comparison with the low attached conditions and the control conditions. There was no interaction between the two factors (both F 's < 1). Thus, independently of one another, being garden pathed and making distant attachments had a detrimental effect on participants' comprehension accuracy.

EXPERIMENT 2

Experiment 2 used eye-tracking on items adapted from those used in Experiment 1. The main aim of this experiment was to test whether the results of Experiment 1 could be replicated with a different methodology. Self-paced reading is more flexible with regard to line length, but also has some disadvantages in relation to eye tracking. First, our self-paced reading procedure involved the artificial segmentation of the sentences into smaller fragments, which might have supported the low attachment analysis. Such a segmentation bias is unlikely, however, as the segmentation in Experiment 1 was comparable to that used in Sturt et al. (2001) for sentences like (4) discussed above, but in those experiments, a *high* attachment preference was obtained. A potentially more serious concern is the unavailability of previously read text for reinspection. Although the noncumulative version of self-paced reading, as used in Experiment 1, is generally accepted to be more sensitive than its cumulative counterpart (Just, Carpenter, & Wooley, 1982), it prevents reinspections. As regressive eye movements are known to be a feature of reanalysis (Frazier & Rayner, 1982), the participants' inability to make regressions in Experiment 1 could have artificially affected their reanalysis ambiguity preferences. For example, during noncumulative self-paced reading, participants might rely on strategically stored information to compensate for the unavailability of previously read text in the visual field. The recency preference of Experiment 1 could be explained if we assume that more recently stored information is more readily available during reanalysis.

The eye-tracking procedure allows us to eliminate the possibility that the preference found in Experiment 1 is due to the unavailability of previous critical regions in the visual field or to segmentation bias. The one drawback with the eye-tracking procedure used in our laboratory is the stricter limitation on the number of characters that can appear within a single line. This means that, unlike in Experiment 1, we were forced to break the line before the appearance of the critical reflexive region, which itself could have caused segmentation biases. However, we

controlled for any biases by counterbalancing the line-break position across items (see below).

Method

Participants

Thirty-two participants from the Glasgow University community were paid to take part in the experiment. The data for an additional participant were discarded because of very slow reading times and a large number of fixations per trial (average 60).

Items

The items were based on those from Experiment 1, but they were converted into short texts by adding a second sentence. The example sentence given above was continued as follows:

- (6) The photographers found the countess who heard the choirboy had really enjoyed himself at the concert in the town hall. The whole evening had been a great success.

The second sentence served to add more distance between the critical region and the end of the trial. This allowed us to analyze the data for the critical sentence without contamination from noisy events associated with the end of a trial, such as wholesale rereading of the stimulus, anticipatory eye movements for reading the comprehension question, or the planning of a button press. The second sentence also made the text more natural and interpretable. To reduce the numbers of fixations close to the right edge of the screen (which can sometimes cause track loss), two of the items were shortened by modifying the last two regions of the first sentence (see Appendix for details).

Procedure

The experiment was run using a Generation 5.5 Fourward Technologies Dual Purkinje Image eye tracker. The tracker has angular resolution of 10' of arc. The tracker monitored only the right eye's gaze location, but viewing was binocular. A PC displayed items on a VDU 70 cm from the participants' eyes. The VDU displayed about four characters per degree of visual angle. The tracker monitored participants'

gaze location every millisecond, and the software sampled the tracker's output to establish the positions of eye fixations and their start and finish times.

Before the experiment started, the participant sat at the eye tracker, and three practice trials were presented to familiarize the participant with the experimental procedure. The experimenter then used bite bars and forehead restraints to immobilize the participant's head. Before each trial, the experimenter had the participant fixate a series of squares at various positions on the screen, to test accuracy of calibration. If this was inaccurate, the eye tracker was recalibrated. The last square fixated in this sequence was in the same position as the first character of the text. When the participant fixated this square, the experimenter pressed a button and the text was displayed. The participant pressed a button when he/she had finished reading the text and then answered a question (as in Experiment 1) using the left or right buttons.

We chose two positions for the line break and quasi-randomly assigned half of the items to each line break position. In the early line break texts, the line break in the first (critical) sentence appeared immediately after the low verb. In the late line break texts, the break in the first sentence appeared immediately after the low noun phrase. In the early line break texts, the second sentence began at the start of the third line. In the late line break texts, the first two or three words of the second sentence appeared on the second line, and the remainder appeared on the third line. The positions of the line breaks (marked with double slashes) are illustrated in the following example.

Early line break: The photographers found the countess who heard//the choirboy had really enjoyed himself at the concert in the town hall.//The whole evening had been a great success.

Late line break: The photographers found the countess who heard the choirboy//had really enjoyed himself at the concert in the town hall. The whole//evening had been a great success.

To aid subsequent analysis of fixation data, two blank lines were inserted between each line of text.

As in Experiment 1, four lists of items were created. Twenty-eight filler texts (adapted from the filler sentences used in Experiment 1) were added to the lists, as well as 27 texts from an unrelated experiment on metaphor comprehension. (Example item: *Chris's day was not going too well. The windows in his office wouldn't open. It was bothering him. The heat was rising. He looked forward to tomorrow.*) The fillers and experimental items from both studies all consisted of at least two sentences. The lists were randomized separately for each participant, with the constraint that no two experimental items from the same study appeared adjacent to each other. The experiment began with at least four fillers.

Data Analysis

An automatic procedure pooled short contiguous fixations. The procedure incorporated fixations of less than 80 ms into larger fixations within one character and then deleted any remaining fixations of less than 80 ms. Readers do not extract much information during such short fixations (Rayner & Pollatsek, 1989). Before analyzing the eye-movement data, we eliminated seven trials where the participant failed to read the sentence, or where tracker loss resulted in a serious loss of data (defined as trials in which the first-pass reading time for a sequence of three or more regions was zero). We will report results for three different eye-movement measures. *First-Pass Reading Times* are the sum of all fixations in a region between the time when the reader first enters the region from the left to the time when the region is first exited to either the right or left. *Regression Path Times* (Brysbaert & Mitchell, 1996; Duffy, Morris, & Rayner, 1988; Konieczny, 1996, Konieczny, Hemforth, Scheepers, & Strube, 1997; Liversedge, 1994; Traxler, Pickering, & Clifton, 1998) are the sum of fixations from the time when the reader first enters the region from the left to the time when the region is first exited to the right. In both of these measures, if the region is never fixated, or if it is only fixated after later regions, the data point is excluded from the

analysis and not counted as zero. Further analyses including these data as zero times showed a very similar pattern of results. Note that regression path times always correspond to first-pass reading times if the region is first exited to the right. However, regression path times differ from first-pass reading times if the first exit from the region is a regression. First-pass reading times and regression path times will be referred to as *first-pass measures*. In addition, *Second-Pass Reading Times* are the sum of fixations made on a region after subsequent regions have already been fixated. They do not include any time spent rereading the region before the reader has gone past it. For second-pass reading times, zero times were included in the analysis. Note that second-pass times can include fixations that occur during regressions from the second sentence.

For the purposes of analysis, the items were divided into 10 regions. These regions followed the self-paced reading segmentation of Experiment 1, except that the first critical region, which immediately preceded the disambiguating reflexive in Experiment 1, was divided into two regions for Experiment 2. The first of these two regions always consisted of two words, and will continue to be referred to as the *first critical region*, and the second of these regions always consisted of one word and will be referred to as the *filler region*. Thus the regions were as follows:

1. The photographers
2. found (that)
3. the countess
4. who heard (that)
5. the choirboy (*precritical region*)
6. had really (*first critical region*)
7. enjoyed (*filler region*)
8. himself (*reflexive region*)
9. at the concert (*spillover region*)
10. in the town hall (*final region*)

Residual reading times were computed and data trimming was performed as in Experiment

1. The residual reading time transformation was calculated only for the first-pass times and not for the regression path times or second-pass reading times. Regression path times often include regressive fixations outside the region of interest, where the length of that region cannot be expected to play a major role. Second path times consist solely of rereading, and so the role of word length is also unclear: our regression analyses indicated that second-pass reading time hardly correlates with length ($r^2 = .04$), whereas length accounts for a substantial proportion of the variance in the first-pass reading times ($r^2 = .32$). Data trimming affected no more than 1.2% of the data in any given measure. For the regression path time data, the trimming procedure was conducted separately for trials on which a first-pass regression was made. This is because the regression path times for these trials are often considerably longer than nonregression trials, and a very large number of them would have counted as extreme values if the trimming procedure had been performed on all the data simultaneously.

Results and Discussion

Reading times calculated according to the three measures described above were entered into ANOVAs. As before, the design of the ANOVA included ambiguity and attachment as repeated measures factors. As an initial check to ensure that our results were not dependent on the position of the line break, we also performed ANOVAs including line break position (Early vs. Late Line Break) as a within-participants and between-items factor. However, the pattern of results was very similar across the two line break conditions, with no significant interactions involving line break in any of the critical regions. Thus analyses including this factor are not reported.

First-Pass Measures

Means for the first-pass measures are given in Tables 2 and 3. In addition, Fig. 3 shows the regression path times from the first critical to spillover regions in graphical format.

TABLE 2

Trimmed Residual (and Untrimmed Raw) First-Pass Reading Times (ms) for Experiment 2

Region	Precritical The choirboy	First critical had really	Filler enjoyed	Reflexive himself	Spillover at the concert	Final in the town hall
High/ambiguous	86 (560)	27 (439)	-57 (298)	-49 (293)	4 (356)	-13 (545)
High/control	81 (552)	16 (432)	-43 (313)	-51 (289)	-1 (357)	-24 (534)
Low/ambiguous	60 (535)	38 (450)	-30 (328)	-49 (285)	1 (359)	-21 (529)
Low/control	12 (483)	-6 (403)	-59 (299)	-70 (264)	-17 (338)	-54 (504)

In the precritical region (*the choirboy* in (6)), first-pass reading times revealed an effect of attachment which was marginal in the participants analysis ($F_1(1,31) = 4.13, p < .06$; $F_2(1,27) = 7.46, p < .05$), with low attached sentences being read more quickly than high attached sentences. There were no other significant effects in the first-pass reading times in this region (all p 's $> .15$). Note that, despite the lack of an interaction, the attachment effect could only have been driven by the control conditions, as these are the only conditions to differ at this point in the sentence. Contrast analyses support this conclusion; high attached sentences took marginally longer to read than low attached sentences in the control conditions ($F_1(1,31) = 3.37, p < .08$; $F_2(1,27) = 8.31, p < .01$), but attachment had no effect in the ambiguous conditions (both F 's < 1.7). These results could have been due to differences in the immediately preceding context, leading to different eye-movement profiles among the four conditions, perhaps due to differences in cloze probability. In the low control condition, which was read relatively quickly, the region immediately follows a complementizer, while in the other three conditions, which were read relatively slowly, the region immediately follows the low verb. There was, however, no evidence for a low attachment advantage in the

regression path times; if anything, there was a trend in the opposite direction, with a nonsignificant tendency for low attached conditions to take longer to read than high attached conditions ($F_1(1,31) = 2.90, p < .1, F_2(1,27) = 3.53, p < .08$). No other effects approached significance in the regression path times (all F 's < 1.7).

The first critical region (*had really*) is the region where reanalysis is initially necessary in the ambiguous conditions. A main effect of ambiguity suggests that this reanalysis took place, with ambiguous sentences taking longer to read than control sentences (first-pass RT, $F_1(1,31) = 6.55, p < .05$; $F_2(1,27) = 4.72, p < .05$; regression path, $F_1(1,31) = 5.99, p < .05$; $F_2(1,27) = 13.07, p < .01$). Attachment site had no effect in either of the first-pass measures (all p 's $> .1$) and the two factors did not interact significantly (all F 's < 1.8).

The only effect to reach significance in the first pass measures in the filler region was an interaction between attachment and ambiguity in the items analysis of the first pass reading times ($F_1(1,31) = 3.11, p < .09$; $F_2(1,28) = 6.88, p < .05$). This effect should be treated with caution, as this short region exhibited a relatively low probability of first-pass fixation (82% vs. an average of 94% for the other regions) and the effect was completely absent in the analysis of

TABLE 3

Trimmed Raw (and Untrimmed Raw) Regression Path Times (ms) for Experiment 2

Region	Precritical The choirboy	First critical had really	Filler enjoyed	Reflexive himself	Spillover at the concert	Final in the town hall
High/ambiguous	724 (728)	544 (544)	391 (421)	535 (538)	832 (832)	1086 (1100)
High/control	692 (692)	494 (498)	363 (416)	367 (378)	442 (452)	811 (848)
Low/ambiguous	794 (797)	531 (536)	374 (382)	379 (388)	523 (538)	842 (847)
Low/control	736 (738)	456 (457)	406 (413)	315 (320)	410 (420)	848 (851)

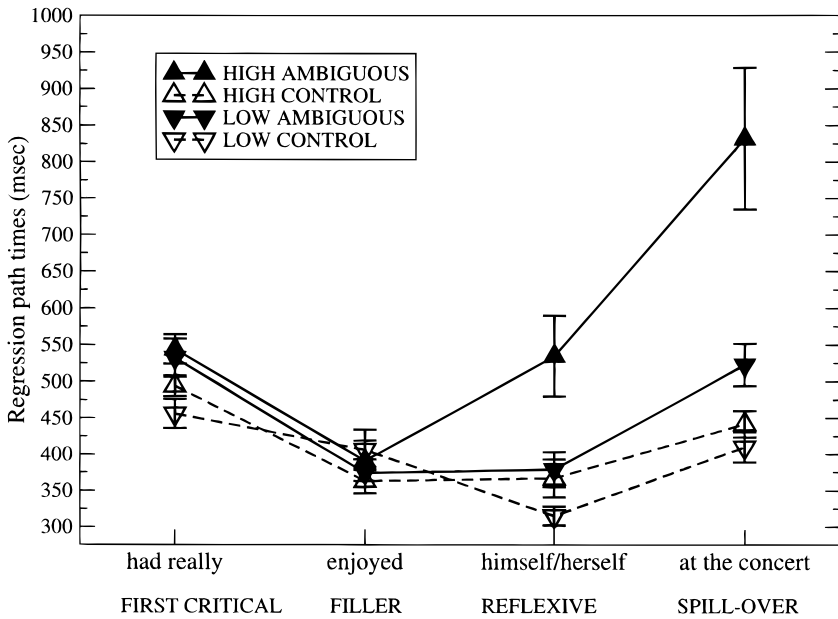


FIG. 3. Trimmed raw regression path times with standard errors for Experiment 2: critical to spillover regions.

first-pass reading times when zero times were included (both F 's < 1).

In the reflexive region (*himself/herself*) there were clear results in the regression path times. Regression path times showed effects of both attachment site, with high attached sentences taking longer to read than low attached sentences ($F_1(1,31) = 9.73, p < .01; F_2(1,27) = 9.81, p < .01$), and ambiguity, with ambiguous sentences taking longer to read than control sentences ($F_1(1,31) = 10.34, p < .01; F_2(1,27) = 19.78, p < .001$). Although the ambiguity penalty was greater for the high attached sentences than the low attached sentences (168 ms vs. 64 ms), this did not lead to a significant interaction between the two factors (both p 's > .15). Planned comparisons showed that the ambiguity disadvantage was significant for both low attached sentences ($F_1(1,31) = 5.13, p < .05; F_2(1,27) = 5.90, p < .05$) and high attached sentences ($F_1(1,31) = 6.55, p < .05; F_2(1,27) = 8.95, p < .01$). The results were less clear in the first-pass reading times. There was a marginal effect of attachment, which was reliable in the participants analysis only ($F_1(1,31) = 4.47, p < .05$;

$F_2(1,27) = 1.22, p > .25$). As in the regression path times, the tendency was for the high attached sentences to be read more slowly than the low attached sentences. There were no other significant effects in the first-pass reading times (all p 's > .09).

In the spillover region (*at the concert*), the two main effects in the regression path times were accompanied by the critical interaction (ambiguity, ($F_1(1,31) = 19.84, p < .001; F_2(1,27) = 20.21, p < .001$; attachment, ($F_1(1,31) = 9.53, p < .01; F_2(1,27) = 11.21, p < .01$ ambiguity \times attachment, ($F_1(1,31) = 5.71, p < .05; F_2(1,27) = 8.77, p < .01$). The interaction occurred because the ambiguity penalty was greater for the high attached sentences than for the low attached sentences. However, as in the previous region, both ambiguous conditions took longer to read than their respective controls (high, ($F_1(1,31) = 13.39, p < .001; F_2(1,27) = 16.22, p < .001$; low, ($F_1(1,31) = 10.84, p < .01; F_2(1,27) = 10.07, p < .01$). No effects approached significance in the first-pass reading times (all F 's < 1.1).

TABLE 4

Trimmed Raw (and Untrimmed Raw) Second-Pass Reading Times (ms) for Experiment 2

Region	Precritical The choirboy	First critical had really	Filler enjoyed	Reflexive himself	Spillover at the concert	Final in the town hall
High/ambiguous	343 (347)	314 (318)	194 (195)	204 (208)	140 (141)	67 (70)
High/control	212 (213)	181 (182)	107 (108)	119 (120)	125 (125)	66 (66)
Low/ambiguous	198 (209)	194 (205)	113 (120)	118 (123)	113 (118)	93 (99)
Low/control	191 (191)	194 (194)	108 (108)	100 (103)	110 (111)	65 (67)

In the final region, the only significant effects in the first-pass measures were for the regression path times. There was a marginal interaction between attachment and ambiguity in the regression path times ($F_1(1,31) = 4.80, p < .05$; $F_2(1,27) = 3.50, p < .08$). Although the interaction missed significance in the items analysis, contrast analyses showed that the ambiguity effect was significant for the high attached sentences ($F_1(1,31) = 7.21, p < .05$; $F_2(1,27) = 6.43, p < .05$), but there were no signs of an ambiguity effect for the low attached sentences (both F 's < 1). The regression path times also showed a marginal main effect of ambiguity, reliable by items only ($F_1(1,31) = 3.39, p < .08$; $F_2(1,27) = 4.53, p < .05$), such that the ambiguous conditions were read more slowly than control conditions.

Second-Pass Reading Time

Second-pass reading times for the analysis regions are presented in Table 4.

From the precritical region until the reflexive region, people spend much more time rereading in the high ambiguous condition than the other three conditions. This results in significant interactions between ambiguity and attachment in all regions from the precritical region to the filler region (precritical; $F_1(1,31) = 7.80, p < .01$; $F_2(1,27) = 6.54, p < .05$; first critical, $F_1(1,31) = 9.56, p < .01$; $F_2(1,27) = 7.50, p < .05$; filler, $F_1(1,31) = 5.76, p < .05$; $F_2(1,27) = 8.73, p < .01$) The same interaction is observed in the reflexive region, though marginal by participants ($F_1(1,31) = 3.94, p < .06$; $F_2(1,27) = 5.12, p < .05$). The interaction occurred because rereading always took longer in the high ambiguous condition than in its control condition

(precritical to reflexive, all p 's $< .01$), but did not take longer in the low ambiguous condition than its control (precritical to reflexive, all F 's < 1.1). In addition, while the two ambiguous conditions differed considerably in rereading time from the precritical region to the reflexive (all p 's $< .01$), the control conditions did not differ (all F 's < 1.5). On the spillover and final regions, the interaction disappeared completely (all F 's < 1.7).³

The simplest interpretation of this pattern is that it reflects the differences in duration and/or frequency of regressions launched from the reflexive pronoun and following regions. The elevated second-pass reading times for the high ambiguous condition in relation to the other conditions shows that this condition was more disrupted by regressions than the other three. This result is consistent with the regression path time results, which also show that regressions were particularly disruptive in the high ambiguous condition in the regions following the reflexive pronoun. However, unlike the regression path times, the second-pass reading time analyses show no evidence at all for difficulty in the low ambiguous condition. Overall, the second-pass reading time analyses show good evidence for a low reanalysis preference.

Summary of Results

In summary, the results of the eye-movement measures are broadly compatible with the results of Experiment 1. The garden path effect at the first critical region, which was found in the

³ The second-pass reading times on the final region result from regressions from the second sentence.

self-paced reading experiment, showed up in both first-pass reading times and regression path times. The interaction between ambiguity and attachment in the spillover region showed up in the regression path times, as in the self-paced reading times in Experiment 1. Finally, the second-pass times give additional evidence that the high ambiguous condition caused more difficulty than the other three.

The main difference between Experiment 1 and Experiment 2 was that in Experiment 2, the regression path times showed that ambiguous conditions were difficult to read not only when the reflexive disambiguated high, but also when it disambiguated low, though the ambiguity penalty was greater for the high attachment conditions. In Experiment 1, there was evidence for difficulty only in the high attachment conditions, though the lack of an ambiguity \times attachment interaction at the reflexive region in that experiment could also have been due to elevated reading times for the low ambiguous condition (see discussion of Experiment 1).

There are at least two possible explanations for ambiguity cost in the regression path times for the low attached conditions in Experiment 2. The first explanation is in terms of spillover effects. According to this account, the ambiguity cost for the low attached conditions in the reflexive and spillover region in Experiment 2 is related to the initial garden path effect in the first critical region. Thus, it is possible that the ambiguity cost for this earlier region continued through to the later regions, elevating the reading times for both low and high attached ambiguous conditions. However, this spillover explanation is difficult to reconcile with the actual pattern of results for the low attached conditions from the first critical region to the reflexive in Experiment 2. Considering the regression path times, for example, there was clear evidence for an ambiguity effect in both the first critical and reflexive regions, but not in the intervening filler region.

A more likely explanation is that the ambiguity effect for the low attached conditions at the reflexive and spillover regions is that this reflects a new garden path effect triggered by the gender marking on the reflexive. This would be

consistent with a claim that the low attachment advantage in reanalysis is a probabilistic preference rather than a purely deterministic one, with a minority of trials exhibiting a dispreferred high attachment in reanalysis. Effects like these, which are due to a small subset of trials, may only be detectable with sensitive techniques such as eye tracking. If so, this would explain the lack of evidence for a low attachment disadvantage in Experiment 1. However, as mentioned above, a probabilistic preference may be responsible for the lack of an interaction at the reflexive region in that experiment, due to elevated reading times for the low ambiguous conditions. At first sight, the results for the second-pass times do not support the claim that the low reanalysis preference is probabilistic, as rereading was never longer in the low ambiguous condition than in its control condition. However, it is possible that the number of trials exhibiting difficulty in the low ambiguous condition was too low for such differences to be picked up by the second-pass reading time measure.

Differences among Eye-Movement Measures

One interesting aspect of the results concerns the differences among the eye-movement measures. Recall that the early garden path effect in the first critical region was established for both of the first pass measures, while in the later reflexive and spillover region, the ambiguity effect and interaction were established only in the regression path times. One possible explanation is that this difference reflects rather low-level differences between the two regions, such as, for example, the fact that region 6 occurs in an earlier position in the sentence than regions 8 and 9. A more interesting (and more speculative) explanation is that the differing eye-movement profiles for these two regions reflect qualitative differences in the type of reanalysis is taking place. In the earlier region, the type of reanalysis (an NP/S reanalysis involving moving from a noun phrase to a sentence interpretation) is relatively easy (Sturt et al., 1999), while the type of reanalysis in the reflexive and spillover region (undoing an NP/S reanalysis, finding the correct antecedent for the reflexive, and performing a new NP/S reanalysis in a different clause) is ob-

viously more complex and would be predicted to cause considerable difficulty in most theories of reanalysis difficulty (Fodor & Inoue, 1998; Gorrell, 1995; Pritchett, 1992). At first sight it seems counterintuitive that the easier reanalysis was detected in two eye-movement measures, but the harder reanalysis in only one. However, it is possible that regressions are more frequent when reanalysis is hard or when more complex grammatical inference is required to diagnose the problem (Fodor & Inoue, 1998). If these regressions tend to be launched soon after the initial fixation in the disambiguating region, processing difficulty may well not be detectable in first-pass times, but would be detectable on the regression path measure, which takes into account the time spent during regressions. Easier reanalysis, on the other hand, may more often occur while the eye remains in the disambiguating region, resulting in detectable differences in first-pass times.

Comprehension Accuracy

Comprehension accuracy results for the four conditions were high ambiguous, 79%; high control, 83%; low ambiguous, 85%; low control, 88%. It can be seen that the descriptive pattern of results resembles those of Experiment 1, with participants performing better on questions about low attached and control sentences than on their high attached and ambiguous counterparts. However, analyses of variance conducted on these percentages did not reveal any significant effects (all p 's > .1), except for a main effect of attachment site in the subjects analysis only ($F_1(1,31) = 4.90, p < .05$; $F_2(1,27) = 2.66, p > .12$). Numerically, the overall comprehension accuracy appears lower than that of Experiment 1. This was confirmed statistically by an analysis of variance combining the scores for the two experiments and treating experiment as an additional between-participants and within-items factor. The main effect of experiment was significant ($F_1(1,62) = 8.11, p < .01$; $F_2(1,27) = 8.52, p < .01$). There were no interactions involving experiment or any other variable (all F 's < 1). One explanation of this is that Experiment 2 included an extra sentence between the critical

sentence and the question. As the question always related to the first sentence, the presence of the intervening sentence could have caused memory interference, resulting in lower scores.

GENERAL DISCUSSION

The purpose of the experiments reported here was to test the claim that a recency preference operates in the disambiguation of reanalysis ambiguities. Both experiments showed evidence to support this claim. In the first experiment, reading times in the region immediately following the reflexive pronoun were longer for the high ambiguous condition than those of the other conditions, which did not differ. In the same region in Experiment 2, regression path times were longer for the ambiguous than the control conditions, and this ambiguity effect was reliably larger for the high conditions. The extra difficulty for the high attached conditions is further supported by the second-pass reading times, which show that participants spent much longer in rereading the reflexive and preceding regions when the sentence was disambiguated high and was ambiguous than in the other conditions, which did not differ from each other. The low reanalysis preference broadly supports the claim that the type of reanalysis involved can be modeled as an attachment operation, which exhibits ambiguity resolution preferences similar to those that operate in initial ambiguity resolution (Fodor & Inoue, 1998; Lewis, 1993; Sturt & Crocker, 1996). However, the theories which make this claim predict the preference to be deterministic. Thus some revision of the models would be needed if the preference is actually probabilistic, as suggested by the results of Experiment 2.

The results of the present paper should be seen in the context of earlier work by Sturt et al. (2001) and Schneider and Phillips (2001), discussed above, which examined sentences like (4), repeated here for convenience:

- (4) The countess who heard the choirboy had really enjoyed himself/herself at the concert (and) organized a charity event afterward.

As mentioned above, these studies showed a high attachment preference for such sentences,

which can be interpreted in terms of a preference to avoid reanalysis (Fodor & Frazier, 1980). This contrasts with the low attachment advantage seen in the present paper. Taken together, the two sets of experiments show that, although the recency preference is general enough to apply to reanalysis, it also interacts with what can be interpreted as a reanalysis avoidance preference. This interactive behavior is predicted by the model of Sturt and Crocker (1996): The processor attempts simpler parsing operations which do not involve reanalysis before more complex operations which do, explaining the high attachment preference for (4). On the other hand, a recency preference applies to the parsing operation required for reanalysis, explaining the low attachment preference reported in the present paper.

As we have seen, the type of reanalysis considered here is relatively simple and can easily be modeled as an attachment operation. One interesting question that can be addressed in future research is whether preferences such as recency also apply when the reanalysis operation involved is more complex and less obviously resembles an attachment operation. A second issue that can be considered is the degree to which reanalysis is encapsulated, for example, whether there is an initial stage of reanalysis during which the information available to the parser is highly restricted (see discussion in Sturt, Pickering, & Crocker, 2000).

The results reported in the present paper are encouraging for the feasibility of such future studies, as they demonstrate not only that systematic preferences operate in resolving reanalysis ambiguities, but also that established experimental techniques can be used to study them.

APPENDIX: EXPERIMENTAL ITEMS

The items for the two experiments are shown here in the low ambiguous condition. To recreate the high conditions, replace *herself* with *himself* and vice versa. To recreate the control conditions, add the word *that* as the last word of the second region (high control) or the fourth region (low control). Single slashes mark region boundaries for the analysis of both experiments (and the segmentation for Experiment 1). Double slashes mark where the line break appeared for each item in Experiment 2 (and one of the segment breaks for Experiment 1). In Experiment 1, the line break always

appeared between the second-last and last regions. Experiment 2 used shorter versions of the last two regions in items 7 and 11. The two alternatives for these items are given in parentheses (square brackets for Experiment 1, and round brackets for Experiment 2). The full materials for both Experiments (including the second sentence of each item) are available at <http://www.idealibrary.com>.

1. The social worker / saw / the foster mother / who accepted // the little boy / didn't really trust / himself / at all / about anything.

2. The vicar / saw / the husband / who accepted / the child bride // didn't really like / herself / very much / at all.

3. The women's refuge / accepted / the call-girl / who saw // the shy monk / had always undervalued / himself / too much / in the past.

4. The sea captain / accepted / the cabin boy / who saw / the mermaid // was always admiring / herself / with a mirror / according to the story.

5. The detectives / remembered / the woman / who recognised // the rapist / had clearly contradicted / himself / on the tape / before the trial.

6. The teachers / remembered / the boy / who recognised / the princess // had cleverly disguised / herself / in a hat / at the event.

7. Everybody / recognised / the old lady / who remembered // the mad uncle / had badly neglected / himself / [for years / and years] (for many / years).

8. The congregation / recognised / the clergyman / who remembered / the duchess // had carefully educated / herself / at home / in Yorkshire.

9. The letter / mentioned / the Dutchman / who understood / the waitress // had really embarrassed / herself / at lunch / in the restaurant.

10. The teachers / mentioned / the girl guide / who understood // the Frenchman / had always fancied / himself / a lot / in the past.

11. The caretaker / understood / the handyman / who mentioned // the dinner lady / had actually confused / herself / [a lot / about the fire procedures] (about the / fire procedures).

12. The manager / understood / the salesgirl / who mentioned / the workman // was now blaming / himself / for the trouble / in the storeroom.

13. The tourists / noticed / the movie man / who discovered // the actress / was always injecting / herself / with drugs / from Afghanistan.

14. The journalist / noticed / the baroness / who discovered / the craftsman // had clearly deceived / himself / seriously / about the issue.

15. The rescue workers / discovered / the cleaning lady / who noticed // the fireman / had badly burned / himself / on the arm / above the elbow.

16. The trainer / discovered / the batsman / who noticed // the female fan / was really enjoying / herself / a lot / during the test match.

17. The villagers / doubted / the barmaid / who reported / the conman // had already betrayed / himself / to the press / the day before.

18. The jury / doubted / the policeman / who reported // the prostitute / had badly hurt / herself / in the fight / outside the pub.

19. The nosy neighbour / reported / the housewife / who doubted / the salesman // could easily justify / himself / to anyone / about the decision.

20. The union / reported / the foreman / who doubted / the chairwoman // had fully committed / herself / to the deal / about the pay agreement.

21. The photographers / found / the countess / who heard / the choirboy // had really enjoyed / himself / at the concert / in the town hall.

22. The reporter / found / the Irishman / who heard / the chorus girl // had badly misbehaved / herself / in Paris / in the night club.

23. The radio operator / heard / the ambulanceman / who found // the trapped girl / had seriously cut / herself / on a nail / with a rusty point.

24. The interviewer / heard / the policewoman / who found // the madman / had somehow hidden / himself / in a bush / near the stream.

25. The locals / recalled / the gypsy girl / who revealed / the Mafia godfather // had fatally shot / himself / in Rome / near a warehouse.

26. The announcer / revealed / the newsgirl / who recalled / the cameraman // had really distinguished / himself / at work / in several ways.

27. The murder investigator / proposed / the butler / who acknowledged // the housemaid / had already deceived / herself / a lot / about the whole business.

28. The director / acknowledged / the spokeswoman / who proposed // the chairman / needed to promote / himself / carefully / to the clients.

REFERENCES

- Altmann, G. T. M., van Nice, K. Y., Garnham, A., & Hestra, J.-A. (1998). Late closure in context. *Journal of Memory and Language*, **38**, 459–484.
- Bader, M. (1998). Prosodic in uences on reading syntactically ambiguous sentences. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 1–56). Dordrecht, The Netherlands: Kluwer.
- Brysaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentence parsing: Evidence from dutch. *Quarterly Journal of Experimental Psychology, Section A—Human Experimental Psychology*, **49**, 664–695.
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). Psyscope: a new graphic interactive environment for designing psychological experiments. *Behavioral Research Methods, Instruments, & Computers*, **25**, 257–271.
- Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition*, **30**, 72–105.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, **27**, 429–446.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, **25**, 348–367.
- Ferreira, F., & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16**, 725–745.
- Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, **30**, 725–745.
- Fodor, J. D., & Ferreira, F. (Eds.). (1998). *Reanalysis in sentence processing*. Kluwer.
- Fodor, J. D., & Frazier, L. (1980). Is the human sentence parsing mechanism an ATN? *Cognition*, **8**, 417–459.
- Fodor, J. D., & Inoue, A. (1998). Attach Anyway. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 101–139). Dordrecht, The Netherlands: Kluwer.
- Fodor, J. D., & Inoue, A. (2000). Garden path reanalysis: Attach (anyway) and revision as last resort. In M. D. Vincenzi & V. Lombardo (Eds.), *Cross linguistic perspectives on sentence processing* (pp. 21–61). Dordrecht, The Netherlands: Kluwer.
- Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies*. Unpublished doctoral dissertation, University of Connecticut, Storrs, CT.
- Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory*, **5**, 519–559.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, **14**, 178–210.
- Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, **37**, 58–93.
- Gorrell, P. (1995). *Syntax and parsing*. Cambridge, UK: Cambridge Univ. Press.
- Hirose, Y. (1999). *Resolving reanalysis ambiguity in Japanese relative clauses*. Unpublished doctoral dissertation, The City University of New York, New York, NY.
- Hirose, Y., & Inoue, A. (1998). *Ambiguity of reanalysis in parsing complex sentences in Japanese*. In D. Hillert

- (Ed.), *Syntax and Semantics: Sentence Processing: A Crosslinguistic Perspective* (Vol. 31, pp. 71–93). San Diego: Academic Press.
- Igoa, J. M., Carreiras, M., & Meseguer, E. (1998). A study on late closure in Spanish: Principle-grounded vs. frequency-based accounts of attachment preferences. *Quarterly Journal of Experimental Psychology. Section A: Human Experimental Psychology*, **51**, 561–592.
- Just, M., Carpenter, P., & Wooley, J. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, **111**, 228–238.
- Kamide, Y. (1999). *The role of argument structure requirements and recency constraints in human sentence processing*. Unpublished doctoral dissertation, University of Exeter.
- Kennedy, A., Murray, W. S., Jennings, K., & Reid, C. (1989). Parsing complements: Comments on the generality of the principle of minimal attachment. *Language and Cognitive Processes*, **4**, 51–76.
- Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, **2**, 15–47.
- Konieczny, L. (1996). *Human sentence processing: A semantics-oriented approach*. Unpublished doctoral dissertation, University of Freiburg, Germany.
- Konieczny, L., Hemforth, B., Scheepers, C., & Strube, G. (1997). The role of lexical heads in parsing: evidence from German. *Language and Cognitive Processes*, **12**, 307–348.
- Lewis, R. (1993). *An architecturally-based theory of human sentence comprehension*. Unpublished doctoral dissertation, Computer Science Department, Carnegie Mellon University.
- Liversedge, S. P. (1994). *Referential contexts, relative clauses, and syntactic parsing*. Unpublished doctoral dissertation, University of Dundee.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, **101**, 676–703.
- Meng, M., & Bader, M. (2000). Ungrammaticality detection and garden path strength: Evidence for serial parsing. *Language and Cognitive Processes*, **15**, 615–666.
- Phillips, C., & Gibson, E. (1997). On the strength of the local attachment preference. *Journal of Psycholinguistic Research*, **23**, 323–346.
- Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24**, 940–961.
- Pickering, M. J., Traxler, M. J., & Crocker, M. W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language*, **43**, 447–475.
- Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. Chicago, IL: Univ. of Chicago Press.
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Prentice Hall.
- Schneider, D., & Phillips, C. (2001). Grammatical search and reanalysis. *Journal of Memory and Language*, **45**, 308–336.
- Stevenson, S. (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research*, **23**, 295–321.
- Sturt, P. (1997). *Syntactic reanalysis in human language processing*. Unpublished doctoral dissertation, Centre for Cognitive Science, University of Edinburgh, Edinburgh, Scotland.
- Sturt, P., & Crocker, M. W. (1996). Monotonic syntactic processing: a cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes*, **11**, 449–494.
- 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.
- Sturt, P., Pickering, M. J., & Crocker, M. W. (2000). Search strategies in syntactic reanalysis. *Journal of Psycholinguistic Research*, **29**, 183–194.
- Sturt, P., Pickering, M. J., Scheepers, C., & Crocker, M. W. (2001). The preservation of structure in language comprehension: Is syntactic reanalysis a last resort? *Journal of Memory and Language*, **45**, 283–307.
- Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language*, **39**, 558–592.
- 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.
- 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.

(Received October 31, 2000)

(Revision received February 28, 2001)

(Published online November 9, 2001)