

RESEARCH

Processing ambiguities in attachment and pronominal reference

Margaret Grant¹, Shayne Sloggett² and Brian Dillon³¹ Department of Linguistics, Simon Fraser University, Burnaby, CA² Department of Language and Linguistic Science, University of York, Heslington, York, UK³ Department of Linguistics, University of Massachusetts Amherst, Amherst, MA, USCorresponding author: Margaret Grant (margaret_grant@sfu.ca)

The nature of ambiguity resolution has important implications for models of sentence processing in general. Studies of structural ambiguities, such as modifier attachment ambiguities, have generally supported a model in which a single analysis of ambiguous material is adopted without a cost to processing. Concurrently, a separate literature has observed a processing penalty for ambiguities in pronominal reference, suggesting that potential referents compete for selection during the processing of ambiguous pronouns. We argue that the apparent distinction between the ambiguity resolution mechanisms in attachment and pronominal reference ambiguities warrants further study. We present evidence from two experiments measuring eye movements during reading, showing that the separation held in the literature between these two ambiguity types is, at least, not uniformly supported.

Keywords: Sentence processing; ambiguity resolution; eye-tracking; reading; pronoun reference; adjunct attachment

1 Introduction

Determining how the language processor handles ambiguous input has been one of the major projects in psycholinguistics over the course of several decades. The processor must be able to handle ambiguities at the level of individual word meanings (*lexical* ambiguity; e.g. Swinney 1979), ambiguities in syntactic structure (*structural* ambiguity, as in Bever's famous sentence *the horse raced past the barn fell*; Bever 1970), and ambiguities about the reference of anaphoric expressions in a given context (*referential* ambiguity). It has proven theoretically insightful to investigate the degree to which the mechanisms that resolve ambiguity are similar across these levels. For example, there has been an active literature focused on the degree to which lexical ambiguity and syntactic ambiguity are processed similarly (see e.g., MacDonald et al. 1994; Traxler et al. 1998). In the present study, we focus on the relationship between the processing of referential ambiguity and structural ambiguity, a comparison that has received relatively less attention in previous work. As we detail below, there are theoretical reasons both for and against the view that referential and structural processing are processed in a similar fashion. At present, there is insufficient evidence to definitively adopt one view over the other. In light of this, our goal in the present paper is to provide an initial evaluation of the degree of similarity between referential and structural processing. To do this, we provide a minimal comparison of the processing of structural and referential ambiguity to determine whether comprehenders react in a qualitatively similar fashion to structural and referential ambiguity in reading. Before we report our experiments, we provide a survey of the literature on both struc-

tural and referential processing. These literatures have largely been separate and seem to support the conclusion that different types of ambiguity are processed differently. In this study we focus on the question of whether ambiguity facilitates or inhibits processing different types of linguistic ambiguity. To preview our major finding, we observe that the processing of both referential and structural relations is facilitated by ambiguity in reading time measures. This similar processing profile suggests that referential and structural ambiguities are processed in a similar fashion.

1.1 Structural processing

Models of structural processing can be divided into two main classes. The first class of model holds that the parser initially constructs a single analysis of incoming material. If this initial analysis is later determined to be incorrect, then reanalysis must occur to construct the appropriate structure. Therefore, such models have been dubbed *two-stage* or *reanalysis* models of sentence processing. While all reanalysis models have the property of assuming a single initial analysis, specific models differ as to how the initial analysis is determined. One classic reanalysis model, the garden-path model (Frazier 1979), holds that structural principles alone guide the initial parse. These principles, including the principles of *Late Closure* and *Minimal Attachment*, were supported by results of early studies of eye movements during reading and self-paced reading (e.g., Frazier & Rayner 1982; Frazier 1987). However, subsequent results indicated that structure-based principles are not the only constraints that influence parsing decisions. In response to the results challenging the syntax-first model, models under which multiple constraints from several information sources (e.g., statistical, semantic, contextual) are simultaneously applied gained favour. One class of such models holds that in processing structural ambiguities, multiple syntactic analyses are activated in a parallel manner (MacDonald et al. 1994; McRae et al. 1998; Spivey & Tanenhaus 1998). Information from a variety of sources, including the current discourse context and the frequency of a syntactic frame for a given lexical item, is used to rank these analyses. In this way, these models propose that syntactically ambiguous material is processed in a fashion similar to what has been proposed for ambiguous lexical items (Gaskell & Marslen-Wilson 1997; Rodd et al. 2002), where the evidence points to activation of multiple meanings with activation levels tracking bias toward one meaning or another (see e.g., Duffy et al. 1988). These lexical representations then compete for selection. In the case of structural ambiguity, the claim is that stored syntactic frames are activated by the input, and when ambiguity is encountered, multiple syntactic representations compete for selection. These *constraint-based* competition models predict that when the information from the constraints at issue does not weigh heavily toward one interpretation over others, processing difficulty increases, and therefore processing time increases as well (although cf. Vosse & Kempen 2009). In general, these models predict the existence of competition effects, processing difficulty or delay upon encountering an ambiguity. Importantly these effects should arise during structural and lexical processing alike (McRae et al. 1998). Although competition effects are a hallmark prediction of constraint-based models, these models do not generally make the stronger claim that all ambiguity creates processing difficulty. For example, Green & Mitchell (2006) show that ambiguity does not create difficulty if the ambiguity is compatible with strong prior biases towards a particular analysis.

Empirical evidence for competition effects in structural processing has proved elusive. Experimental results on the processing of structural ambiguities have generally failed to show such competition effects. In a detailed review, Clifton & Staub (2008) examine reading times for ambiguous regions of text both in experiments that were intended to look

for such effects and in experiments that were designed primarily to test reading times on subsequent disambiguating regions. Clifton and Staub find no convincing evidence for slowed processing time of ambiguous regions, contrary to the predictions of constraint-based competition models of parsing (c.f., Green & Mitchell 2006; but see Clifton & Staub 2008 for further discussion).

Especially pertinent to the present investigation, some studies in fact find less disruption in reading times on ambiguous regions of text than on corresponding unambiguous regions. For example, Traxler et al. (1998) examine the processing of ambiguous versus unambiguous regions of written material in examples such as (1). We note that here, we discuss Traxler et al. (1998)'s results from relative clauses, although prepositional phrase (PP) attachment was also tested. Although later in the paper we show an ambiguity advantage for PP attachment, the advantage for ambiguous PPs was not significant in Traxler et al.'s study.

- (1) Traxler et al. (1998): Experiment 1
- a. The driver of the car that had the **moustache** was pretty cool. (high attachment only)
 - b. The car of the driver that had the **moustache** was pretty cool. (low attachment only)
 - c. The son of the driver that had the **moustache** was pretty cool. (ambiguous)

In an off-line judgment task, Traxler et al. report relatively balanced preference for attachment of relative clauses, with participants indicating a low attachment interpretation 68% of the time. Measures of eye movements during reading, however, showed longer processing times for disambiguated conditions (1a)–(1b) than for the globally ambiguous condition (1c). This *ambiguity advantage* was subsequently replicated using a variety of materials (see van Gompel et al. 2001; 2005).

To account for the ambiguity advantage in attachment, van Gompel et al. (2000) (see also Traxler et al. 1998; van Gompel et al. 2001; 2005) proposed a new reanalysis model. In their *Unrestricted Race Model* (URM) of syntactic ambiguity resolution, potential syntactic analyses engage in a race before a single analysis is adopted. In this race, the information biasing the outcome of the race for one analysis over another can come from any source (syntactic or non-syntactic), constituting the *unrestricted* nature of the race. According to the URM, the ambiguity advantage is the reflection of occasional reanalysis that occurs in either of the unambiguous conditions in (1); when the “wrong” parse wins the race, it will trigger reanalysis, and so slow reading time measures. In the case of global ambiguity, the parser is never revealed to be wrong, and overall processing times are predicted to be fast compared to merely temporarily ambiguous conditions. In this model, critically only information that precedes the ambiguous element can influence the outcome of the race, with the consequence that an input word that is semantically implausible or syntactically unlicensed under the winning syntactic analysis is likely to give rise to a reanalysis (or *garden path*) effect that can be measured through reading time measures.

Taken together, the existing literature on processing structural ambiguity suggests that there is no parallel consideration of multiple structures in a way that gives rise to competition effects in processing. Instead, there is evidence that the availability of multiple structures facilitates reading, either due to a variable-choice structure selection mechanism (van Gompel et al. 2000), or to underspecification (Ferreira & Patson 2007).

1.2 Processing ambiguous reference

At a theoretical level, a number of researchers have used competition-based mechanisms to model the processing of referential dependencies (e.g., Arnold et al. 2000; Badecker & Straub 2002; Kaiser 2011), tacitly assuming that the mechanisms used to resolve referential ambiguity are different in kind than those used to resolve structural ambiguity (and instead bear more similarity to the competition-based mechanism associated with lexical ambiguity). However, to our knowledge there has not been a direct comparison of the real-time processing of structural and referential ambiguities using similar sentence materials (although cf. Green 1995, for a theoretical link between these types, and Hemforth, Konieczny & Scheepers 2000, for an interpretation questionnaire on pronoun reference and attachment). Therefore we regard it as an open empirical question whether referential and structural ambiguities are processed similarly.

In the literature on referential processing, some findings indicate competition effects. For example, MacDonald & MacWhinney (1990) found longer reaction times in a cross-modal probe recognition task for ambiguous pronouns as compared to unambiguous pronouns, which they attributed to increased difficulty associated with resolving ambiguous reference. On the basis of this finding and similar findings (Gernsbacher 1989; Badecker & Straub 1994), Badecker & Straub (2002) argued that pronominal reference is resolved via competitive constraint satisfaction. They reasoned that if pronominal resolution is resolved with a competitive mechanism, competition effects should be evident on the pronoun when it is perceived as ambiguous. In support of this view, they cited findings from Badecker & Straub (1994),¹ who used self-paced reading to investigate sentences such as those in (2).

- (2) Badecker & Straub (1994):
- a. Kenny assured Lucy that he was prepared for the new job.
 - b. Julie assured Harry that he was prepared for the new job.
 - c. Kenny assured Harry that he was prepared for the new job.
 - d. Julie assured Lucy that he was prepared for the new job.

Badecker & Straub (2002) report (p. 750) that reading times at the pronominal region were fastest in (2a), slowest in (2b), and intermediate in (2c)–(2d). They attribute the increased reading time in (2b) relative to (2a) to a competition effect (what they refer to as the *multiple-candidate effect*).

Building on this observation, Badecker and Straub used self-paced reading to test reading times following pronouns in the environment of multiple gender-matched antecedents. In their examples, the pronouns were not globally ambiguous because reference to one of the antecedents was ruled out by Principle B of Chomsky's Binding Theory (Chomsky 1981). Nonetheless, Badecker and Straub found that reading times immediately following an object pronoun were increased in sentences that had multiple gender matched referents, compared to conditions that had only a single grammatical gender-matched antecedent. These results suggest that i) a competitive mechanism resolves pronominal ambiguity and ii) antecedents that are formally ruled out by Binding Theory nonetheless compete for selection and contribute to processing difficulty. However, the strength of these conclusions is called into question by other studies that have failed to find evidence that grammatically illicit antecedents compete for selection (Clifton et al. 1999; Chow et al. 2014). Because this empirical landscape is mixed (see Sturt 2013, for a review), we limit our focus only to studies that contrast the number of grammatically licit antecedents.

¹ Unfortunately, we do not have access to a full report of these findings.

One such study is Garnham et al. (1992), who used a clause-by-clause self paced reading task. They observed slower reading times for clauses containing a pronoun when the preceding context contained multiple gender matched antecedents than when the preceding context contained only a single gender-matched antecedent. Interestingly, Garnham et al. observed this multiple match effect only when the filler sentences in their experiment consistently had pronouns in them; when pronouns were removed from the filler sentences, this effect was not observed. From this pattern, Garnham et al. concluded that the use of gender information that led to the multiple match effect was strategic in nature. Rigalleau et al. (2004) observed similar results in French.

Perhaps most directly relevant for present purposes, Stewart et al. (2007) tested whole-sentence self-paced reading with an experimental design that closely paralleled the designs that have been used to establish the ambiguity advantage effect for structural attachment (Traxler et al. 1998; van Gompel et al. 2000; 2001; 2005). Stewart et al. investigated sentences like *Paul lent Rick/Kate the CD before he/she left for the holidays*. In task contexts where participants were given comprehension questions that explicitly targeted the pronoun's reference (their *deep processing* condition), reading times on the second clause indicated substantially longer reading times for their ambiguous pronoun condition than for either of the unambiguous conditions. This finding is consistent with competition effects in referential processing. In their shallow processing condition, however, they observed that the ambiguous condition was read the most quickly (although this effect did not reach statistical significance). However, because Stewart and colleagues measured whole sentence reading times, it is difficult to interpret their data as unambiguous evidence for a competitive mechanism for resolving pronominal reference. In order to establish that *initial* referential processing is subject to competition effects, one requires a methodology that yields a more detailed picture of the time course of processing.

In sum, there is evidence (along with some lack of evidence) from a range of methodologies that suggests that referential ambiguity causes processing difficulty, possibly implicating competition among competing alternatives as a mechanism for resolving referential dependencies.

1.3 Structural and referential ambiguities: Of a kind?

To summarize: existing literature offers fairly clear evidence that ambiguity facilitates structural processing, but there is some evidence that it hinders referential processing. These different processing profiles seem to imply distinct processing mechanisms for resolving referential and structural dependencies: one grounded in competition, the other not.

However, there is good reason to hypothesize that both attachment and pronominal reference would draw on shared processing mechanisms, and thus show a similar processing profile in real-time processing measures. Both referential and structural processing impose similar functional demands on the processor. Both types of dependency require the comprehender to form a dependency with representations that may be relatively inaccessible in working memory, and thus both require a memory retrieval mechanism to restore these representations to a state where they are available for active processing. Furthermore, it is not clear why the processing mechanisms that underlie the ambiguity advantage in reading (underspecification, or variable-choice selection processes in a race-based framework) would not also be deployed to resolve referential ambiguities. Although some of the evidence indirectly suggests qualitatively distinct mechanisms, we believe that it is premature to draw this conclusion without a direct comparison. And because of the theoretical weight assigned to the studies suggesting separate mechanisms for structural and

referential ambiguity, we believe that a direct comparison is all the more important to the literature on processing ambiguity. To fill this gap in our knowledge, we now turn to our experiments, which were designed to provide a direct comparison of how referential and structural ambiguity are handled by the processor.

1.4 The present studies

Here, we present a study of eye movements during reading that directly compares the processing of attachment and pronoun ambiguities. Based on the studies by Traxler, van Gompel and colleagues discussed above, we expect to find an ambiguity advantage in eye movements during reading for globally ambiguous attachment over conditions that are disambiguated toward either high or low attachment. For pronoun reference, the prediction from much of the literature would be that we should see a slowdown or other evidence of difficulty in reading when a pronoun is ambiguous with respect to its referent. However, as we discussed, the results leading to this prediction are equivocal, and a penalty for ambiguous reference has not been universally found. A competing hypothesis is that in fact, referential ambiguity is like attachment ambiguity in processing. Under this hypothesis we would expect to find an ambiguity advantage for both ambiguity types, in the form of faster reading times and/or fewer regressions from the ambiguous region. In Experiment 1, we find that this second possibility is supported by the evidence from eye movements during reading. We show that the processing of structural and referential ambiguities is facilitated by ambiguity; in this respect, both types of dependency are processed differently than lexical ambiguity.²

In a second experiment, we partly replicate a design from Badecker & Straub (1994) (as cited in their 2002 article), in order to determine whether our lack of competition-based penalty for pronoun ambiguity is due to differences between our experimental materials and their original study. The results of Experiment 2 fail to show a competition penalty for pronouns with ambiguous reference, and instead show a clear processing advantage for referential ambiguity. We conclude that the ambiguity advantage found for the pronoun conditions of Experiment 1 is unlikely to be due to the particular properties of the sentence materials in those experiments.

2 Experiment 1

2.1 Method

2.1.1 Participants

Fifty-nine University of Massachusetts undergraduates were tested individually for course credit. Two participants were excluded from the final analyses due to excessive data loss (due to blinks, etc.) and three were excluded due to poor accuracy (less than 65%) on a subset of comprehension questions that did not target the critical potentially ambiguous material (see description of materials below).

2.1.2 Materials

Thirty items were created following the pattern shown in Table 1. Two variables were manipulated in the item sets: dependency type (attachment versus pronoun reference),

² In order to test for differences across experimental tasks, we also ran a version of Experiment 1 using self-paced reading. The results of the self-paced reading are broadly consistent with the results of Experiment 1: we failed to observe any evidence of any ambiguity penalty in our materials for either PP or pronominal conditions. Instead, we observed a marginal ambiguity advantage on a spillover region in our materials. On its own, this marginal result does not provide strong evidence for an ambiguity advantage effect in self-paced reading for either dependency. However, these results do converge with the findings reported in detail here. For a full report of the self-paced reading results, see http://grant-psycholinguistics.com/grantsloggettdillon_supplement/, password "Ambiguity".

Table 1: Sample item set. Slashes (/) indicate regions of interest for eye movement measures but were not visible during the experiment.

Condition	Example
Ambiguous, PP attachment	We met/ the brother/ of the waiter/ with a beard/ but we didn't talk for long.
High, PP attachment	We met/ the brother/ of the waitress/ with a beard/ but we didn't talk for long.
Low, PP attachment	We met/ the sister/ of the waiter/ with a beard/ but we didn't talk for long.
Ambiguous, Pron. reference	We met/ the brother/ of the waiter/ when he visited the restaurant,/ but we didn't talk for long.
High, Pron. reference	We met/ the brother/ of the waitress/ when he visited the restaurant,/ but we didn't talk for long.
Low, Pron. reference	We met/ the sister/ of the waiter/ when he visited the restaurant,/ but we didn't talk for long.

and what we will call height (ambiguous, unambiguous high or unambiguous low attachment/reference). Each sentence contained a complex determiner phrase (DP) containing an *of* genitive, for example *the brother of the waiter*. We assume that in this example, the noun *brother* both linearly precedes the second and is also higher in the hierarchical structure. This complex DP was either the direct object of a transitive verb or the object of a preposition. In the PP-attachment conditions, the complex noun phrase was followed by a PP modifier. The ambiguity manipulation for these conditions was achieved by manipulating the match between the genders of the potential modifier hosts and the gender-biased attribute described in the modifier (for example, *with a beard*, which stereotypically describes a man). In the ambiguous condition, both nouns (*brother* and *waiter*) are appropriate hosts for the modifier according to stereotypical gender. In the high condition, the second (lower in the syntactic structure) noun was changed such that its gender was not a stereotypically appropriate host for the modifier. In the low condition, the first (higher in the syntactic structure) noun was changed such that it was not a stereotypically appropriate host for the modifier. The gender of the high and low nouns was either unambiguous due to lexical-semantic and/or morphological cues (e.g., *sister*, *waitress*) or rested on stereotypical bias (e.g., *beautician*). For those nouns that relied on stereotypical gender, the mean rating for male-biased nouns in Kennison & Trofe (2003)'s norms was 5.61 on a scale from 1 ("mostly female") to 7 ("mostly male") (range: 4.4 to 6.8; standard deviation: 0.73), while the mean rating for female-biased nouns was 1.93 (range: 1.0 to 3.0; standard deviation: 0.43).³ Using this design, the critical region was the same for all PP-attachment conditions. For the pronoun reference conditions, the same manipulation of the higher and lower nouns was used. In these cases, the complex DP was followed by a second clause beginning with *when* or *after* and a third person nominative pronoun that could appropriately refer to both nouns in the ambiguous condition, to the higher noun only in the high condition and to the low noun only in the low condition. The main clause subjects were either first- or second-person pronouns, plurals or inanimates so that the pronouns in the pronoun reference conditions could not refer to the main clause subject. All items are included in Appendix ??.

To determine the preferred interpretations of our experimental items, we conducted a norming study asking a separate group of English native speaker participants (57 participants residing in the United States recruited via Amazon's Mechanical Turk, age range

³ A small number of nouns that were not lexically disambiguated were not included in Kennison & Trofe (2003)'s norms. These included *masseuse*, used as a female-biased noun, and *actor*, *mechanic* and *driver* used as male-biased nouns.

21–60) to explicitly answer which of the two potential DPs served as the attachment site or reference for the ambiguous region. The ambiguous PP-attachment condition showed 25% high attachment interpretations (comparable to Traxler et al. 1998, who found 26% high attachment for PPs), while the ambiguous pronoun reference condition showed 72% high reference. Therefore, although each meaning was chosen on a proportion of trials, the two dependency types do appear to have opposite preferences.

Another important feature of our PP-attachment conditions is that for some of our materials, there is an additional ambiguity wherein the PP-modifier may initially be taken as a modifier of the main verb phrase (we thank a reviewer for pointing this out). We ran an additional norming study to assess the salience of this interpretation, wherein we asked participants to choose a paraphrase that best matched their initial interpretation of the sentence from high-, low-, and verb phrase (VP) attachment options. The study (with 21 participants) showed an overall VP-attachment rate of 10.3%, with thirteen of those participants choosing the VP option less than 10% of the time. Despite the fact that this interpretation was chosen in some trials on the norming study, we believe that this was not a very salient option in our eye movement study. If this interpretation is initially pursued, then readers should experience a garden path effect at, e.g., *beard* in all of our PP-attachment conditions. Such a garden path would likely overwhelm any effects of our attachment manipulation, as readers should experience the garden path to the same extent in all versions of the PP-attachment materials. Thus this alternative possible analysis of the PP works against our ability to observe any differences between our conditions. Even still, as we will show in the sections to follow, we do find differences between PP-attachment conditions. Therefore, we do not believe that this additional ambiguity complicates our interpretation of the results.

2.1.3 Procedure

Participants were tested individually using an Eyelink 2000 eye-tracker (SR Research: Mississauga, Canada) sampling at 1000Hz. Viewing was binocular, but only the movements of one eye were recorded. After an initial set-up and calibration phase, each experimental session consisted of a series of randomized trials in which participants triggered the appearance of a sentence with their eye movements to a start box at the lefthand side of the screen. Participants were instructed to read normally for comprehension, and to indicate that they had finished reading the sentence with a button press. After each sentence, a two-choice comprehension question appeared, which participants answered with a button press. A portion (12/30) of the comprehension questions targeted the potential ambiguity of the sentence by asking a yes/no question about one of the potential interpretations (we will refer to these as *targeted* questions, e.g., *Was the heiress wearing high heels?*), and the rest targeted other aspects of the meaning of the sentence (we will refer to these as *general* questions, e.g., *Did we meet the waiter?*). Our experimental items were counterbalanced 6 across lists in a latin-square design and intermixed with 52 unrelated filler items. The entire experimental session lasted approximately 45 minutes.

2.2 Results

2.2.1 Statistical analysis

Prior to data analysis, trials with track losses or blinks on the critical region were eliminated (5.1% of total data following participant exclusions). Fixations with durations under 80ms were removed from analyses. The experimental conditions were coded according to two factors: dependency type (PP-attachment vs. pronoun reference), and height (ambiguous, unambiguous high or unambiguous low). In statistical analyses the dependency type factor was coded using sum coding (we refer to factors used in our statistical

analysis as in caps, e.g., dependency type), with coefficients -0.5 for pronoun reference and 0.5 for PP-attachment. The height factor was Helmert-coded to create two contrasts: ambiguity (coefficients: ambiguous conditions = 2, high and low conditions = -1) and height (coefficients: high conditions = 1, low conditions = -1 , ambiguous conditions = 0). Interactions between dependency type and ambiguity as well as dependency type and height were included in the models. Because only the unambiguous conditions were coded as high or low, an interaction between dependency type and height is impossible. Significance of the contribution of these factors to the outcome was assessed using linear-mixed effects in the case of continuous outcomes (such as fixation duration measures) and logistic mixed-effects models for categorical outcomes (e.g., regression data, question accuracy). All models were run using the lme4 package (Bates 2005) in R (R Core Team 2018). Due to convergence problems with models including maximal random-effects structure (Barr et al. 2013), we ran initial models with random intercepts and random slopes for dependency type, ambiguity, height, and the interactions between dependency type and ambiguity and between dependency type and height, excluding the correlation parameter between random effects. In cases where such models would nonetheless not converge, we eliminated random slopes for the interactions. In case of a singular fit, we removed random slopes with zero or near zero variance until a non-singular fit was obtained. Thus, the reported p -values in all cases represent the most maximal possible non-singular linear mixed effects model. For linear models, we report p -values estimated using the Satterthwaite approximation (see Luke 2017), implemented in the lmerTest package in R (Kuznetsova et al. 2017). For categorical measures, we report Wald's z and an associated p -value.

2.2.2 Question accuracy

Mean proportion correct for all questions is reported in Table 2, separated into questions that targeted the ambiguity and general questions. For ambiguous sentences, a “correct” response in the targeted questions corresponds to participants accepting the high attachment or highest reference interpretation (although both interpretations are possible). Participants accepted the high interpretation roughly half of the time for both attachment and pronominal reference ambiguities (52.9 vs. 45.5 percent, respectively). We note that these rates differ from the norming study conducted on our ambiguous items, potentially due to the difference in question structure (accepting or rejecting a single interpretation versus choosing between two options) or could reflect the particular subset of items for which targeted questions were presented. For the unambiguous items, accuracy was higher presumably because there was one intended interpretation which was marked as correct. In a mixed-effects logistic regression analysis, the effect of ambiguity (interpreted

Table 2: Mean accuracy on comprehension questions, Experiment 1. *NB: For the ambiguous ambiguity-targeting items, “correct” corresponded to accepting a high interpretation.*

Question Type	Ambiguity Type	Attachment		
		High	Low	Ambiguous
Ambiguity-targeting items				
	PP-attachment	65.7	88.1	52.9
	Pron reference	81.0	63.8	45.5
General questions				
	PP-attachment	90.8	89.5	84.4
	Pron reference	88.3	80.4	87.7

as the effect of having one correct answer) was unsurprisingly significant ($\hat{\beta} = -0.53 (\pm 0.110)$, Wald's $z = -5.052$, $p < .001$). For the unambiguous conditions, accuracy rates for targeted questions were numerically higher for the preferred interpretation of each sentence type, as supported by a significant interaction between dependency type and height ($\hat{\beta} = -1.50 (\pm 0.44)$, Wald's $z = -3.41$, $p < .001$). This pattern could mean that on some number of trials, participants maintained an implausible or ungrammatical interpretation rather than revising their representation toward a dispreferred interpretation. No other effects were significant for targeted questions. General questions showed a high overall accuracy rate, and the only effect that approached significance was the interaction between ambiguity and ambigtype ($\hat{\beta} = -0.40 (\pm 0.23)$, Wald's $z = -1.751$, $p = 0.08$).

2.2.3 Eye movement measures

Here we report analyses of the pre-critical region, consisting of the prepositional phrase containing the lower NP, the critical region, containing the ambiguous PP or an adverbial clause that included the critical pronoun, and the post-critical region, containing the completion of the sentence (see Table 1 for sample regioning). Table 3 contains the means and standards deviations in each region for each experimental condition. Because the critical pronoun was the subject of an embedded clause in each item, the critical region between the lower NP and the comma prior to the sentence continuation was longer for pronoun reference than for PP-attachment conditions. In the main results section, we maintain the critical region as one unit. However, we conducted additional analyses in which we analyzed the critical region up to the pronoun plus one subsequent word. The mean values for each eye movement measure, included in Appendix A, show no major differences in their pattern from analyses of the critical region as a whole.

We conducted statistical analyses for five eye movement measures: first pass time, go-past time, total time, re-reading time and proportion of regressions out. First pass time is defined as the sum of all fixations from first entering a region until leaving it to the left or the right. We define go-past time as the sum of all fixations from first entering a region until leaving it to the right. This measure is sometimes called regression path duration (see Duffy et al. 1988; Brysbaert & Mitchell 1996; Konieczny et al. 1997). Total time is defined as the sum of all fixations on a region. Re-reading time is the sum of all fixations on a region having previously gone past it (excluding zero values). Proportion of regressions out is a categorical measure encoding whether or not a participant made a regressive (leftward) eye movement from a given (fixated) region during first pass reading. For an in-depth review of measures of eye movements during reading, see Staub & Rayner (2007). We can think of first pass time and go-past time measures as reflective of initial processing, while total time is a later measure that also includes time spent on a region after having read further along in the sentence or even during a second pass of the sentence as a whole (but see Staub & Rayner, 2007, for important qualifications to this view). For eye movement measures, raw values are shown in table 3, and go-past times across the entire sentence are pictured in Figure 1. However, a Box-Cox transform on the eye movement measures analyzed suggested that a log transform of the RTs would most closely approximate a normal distribution of our reading time measure, and therefore analyses of log eye movement measures are reported.

Examining the dependency type manipulation, we found that pronoun reference conditions had consistently longer values in eye movement measures than PP-attachment conditions. This advantage for PP-attachment conditions held on the critical region in first pass time, go-past time, and total time, and re-reading time. This main effect, however, could be due merely to the difference in words and in length of the region between PP-attachment and pronoun reference conditions, as PP-attachment conditions were

Table 3: By-participant means (standard errors) in milliseconds, Experiment 1.

Measure	Condition	Pre-critical	Critical	Post-critical
First pass time	Ambig, PP	551 (25)	494 (21)	818 (31)
	High, PP	538 (25)	559 (23)	844 (35)
	Low, PP	599 (30)	520 (22)	867 (37)
	Ambig, Pro	568 (30)	790 (30)	828 (30)
	High, Pro	609 (29)	844 (36)	864 (33)
	Low, Pro	604 (27)	809 (35)	845 (35)
Go-past time	Ambig, PP	726 (45)	789 (55)	2474 (151)
	High, PP	713 (41)	935 (64)	2603 (170)
	Low, PP	772 (41)	887 (60)	2475 (156)
	Ambig, Pro	758 (39)	972 (43)	2524 (162)
	High, Pro	760 (38)	1141 (60)	2523 (181)
	Low, Pro	794 (41)	1093 (43)	2603 (156)
Total time	Ambig, PP	1216 (60)	958 (51)	1178 (48)
	High, PP	1323 (75)	1069 (55)	1177 (48)
	Low, PP	1300 (68)	953 (48)	1218 (54)
	Ambig, Pro	1083 (52)	1304 (57)	1136 (50)
	High, Pro	1141 (60)	1441 (57)	1179 (62)
	Low, Pro	1170 (52)	1389 (62)	1199 (55)
Re-reading time	Ambig, PP	807 (55)	709 (45)	679 (68)
	High, PP	943 (69)	722 (51)	686 (47)
	Low, PP	833 (48)	619 (39)	680 (46)
	Ambig, Pro	646 (38)	770 (49)	614 (54)
	High, Pro	682 (41)	906 (60)	623 (43)
	Low, Pro	680 (42)	772 (49)	731 (50)
Proportion of regressions out	Ambig, PP	0.18 (0.03)	0.27 (0.03)	0.81 (0.03)
	High, PP	0.17 (0.03)	0.28 (0.03)	0.80 (0.04)
	Low, PP	0.19 (0.03)	0.30 (0.04)	0.80 (0.03)
	Ambig, Pro	0.18 (0.03)	0.14 (0.02)	0.82 (0.03)
	High, Pro	0.16 (0.03)	0.16 (0.03)	0.77 (0.04)
	Low, Pro	0.18 (0.03)	0.19 (0.02)	0.84 (0.02)

shorter overall (e.g., *with a beard* vs. *when he entered the restaurant*). In contrast, there were more regressive eye movements out of the critical region, for PP-attachment conditions than for pronoun reference conditions. Interestingly, some differences based on dependency were already present in the pre-critical region, as determined in first pass time and (marginally) go-past time. Given that there is no difference between PP-attachment and pronoun reference conditions at this point in the sentence, this may either be a spurious finding or an effect of parafoveal preview. Total time on the pre-critical region, in contrast, showed the opposite pattern, with shorter total times for pronoun reference as compared to PP-attachment, as did re-reading time. None of our predictors were significant in analysis of regressions out on the pre-critical region.

Now we turn to the effect of ambiguity, high or low attachment/reference on reading times in each dependency type. On the critical region, first pass times show an overall advantage for ambiguous conditions over unambiguous conditions. The interaction

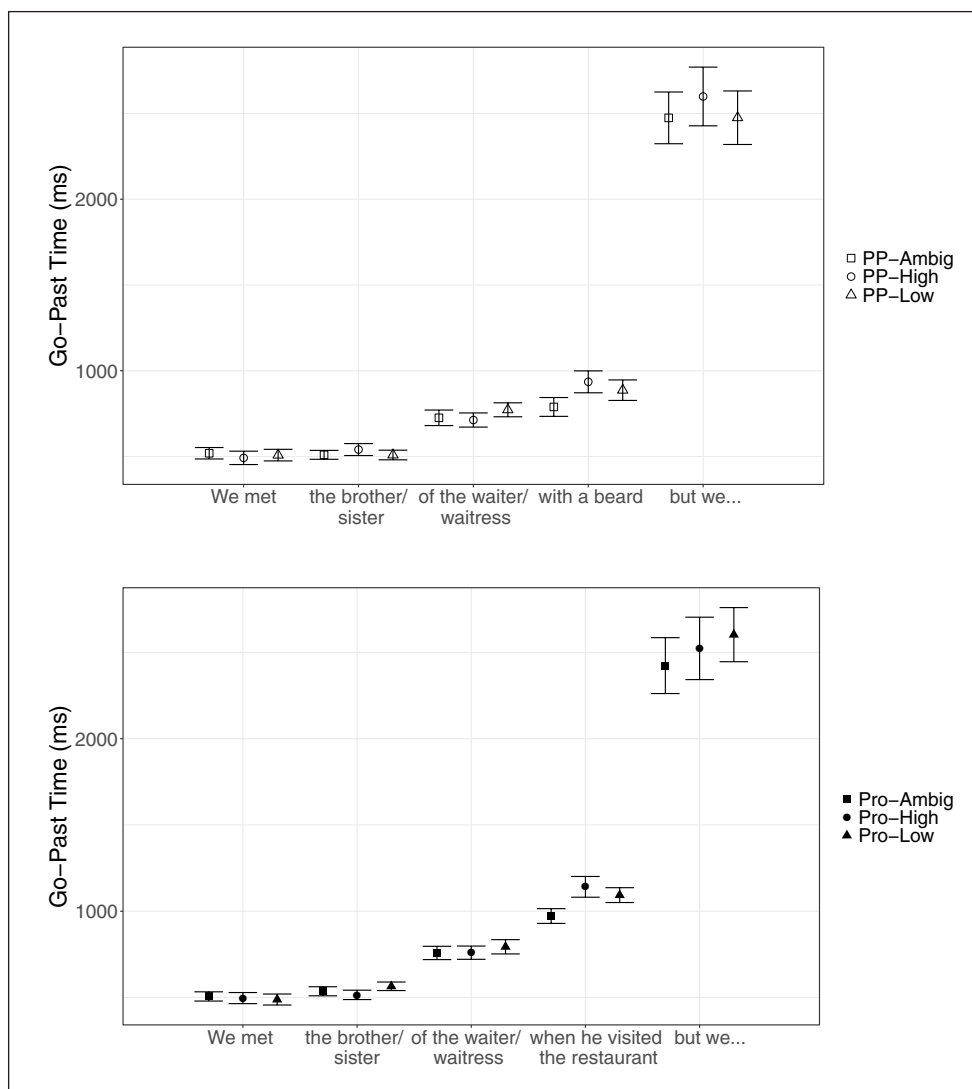


Figure 1: Go-past times across the entire sentence, Experiment 1. PP-attachment conditions are shown in the upper panel, and pronoun reference reference conditions are shown in the lower panel.

between ambiguity and dependency type factors was not significant. However, because we are comparing the mechanisms for processing ambiguous material of both types, we ran separate models including only the PP-attachment conditions or only the pronoun reference conditions. For PP-attachment, the ambiguity advantage was significant ($\hat{\beta} = -0.03 (\pm 0.013)$, $t = -2.173$, $p = 0.040$), and there was an effect of height such that high attachment was associated with longer first pass times than low attachment ($\hat{\beta} = 0.042 (\pm 0.021)$, $t = 2.08$, $p = 0.040$). However for pronoun reference items alone, there was no significant ambiguity advantage ($\hat{\beta} = -0.02 (\pm .01)$, $t = -1.45$, $p = 0.149$). First pass times further showed an effect of height such that high conditions had marginally longer first pass times than low conditions. The ambiguity advantage in first pass times carried over numerically to the post-critical region, however in the statistical model this was only marginal.

Go-past times showed a similar pattern of results on the critical region, with shorter times for ambiguous than unambiguous conditions. In go-past time, however, the effect of ambiguity was significant when the pronoun reference conditions were considered in isolation ($\hat{\beta} -0.04 (\pm 0.01)$, $t = -3.68$, $p = 0.001$) as well as the PP-attachment conditions ($\hat{\beta} = -0.04 (\pm 0.02)$, $t = -2.31$, $p = 0.026$). Both first pass time and go-past time suggest

an overall ambiguity advantage, rather than an ambiguity penalty or a pattern where the globally ambiguous condition patterned with one preferred interpretation (high or low).

In total time, there was also a significant ambiguity advantage on the critical region. However, the numerical pattern is such that for pronoun reference, unambiguous conditions were longer than ambiguous conditions, while PP-attachment cases showed that only high attachment diverged from the ambiguous condition, with low attachment on a par with ambiguous sentences. In the main model for total time, the interaction between dependency type and ambiguity was not significant, and the interaction between dependency type and height was marginal. While the numerical pattern suggested that there may be an effect of height, with high conditions had numerically longer total times than low conditions, the contribution of height to the model did not reach significance. Therefore, any low advantage cannot be taken as completely reliable. To further investigate the apparent difference in pattern between PP-attachment and pronoun reference conditions in total time on the critical region, models were run separately on conditions of each dependency type. Pronoun reference conditions showed a significant effect of ambiguity ($\hat{\beta} = -0.03 (\pm 0.01)$, $t = -2.62$, $p = 0.012$), but the height factor was not significant ($\hat{\beta} = 0.01 (\pm 0.02)$, $t = 0.465$, $p = 0.642$). PP-attachment conditions showed a different pattern. For PP-attachment, contrast between low and high attachment was significant ($\hat{\beta} = 0.06 (\pm 0.02)$, $t = 3.02$, $p = 0.003$), while there was no general ambiguity advantage ($\hat{\beta} = -0.02 (\pm 0.01)$, $t = -1.48$, $p = 0.147$). In sum, total times on the critical region did show a significant ambiguity advantage, however the numerical pattern was more complex, with some evidence for a low attachment advantage (as would be predicted by a garden path model) for PPs.

To complete our results section, it is worthwhile to mention that in re-reading time, there was a marginal effect of height such that high conditions overall had longer re-reading times than low conditions. Analyses of regressions out on the post-critical region showed a marginal interaction between dependency type and height in regressions out. While the proportion of regressive eye movements for PP-attachment conditions were flat across ambiguity and height, for pronoun reference conditions, the high condition showed numerically fewer regressions out than the low or ambiguous conditions.

2.3 Discussion

The results of Experiment 1 show an overall ambiguity advantage for both dependency types. This advantage was particularly clear in go-past time, which showed an ambiguity advantage on the critical region that held for each dependency type when tested in isolation. The ambiguity advantage for attachment ambiguities is consistent with the previous literature supporting the Unrestricted Race Model model (van Gompel et al. 2001; 2005). Our results are thus in line with the review by Clifton & Staub (2008), who show that studies of structural ambiguity in the psycholinguistics literature have typically failed to find competition penalties. In our study, neither referential nor structural ambiguities showed evidence of a competition penalty. This empirically distinguishes structural and referential ambiguities from lexical ambiguities, which reliably show competition effects (again see Clifton & Staub 2008 for a review). The present results are inconsistent with any parsing model that posits a processing cost due to competition among alternative analyses constructed in parallel, such as constraint satisfaction models (e.g., MacDonald et al. 1994; McRae et al. 1998). However, the lack of competition effects does not provide conclusive evidence against the entire class of constraint-based parsing models. Green & Mitchell (2006) show that if the preceding context can set up a bias in favor of one interpretation over another, McRae and colleagues' constraint-based model can capture the ambiguity advantage effect.

Table 4: Model parameters (fixed effect estimates (standard errors), t-values and p-values) for eye movement measures, Experiment 1.

Effect	First pass time								
	Pre-critical region			Critical region			Post-critical region		
	$\hat{\beta}$	t	p	$\hat{\beta}$	T	P	$\hat{\beta}$	t	P
Dep	-0.07 (0.03)	-2.52	0.018	-0.44 (0.06)	-7.71	<0.001	-0.02 (0.03)	-0.71	0.483
Ambig	-0.01 (0.01)	-1.45	0.161	-0.02 (0.01)	-2.53	0.018	0.01 (0.01)	-1.68	0.093
Height	-0.02 (0.02)	-1.03	0.314	0.03 (0.02)	2.31	0.021	-0.01 (0.01)	-0.52	0.605
Dep*Ambig	0.01 (0.02)	0.29	0.775	-0.01 (0.02)	-0.69	0.489	-0.01 (0.02)	-0.34	0.734
Dep*Height	-0.03 (0.03)	-1.14	0.254	0.02 (0.03)	0.54	0.587	-0.03 (0.03)	-0.99	0.323
Effect	Go-past time								
	Pre-critical region			Critical region			Post-critical region		
	$\hat{\beta}$	t	P	$\hat{\beta}$	t	P	$\hat{\beta}$	t	P
Dep	-0.06 (0.03)	-2.02	0.055	-0.30 (0.05)	-5.67	<0.001	0.03 (0.03)	0.83	0.405
Ambig	-0.01 (0.01)	-1.18	0.239	0.04 (0.01)	-3.85	0.001	-0.01 (0.01)	-0.51	0.611
Height	-0.02 (0.02)	-1.57	0.117	0.02 (0.02)	1.37	0.172	-0.004 (0.02)	-0.22	0.824
Dep*Ambig	-0.01 (0.02)	-0.33	0.738	0.003 (0.02)	0.16	0.874	0.01 (0.02)	0.50	0.620
Dep*Height	-0.02 (0.03)	-0.57	0.571	0.02 (0.03)	0.76	0.449	0.04 (0.04)	1.07	0.283
Effect	Total time								
	Pre-critical region			Critical region			Post-critical region		
	$\hat{\beta}$	T	p	$\hat{\beta}$	t	P	$\hat{\beta}$	t	P
Dep	0.10 (0.03)	3.23	0.003	-0.36 (0.05)	-6.75	<0.001	0.04 (0.02)	1.65	0.104
Ambig	-0.02 (0.01)	-2.28	0.031	-0.02 (0.01)	-2.85	0.004	-0.01 (0.01)	-0.65	0.518
Height	-0.01 (0.01)	-0.81	0.418	0.04 (0.01)	2.58	0.010	-0.02 (0.01)	-1.29	0.198
Dep*Ambig	-0.001 (0.02)	-0.08	0.937	0.01 (0.02)	0.63	0.529	0.01 (0.01)	0.81	0.417
Dep*Height	0.03 (0.03)	1.17	0.244	0.05 (0.03)	1.88	0.061	-0.01 (0.02)	-0.51	0.612
Effect	Re-reading time								
	Pre-critical region			Critical region			Post-critical region		
	$\hat{\beta}$	t	p	$\hat{\beta}$	T	P	$\hat{\beta}$	T	P
Dep	0.21 (0.04)	4.97	<0.001	-0.16 (0.06)	-2.71	0.012	0.01 (0.06)	0.24	0.812
Ambig	-0.02 (0.02)	-1.31	0.204	-0.01 (0.02)	-0.39	0.699	-0.01 (0.02)	-0.65	0.520
Height	0.03 (0.02)	1.08	0.282	0.06 (0.03)	2.03	0.053	-0.01 (0.03)	-0.35	0.725
Dep*Ambig	-0.03 (0.03)	-1.17	0.241	0.04 (0.03)	1.45	0.147	0.01 (0.04)	0.33	0.738
Dep*Height	0.09 (0.05)	1.89	0.059	0.02 (0.05)	0.39	0.700	0.04 (0.06)	0.70	0.484
Effect	Regressions out								
	Pre-critical region			Critical region			Post-critical region		
	$\hat{\beta}$	Z	P	$\hat{\beta}$	Z	P	$\hat{\beta}$	Z	P
Dep	0.07 (0.15)	0.46	0.644	0.81 (0.19)	4.28	< 0.001	-0.08 (0.14)	-0.60	0.546
Ambig	-0.01 (0.06)	-0.14	0.892	-0.08 (0.05)	-1.60	0.110	0.04 (0.05)	0.89	0.371
Height	-0.04 (0.08)	-0.50	0.616	-0.08 (0.09)	-0.90	0.367	-0.13 (0.09)	-1.50	0.135
Dep*Ambig	0.002 (0.10)	0.02	0.987	0.09 (0.10)	0.97	0.331	-0.02 (0.10)	-0.22	0.829
Dep*Height	-0.02 (0.17)	-0.12	0.901	0.02 (0.16)	0.14	0.89	0.28 (0.17)	1.65	0.099

In some measures, the numerical pattern of results suggested a low attachment or recent referent advantage for both dependency types. Although this effect was only significant in total time when the PP conditions were analyzed in isolation, the numerical pattern held across measures and across the dependency types. A low attachment preference has been found for English (e.g., Clifton & Carreiras 1999; and offline judgments of our stimuli), which has been taken as evidence for a principle of late closure (Frazier 1979). However, the low-attachment preference has not universally been found for English (see e.g., Carreiras & Clifton 1993) and a high-attachment preference has been found for some languages (e.g., for Spanish, Cuetos & Mitchell 1999; although see Gilboy, Sopena, Clifton & Frazier 1995 for arguments that the differences between English and Spanish may be smaller than these studies suggest). In the literature on pronoun resolution, recency effects have not been paramount, with preferences instead leaning toward reference to the topic or subject where possible (Crawley et al. 1990; Cowles et al. 2007). Because the potential referents for the ambiguous pronouns in our items both made up a part of the complex direct object of the sentence, topic or subject-hood could not play a role in the choice of referent in our materials. Interestingly, the trend towards a recency preference in the reading time data seems to run counter to offline judgments of our stimuli, which suggested that there was a preference to take the higher NP as an antecedent (70% high NP reference).

For a subset of items, the comprehension question following the sentence presentation is informative in determining the final representation of the ambiguous material that was achieved. We found that on this subset of items, participants accepted a high-attachment interpretation on 52.9% trials and for pronoun reference, on 45.5% of the trials. Therefore there was no strong bias toward either interpretation in sentence-final judgments for participants in this experiment, at least for this subset of items. We note that the results on this subset of comprehension questions differ from our off-line norms, which may be due to the difference in question format (a yes/no question about the high interpretation as compared to a two-answer forced choice question) or could be due to the particular subset of items with targeted questions in the eye movement study as compared to the full set in our norming study. Nonetheless, these interpretation results are intuitive with respect to the ambiguity advantage observed. For attachment under an unrestricted race model; on approximately half of the trials the initial interpretation was high-attachment, causing difficulty if the material was disambiguated toward low-attachment, and vice-versa.

For pronoun reference, a similar conclusion can be drawn, although some extra assumptions are necessary. The limited set of interpretation results shows that for ambiguous pronouns, the higher referent was chosen on approximately half of trials. Therefore, when gender cues indicate that the lower referent is in fact the pronoun's referent, a processing cost should appear. Likewise, the cost should appear on the continuation disambiguated toward the higher NP when a low interpretation was initially formed. This conception of the results, however, requires the assumption that conflicting gender cues are not enough to prevent an initial referential interpretation. While this assumption may be counterintuitive, there is independent evidence supporting a failure to use gender cues in a task targeting "automatic" processing (Greene et al. 1992; Rigalleau & Caplan 2000; Rigalleau et al. 2004; Stewart et al. 2007). We believe that measuring eye movements during reading is an example of an automatic task in this sense. Other than the subset of questions targeting the reference of the pronoun (which made up 6 questions per participant out of 82 total trials), nothing about the study invited anything other than natural reading. By contrast, Greene et al. (1992) found that a unique referent was identified using gender

cues in a task targeting “strategic” processing, wherein an explicit decision had to be made by the participants.

The results of Experiment 1 suggest that pronominal ambiguities and structural ambiguities are similar in the sense that the ambiguity advantage effect is seen for both dependencies in similar sentential contexts. Nonetheless, the finding that pronominal ambiguity speeds, rather than slows reading conflicts most directly with the results of Badecker & Straub (1994). To address the possibility that the difference between our results and those of Badecker & Straub is due to a difference in the construction of materials, for example in the salience of the two potential referents, we conducted a second study of eye movements during reading that replicates their design. As mentioned above, Badecker & Straub (2002) argue for the multiple candidate effect based on evidence from sentences with one syntactically licensed and one unlicensed referent. However, they mention an earlier experiment (Badecker & Straub 1994), in which the design included an ambiguous condition with two syntactically licensed referents for a subsequent pronoun (see example (2)). Experiment 2 not only addresses the difference in salience asymmetry between the two potential referents, but it also allows us to test whether the difference in methods between our Experiment 1 and the Badecker & Straub studies, which used self-paced reading rather than eyetracking, could make the difference in whether evidence for competition is found or not.

3 Experiment 2: Partial replication of Badecker & Straub (1994)

Experiment 2 was conducted to determine, beyond our arguments from interpretation data, whether our lack of competition penalty for pronouns with multiple potential referents in Experiment 1 were due to a lack of salience of one candidate referent. Further, Experiment 2 tests the possibility that the competition penalty found by Badecker & Straub (1994; 2002) in self-paced reading may not be the same in eye movements during reading. Thus Experiment 2 presents a partial replication of the design of the Badecker & Straub (1994) study, which included two syntactically licensed antecedents for a pronoun in the ambiguous condition (see Table 5). The predictions for Experiment 2 are similar to those for the pronoun reference conditions in Experiment 2. If there is competition between potential referents for a pronoun, then we would expect the pronoun region (and potentially spillover regions) to be longer in the ambiguous reference condition than either of the highest referent or lowest referent conditions. If pronoun reference ambiguity is like structural ambiguity in its processing, then we should find an ambiguity advantage in Experiment 2 like that in Experiment 1. The structure of the items in Experiment 2 differs from Experiment 1 in the sense that one potential referent NP is the main clause subject, which should make that NP particularly salient as the referent of the subsequent pronoun. Because of this asymmetry in salience, one could further imagine that apart from the question of processing ambiguous reference, we would find an advantage for the unambiguous highest referent condition over the unambiguous lowest referent condition.

Table 5: Sample item set, Experiment 2. Slashes demarcate analysis regions. Indices are shown for the reader’s benefit but were not visible to participants.

Condition	Example
Highest referent	The young prince _i showed the revered queen _j / that he _{i,j} / would be/ a fine leader/ of the Tharassian empire.
Lowest referent	The revered queen _i showed the young prince _j / that he _{i,j} / would be/ a fine leader/ of the Tharassian empire.
Ambiguous reference	The young prince _i showed the revered king _j / that he _{i,j} / would be/ a fine leader/ of the Tharassian empire.

3.1 Method

3.1.1 Participants

85 participants were recruited from the University of Massachusetts community. Participants were self-reported native speakers of English who had normal or corrected-to-normal vision. Of these, 7 were excluded from further analysis because more than 25% of their data was lost during artifact rejection, and one additional participant was rejected due to poor question-answering accuracy (less than 65%), leaving 77 participants in the analyses reported below. Participants were tested individually, and received extra credit in either a Psychology course or a Linguistics course as compensation for their participation.

3.1.2 Materials

Twenty-four item sets were created following three of the conditions from Badecker & Straub (1994)'s design (see Table 5 for a sample item set). The fourth condition from Badecker & Straub's design, wherein none of the referents preceding the critical pronoun matched the pronoun's gender, is not relevant to our investigation and therefore was not included in the study. In each item, there were two possible referents introduced in the matrix clause. The critical pronoun was introduced in the subject position of an embedded clause that followed the direct object of the matrix clause. All referents introduced in the matrix clause were either definite descriptions that had definitional gender (e.g. *king*), or were proper names that were gender unambiguous (e.g. *Olivia*). Half of the items had two proper names, and the other half had two definite descriptions in the matrix clause. The pronoun itself was identical across all forms of a single item set; the referent of the pronoun, or whether it was ambiguous, was manipulated by changing the arguments inside the matrix clause (see 5). Half of the items had *he* as the critical pronoun, half had *she*.

For purposes of analysis, we defined three regions of interest. The pre-critical region consisted of the entire matrix clause. The critical region included the complementizer *that* and the critical pronoun, and the post-critical region included all material following the pronoun up to and including the verb (including any auxiliary verbs, if present).

3.1.3 Procedure

The procedure was largely identical to Experiment 1. In Experiment 2, only half of the items presented to participants contained a comprehension question. In addition, all twelve comprehension questions probed the comprehension of the pronoun (e.g. *Who would be a fine leader of the empire?*). In the Ambiguous Reference condition, both answers reflected grammatical potential interpretations of the pronoun (although salience might favour one over the other); in the remaining two conditions there was only one correct answer.

3.2 Results

3.2.1 Statistical analysis

We followed the same data rejection procedure as Experiment 1; trials with blinks or track losses on the critical region were excluded (6.3% of total data for included participants). The statistical analysis performed was identical to Experiment 1. The three conditions in Experiment 2 are parallel to the levels of the Ambiguity Type factor in Experiment 1. For this reason, we used the same Helmert coding as in Experiment 1 to create two contrasts: ambiguity and height.

3.2.2 Question accuracy

Table 6 summarizes the question-answering behavior in Experiment 2. Accuracy questions in the two unambiguous conditions was overall quite high, with an average accuracy rate of 90.4%. Accuracy did not differ significantly across these conditions. For ambiguous

questions, we found that answers were overall quite evenly balanced across the experiment with (49.9% high interpretations), suggesting that across the experiment there was not a substantial bias towards choosing either the higher or lower referent as the antecedent for an ambiguous pronoun.

3.2.3 Eye movement measures

We report eye movement measures for the pre-critical, critical, and post-critical regions. Raw times in milliseconds are presented in Table 7, and go-past time on the regions of interest are shown in Figure 2. As in Experiment 1, we report statistics based on the log values of each eye movement measure. A summary of fixed effects for each eye movement measure is shown in Table 8.

We observed no significant differences on the pre-critical region. On the critical region, total times were marginally longer ambiguous sentences. No other measure showed an effect on this region. However, in the post-critical region we observed a marginal effect of ambiguity in first pass time, and a significant effect in go-past time and total time. In all cases, this reflected faster reading times in the ambiguous condition. No significant effect of ambiguity was observed in regressions out; in re-reading time, the effect was marginal. In no measure did we see a significant effect of height.

3.3 Discussion

The results of Experiment 2 are straightforward to summarize: in three out of five reading measures (first pass, go-past, and total time) in the post-critical region, we saw an ambiguity advantage effect, with lower reading times for the ambiguous condition com-

Table 6: Mean accuracy (%) on pronoun-targeted comprehension questions over participants, Experiment 2. NB: For the ambiguous items, “correct” corresponded to accepting a high interpretation.

High	Low	Ambiguous
89.2 (1.9)	91.7 (1.5)	49.9 (3.3)

Table 7: By-participant means and standard errors in milliseconds, Experiment 2.

Measure	Condition	Pre-critical	Critical	Post-critical
First pass time (ms)	Ambig	1291 (40)	294 (10)	380 (16)
	High	1265 (40)	293 (11)	398 (17)
	Low	1327 (45)	305 (12)	418 (16)
Go-past time (ms)	Ambig	1299 (41)	334 (12)	455 (19)
	High	1271 (41)	330 (12)	479 (22)
	Low	1333 (45)	344 (13)	505 (21)
Total time (ms)	Ambig	1567 (63)	375 (14)	502 (19)
	High	1560 (58)	389 (16)	552 (23)
	Low	1607 (63)	398 (15)	553 (22)
Re-reading time (ms)	Ambig	620 (51)	315 (20)	354 (19)
	High	641 (46)	326 (17)	438 (23)
	Low	686 (43)	314 (13)	387 (20)
Proportion of regressions out	Ambig	0 (0)	0.07 (0)	0.11 (0)
	High	0 (0)	0.08 (0)	0.12 (0)
	Low	0 (0)	0.08 (0)	0.12 (0)

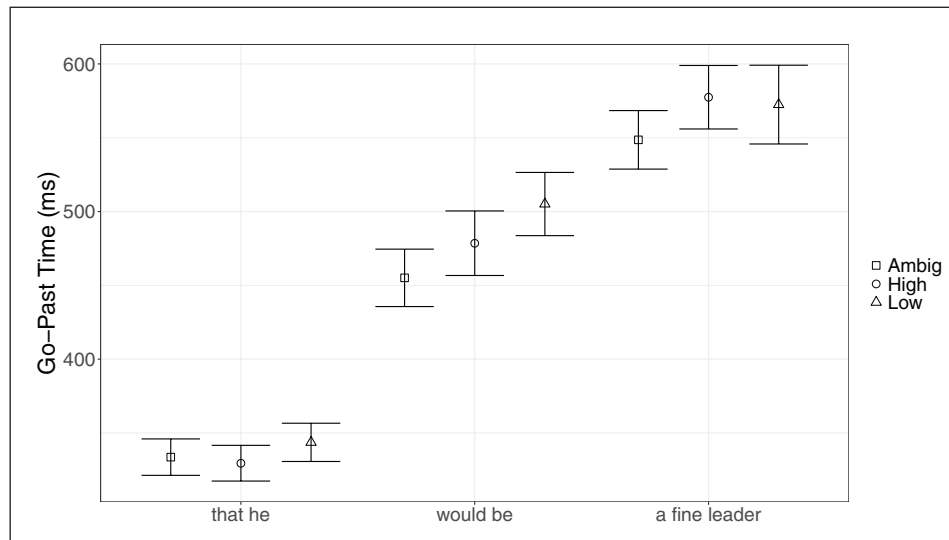


Figure 2: Go-past time in the critical (pronoun) region and two subsequent regions, Experiment 2. Bars indicate standard errors.

Table 8: Model parameters (fixed effect estimates (standard errors), *t*-values and *p*-values) for eye movement measures, Experiment 2.

Effect	First pass time								
	Pre-critical region			Critical region			Post-critical region		
	$\hat{\beta}$	<i>t</i>	<i>p</i>	$\hat{\beta}$	<i>t</i>	<i>p</i>	$\hat{\beta}$	<i>t</i>	<i>P</i>
Ambig	0.01 (0.02)	0.41	0.686	-0.03 (0.02)	-1.04	0.302	-0.06 (0.03)	-1.99	0.057
Height	-0.03 (0.02)	-1.46	0.158	-0.03 (0.02)	-1.19	0.239	-0.04 (0.03)	-1.41	0.173
Effect	Go-past time								
	Pre-critical region			Critical region			Post-critical region		
	$\hat{\beta}$	<i>t</i>	<i>p</i>	$\hat{\beta}$	<i>t</i>	<i>p</i>	$\hat{\beta}$	<i>t</i>	<i>P</i>
Ambig	0.01 (0.02)	0.42	0.678	-0.02 (0.03)	-0.72	0.476	-0.08 (0.03)	-2.38	0.023
Height	-0.03 (0.02)	-1.50	0.148	-0.02 (0.03)	-0.64	0.527	-0.05 (0.03)	-1.49	0.151
Effect	Total time								
	Pre-critical region			Critical region			Post-critical region		
	$\hat{\beta}$	<i>t</i>	<i>p</i>	$\hat{\beta}$	<i>t</i>	<i>p</i>	$\hat{\beta}$	<i>t</i>	<i>P</i>
Ambig	-0.01 (0.02)	-0.32	0.753	-0.05 (0.03)	-1.89	0.062	-0.09 (0.03)	-3.41	0.003
Height	-0.03 (0.03)	-1.09	0.285	-0.02 (0.03)	-0.84	0.403	-0.03 (0.04)	-0.85	0.405
Effect	Re-reading time								
				Critical region			Post-critical region		
				$\hat{\beta}$	<i>t</i>	<i>P</i>	$\hat{\beta}$	<i>t</i>	<i>P</i>
Ambig				-0.02 (0.05)	-0.43	0.671	-0.10 (0.06)	-1.75	0.085
Height				0.02 (0.07)	0.28	0.782	0.08 (0.07)	1.02	0.319

pared to either unambiguous condition. In re-reading time, this effect was marginal. We observed no evidence for any difference between high or low reference in our experiment. Thus, Experiment 2 provides additional evidence that there is an ambiguity advantage for pronominal processing, extending the findings of Experiment 1 to a new configuration: pronouns in subject position with syntactically prominent antecedents.

The observation of an ambiguity advantage effect in Experiment 2 serves to establish the generality of the ambiguity advantage for pronominal processing. If the ambiguity advantage in Experiment 1 were due to the lack of salience for the referents in that Experiment 1, then we would expect a different pattern of results in Experiment 2, in which the candidate referents were the main clause subject and object. In particular, we would have expected an ambiguity penalty, reflecting a costly competitive process for selecting an antecedent in the presence of two, relatively salient antecedents. However, we observed no evidence of an ambiguity penalty anywhere in our data; instead, we saw consistent evidence of an ambiguity advantage effect in the post-critical region.

In addition, the question answering data indicates that the ambiguity is fairly balanced across our materials, although the strength of this bias did vary item to item. This aligns well with the eye-tracking data, which found no evidence for differential processing of the high and low conditions. This pattern in itself is interesting, as one might have plausibly predicted an overall bias towards high reference in our experiment: the highest referent is the main clause subject, which is generally considered to be a likely referent for a subsequent pronoun (see e.g., Gernsbacher 1989). We hypothesize that the preference for a subject as a pronoun referent may have been diminished by the semantic relationships introduced by our choice of verbs in the study. Many of our verbs were verbs of transfer of information, such as *tell*, *inform*, or *notify* (see Appendix B for a full list of experimental materials used). It is possible that these verbs increased the salience of the object, as the receiver of the critical information (see Kaiser et al. 2009). On this view, the overall balance in the ambiguity across the experiment reflects a compromise between a subject bias, and a receiver bias; we note that this pattern is consistent with the results reported in Kaiser et al. (2009).

Experiment 2 provides additional evidence against a competition mechanism for the resolution of pronominal ambiguity during reading. Even with candidate referents that are salient, and would certainly be considered during the selection process under Badecker & Straub (2002)'s theory, we do not find a robust penalty for ambiguity.

4 General discussion

4.1 Similarity in processing across ambiguity types

The experiments we report here show a notable similarity in the processing of attachment and pronominal dependencies: for both, ambiguity speeds, rather than slows, reading. While many prior results in the literature would have led us to expect an ambiguity advantage in the case of attachment ambiguity, and an ambiguity penalty in the case of pronominal reference ambiguity, we find no evidence of this pattern in our experiments. In Experiment 1, we found an ambiguity advantage for both ambiguity types. Experiment 2 indicates that our failure to find an ambiguity penalty in our original studies is not likely to be due to the structures we chose for our materials. We believe that the similarities between the two ambiguity types point to a picture in which the parser's recognition and initial treatment of ambiguities in both attachment and pronoun reference shares a fundamental nature. Furthermore, our failure to find any evidence for competition effects in reading fails to support the predictions of competition-based constraint satisfaction models of sentence processing (MacDonald, Pearlmutter & Seidenberg 1994; but cf. Green & Mitchell 2006).

The observation that both structural and referential dependencies show an ambiguity advantage in similar contexts supports the view that the computational processes engaged to resolve both dependency types are highly similar. One possibility that is consistent with the results of Experiment 1 is that the unrestricted race model applies to resolving both

structural and referential ambiguities alike. On this view, different referents race to be selected as an antecedent for a pronoun, with the one that becomes available most quickly being adopted as the referent for the pronoun. If the antecedent that happens to win mismatches the pronoun's gender features, then an error signal is generated and the reference of the pronoun is re-evaluated. Under this model, the slower reading times on either of the unambiguous conditions reflect a penalty due to reanalysis: on some portion of trials, the processor selects the ultimately incorrect reference or attachment, which then must be reanalyzed. This interpretation of our results thus holds that both referential and structural dependencies are negotiated using a processor that is effectively serial, stochastic, and repair-driven. This conclusion aligns well with the literature on cue-based parsing, which holds that comprehenders must rely on retrieval of information from memory to form linguistic dependencies and support real-time sentence comprehension (for reviews, see Lewis & Vasishth 2005; Lewis et al. 2006; McElree 2006; Foraker & McElree 2011; Van Dyke & Johns 2012; Wagers & McElree 2013). In positing a noisy retrieval mechanism for the creation of linguistic dependencies, these models realize a processing model that has the key characteristics of the unrestricted race model: serial processing and stochastic selection of analyses (see discussion in Lewis & Vasishth 2005). It has been proposed that similar retrieval mechanisms subserve both referential and structural processing. The finding that pronominal reference patterns like structural attachment with respect to the ambiguity advantage effect is broadly consistent with this position.

An important point is that this explanation of our data ascribes the reading time slowdown in unambiguous pronominal conditions to a sort of garden path: the processor has selected the incorrect, gender-inappropriate antecedent, prompting reanalysis to identify an appropriate antecedent. This conclusion may seem surprising: unlike the PP attachment conditions, there is no portion of the string which is actually ambiguous. In our pronoun conditions, the pronoun was either ambiguous, or it wasn't.

In this sense, our pronoun conditions are similar to Experiment 2 of van Gompel et al. (2005). In this experiment, van Gompel and colleagues looked at RC attachment ambiguities that were immediately disambiguated, as in (3):

- (3) van Gompel et al. (2005)
- a. I read that the bodyguard of the governor retiring after the troubles is very rich. (globally ambiguous)
 - b. I read that the governor of the province retiring after the troubles is very rich. (high attachment)
 - c. I read that the province of the governor retiring after the troubles is very rich. (low attachment)

Even in sentences like (3b)–(3c), where the word that introduced the relative clause (*retiring*) also immediately disambiguated the attachment position, van Gompel and colleagues observed an ambiguity advantage effect. They offered an explanation of this finding in terms of the unrestricted race model, suggesting that on at least some trials, the inappropriate parse can win the race despite being incompatible with the semantic cues on *retiring*. In other words, the semantic cues are not available fast enough, or are not strong enough, to completely tilt the race in favor of the appropriate parse on every trial.

If the ambiguity advantage effect for pronouns reflects reanalysis on the unambiguous trials, it seems necessary to adopt a similar position: the gender cues on the pronoun must not be sufficient to categorically determine the outcome of the race on every trial. This could arise because the gender cues are not available quickly enough, or aren't strong

enough, to definitively settle the race in favor of the correct antecedent. This position has some measure of independent support. On the basis of several probe recognition studies, Greene et al. (1992) argued that gender cues were not automatically used to identify a referent for a pronoun. Instead, they argued that, under normal circumstances, pronouns simply attached to referents that were focused in the discourse. On their view, participants only used gender cues to identify a referent when the task demands rendered this necessary. Other reading studies, which we mentioned in the Introduction, have reported that the use of gender cues appears to depend on the task context. In one study, participants used gender cues during antecedent selection only if the content of the filler sentences reliably contained pronouns (Garnham et al. 1992), or the comprehension questions required participants to resolve the reference of the pronoun (Garnham et al. 1992; Rigalleau et al. 2004; Stewart et al. 2007). Further, van Gompel & Liversedge (2003) found delayed use of gender cues in an experiment examining cataphoric pronoun comprehension. However, at least some studies using the visual world methodology suggest that normal referential processing can make full and immediate use of gender cues on pronouns to find an antecedent (Arnold et al. 2000). Nonetheless, if the processor's real-time use of the pronoun's gender cues is less than perfect, then the processor could be effectively garden-pathed (i.e., it could select the wrong referent) in sentences that do not have any appreciable ambiguity.

Another possible explanation for our results is that readers are constructing *Good-Enough* representations of both structural attachments and pronominal reference to a similar extent, leading to an ambiguity advantage for both ambiguity types in Experiment 1. Swets et al. (2008) interpret the ambiguity advantage as underspecification of globally ambiguous attachment, with support from differences in reading-time patterns depending on an experimental task manipulation. Using a non-cumulative moving window self-paced reading paradigm, Swets et al. (2008) found an ambiguity advantage when comprehension questions were sparse and/or superficial, and a penalty for high attachment when comprehension questions were frequent and targeted the ambiguity. Swets et al. interpret their results as showing that when comprehension questions are sparse or superficial, readers underspecify attachment of the ambiguous constituent, in their case a relative clause. However, when comprehension questions targeted the ambiguity, the results support an interpretation under which the parser attaches incrementally, in this case with a preference for low attachment. Therefore, ambiguous sentences behave like low attachment in reading, with a penalty due to a garden path effect for high attachment. These results show that the processing of attachment ambiguities is sensitive to the experimental context, which includes all aspects of the procedure. For our results to be consistent with this Good-Enough account, one would have to claim that our experimental procedure was like the superficial or sparse question conditions of Swets et al. (2008). While it is an empirical question as to what proportion of comprehension questions must target a key ambiguity in order to induce more thorough processing (we thank a reviewer for pointing this out), we believe that it is not unreasonable to think that our 12 targeted questions out of 30 experimental items and 52 fillers (in Experiment 1) might not be enough to do so.

Recently, Karimi & Ferreira (2015) proposed an explanation of the ambiguity advantage in attachment that rests on Good-Enough *encoding* of the complex noun phrase (e.g., *the daughter of the colonel*), rather than Good-Enough attachment of a relative clause (we thank an anonymous reviewer for pointing this work out to us). Karimi and Ferreira suggest that readers merge the two nouns into a single, sketchy representation when the processing circumstances do not require more detailed encoding. They discuss the case of relative clause attachment as disambiguated by a reflexive, as in Traxler et al. (1998) among others. If readers encode complex noun phrases in a single syntactic or discourse

representation, then processing a reflexive that forces the division of this single entity into two should be more difficult than processing a reflexive that is compatible with the single representation (the globally ambiguous case; see also Albrecht & Clifton (1998), for a similar proposal on reference to coordinated NPs). In this way, the ambiguity advantage is predicted. This account would, we believe, also predict that a pronoun with ambiguous reference should be easier to process than an unambiguous pronoun, for sentences such as those in Experiment 1. To account for the ambiguity advantage in Experiment 2, Karimi and Ferreira would have to propose, however, that the two subject noun phrases in the matrix and embedded clause are encoded as one sketchy representation. We believe that this possibility is less likely, but it remains as an open research question.

One influential theory that proposes parallel activation of structural analyses, namely surprisal theory (Hale 2001; Levy 2008), has also been proposed to account for the ambiguity advantage in attachment. Levy (2008) discusses the results of Traxler et al. (1998) as an example of the ambiguity advantage. Levy shows that surprisal theory predicts that the globally ambiguous item (*The son of the colonel who shot himself...*) has lower surprisal because it is compatible with either of the high or low attachment analyses, and therefore receives essentially the sum of their two probabilities. Linguistic input with lower surprisal gives rise to faster processing times, leading to the ambiguity advantage effect. In principle one might be able to extend surprisal theory to reference, and predict the ambiguity advantage for pronoun assignment as well. This remains a question for future research.

At present, our results do not favour one theory, either the URM, Good-Enough Processing or surprisal theory. What we can conclude is that there is no convincing reason, based on our current data, to believe that the resolution of attachment and pronominal reference ambiguities are subserved by distinct mechanisms in the language processing system. There is no evidence from our eye-movement data that either dependency is subject to competition effects in processing; for both, the opposite seems to be true. Ambiguity facilitates structural and referential processing alike.

4.2 The apparent conflict between our results and the previous pronoun literature

The results of our experiments are in conflict with the previous literature reporting competition effects in pronoun reference (e.g., Badecker & Straub 1994; 2002). There are several possible reasons for this discrepancy we see at present. First, as we noted in the introduction, not all studies have found competition effects due to the presence of morphologically matching, but structurally inaccessible antecedents (Clifton et al. 1999; Chow et al. 2014). The difference in the observed results is unlikely to be due to differences in materials or procedure, as Chow et al. (Experiment 2) replicated the procedure and materials used by Badecker and Straub, but still failed to find competition effects. This raises the unsatisfactory possibility that previous reports of competition effects were actually Type I errors.

However, a possibility that strikes us as more likely is that the difference between our results and previous results stems from differences in the nature of the dependent measure and the task context. Recall that Stewart et al. (2007) found a penalty for ambiguous pronouns only in “deep processing” conditions. This task prompted participants to answer a question targeting the pronominal ambiguity after each trial. In a similar fashion, Garnham et al. (1992) and Rigalleau et al. (2004) found ambiguity penalties only when the task content encouraged participants to strategically use gender cues to resolve pronominal ambiguity. This may explain the differences between the results: it is possible that, despite our inclusion of some targeted questions, our task did not require the extent or type of strategic processes that these previous studies required. Alternatively, it is possible that our dependent measure (eye-tracking) is more robust to intrusion from strategic

processing than whole-sentence self-paced reading, allowing the ambiguity advantage for pronouns to surface in our reading time measures. We cannot decide between these possibilities at present. From our studies, we can only conclude that under the same experimental conditions, referential ambiguities and attachment ambiguities do in fact show a similar ambiguity advantage effect in reading measures.

We show in this paper that, when compared directly, structural and referential ambiguity show a similar pattern in processing in eye movements during reading. We do not claim that one should never find a penalty for ambiguity of either kind. For example, in a behavioural study, Logačev & Vasishth (2015) showed that, under the right experimental conditions, one could even find an ambiguity *disadvantage* for structural ambiguity. What we need then, is a model of ambiguity resolution that can capture the range of effects that researchers find in different tasks and experimental contexts. Here, we argue that more evidence is needed before we can assume that models of structural and referential ambiguity resolution should be separate.

4.3 Interpretive differences between attachment and pronoun reference

We have concluded that the computational mechanisms that underly the formation of referential and structural dependencies share an important similarity, and that as a consequence, readers react to referential and structural ambiguity in a similar fashion in processing. However, we cannot endorse the stronger claim that the two rely on exactly the same processing mechanism on the basis of our data. While this remains a logical possibility, it is also possible that there are two distinct cognitive mechanisms engaged for each dependency type that merely share similar computational properties. On this latter view, the referential and structural processing subsystems would be constrained by similar cognitive or computational principles, which would in turn lead to similar processing profiles (see discussion of this possibility in Lewis & Vasishth 2005).

Supporting this latter view, there are many differences in how referential ambiguity and structural ambiguity are resolved (Frazier & Clifton 2005). A model of ambiguity resolution that accounts for both attachment and pronoun reference ambiguities must allow for the different factors that influence comprehenders' ultimate interpretations of each ambiguity type, potentially giving rise to disparate preferences such as we found in the norming data for Experiment 1. For attachment, previous studies have found several factors that affect attachment of modifiers. For example, Hemforth et al. (2015) found that the length of a relative clause had an effect on attachment decisions across several languages, with a greater proportion of high attachment decisions for long relative clauses. For prepositional phrases, Spivey-Knowlton & Sedivy (1995) found that definiteness of an NP affected whether a PP was attached as a modifier of that NP or of the VP that contained it. A large body of research has also uncovered factors that make a particular referent more salient for the reference of a subsequent pronoun, including being in subject position and information structure influences such as topicality (Kaiser 2011). Discourse coherence factors also influence ultimate pronoun reference, for example the implicit causality encoded in a verb (Garvey & Caramazza 1974) among other factors (see e.g., Kehler 2008). We view the factors that influence the ultimate attachment or reference decision as a separate question from the mechanism by which an ambiguity is resolved. The Unrestricted Race Model allows for a number of influences at several levels of representation to influence attachment decisions. Depending on the linguistic and non-linguistic context of an utterance, the relative importance of various factors may vary in influencing an initial attachment decision. To suggest that a similar process could hold for pronoun reference does not necessarily mean that the factors influencing an initial decision must be identical to those influencing attachment.

5 Conclusions

We have reported the results of two experiments of eye movements during reading. Experiment 1 examined the processing of structural (PP-attachment) and referential (pronoun) ambiguities using eye movements during reading, and Experiment 2 sought to replicate previous results in pronominal reference ambiguity resolution Badecker & Straub (2002). In neither experiment do we find evidence for processing difficulty due to ambiguity of either type as compared to corresponding unambiguous materials. Rather, we find a processing advantage for both types of ambiguity. Importantly, we find no evidence for a pattern that has been commonly assumed based on separate studies of structural and referential ambiguity resolution, namely that there is no penalty (or potentially an advantage) for structurally ambiguous material but that there is a penalty for pronouns with multiple potential referents. This penalty has been interpreted as a competition effect for reference, whereas no competition between potential attachment sites occurs for attachment ambiguity. Our results suggest that the language processing system may react to each ambiguity type in a similar way, and that models of sentence processing should address this similarity.

Additional File

The additional file for this article can be found as follows:

- **Appendices.** Processing ambiguities in attachment and pronominal reference.
DOI: <https://doi.org/10.5334/gjgl.852.s1>

Ethics and Consent

All human subjects research in this paper was approved by the Institutional Review Board at the University of Massachusetts, protocol #2014-2181.

Acknowledgements

Much of this work was completed while the corresponding author held positions as a Postdoctoral Fellow at McGill University, under the advisement of Yosef Grodzinsky and Michael Wagner, and as a visiting Assistant Professor at the University of Toronto, and as a Post-Doctoral Researcher at the Humboldt-Universität zu Berlin funded by the Gottfried Wilhelm Leibniz-Preis 2014 awarded to Prof. Dr. Artemis Alexiadou (AL554/8-1). We are grateful for feedback from the reviewers on this manuscript as well as audiences at Architectures and Mechanisms for Language Processing XX at the University of Edinburgh and the 28th CUNY Conference on Human Sentence Processing at the University of Southern California. We thank Chuck Clifton for comments on an earlier draft of the paper. We would like to thank Vishal Arvindam, Russell Cavallaro, Amanda Doucette, and Christopher Hammerly for their help with data collection at the University of Massachusetts, Amherst, and Elena-Cristina Feraru for help at the University of Toronto. All errors and omissions remain our own.

Competing Interests

The authors have no competing interests to declare.

References

- Albrecht, Jason E. & Charles Clifton, Jr. 1998. Accessing singular antecedents in conjoined phrases. *Memory & Cognition* 26(3). 599–610. DOI: <https://doi.org/10.3758/BF03201166>
- Arnold, Jennifer E., Janet G. Eisenband, Sarah Brown-Schmidt & John C. Trueswell. 2000. The rapid use of gender information: evidence of the time course of pronoun resolu-

- tion from eyetracking. *Cognition* 76. B13–B26. DOI: [https://doi.org/10.1016/S0010-0277\(00\)00073-1](https://doi.org/10.1016/S0010-0277(00)00073-1)
- Badecker, William & K. Straub. 1994. Evidence that binding principles participate in a constraint satisfaction process. Poster session presented at the Seventh annual CUNY Sentence Processing Conference.
- Badecker, William & Kathleen Straub. 2002. The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory and Cognition* 28(4). 748–769. DOI: <https://doi.org/10.1037/0278-7393.28.4.748>
- Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68. 255–278. DOI: <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, Douglas M. 2005. Fitting linear mixed models in R: Using the lme4 package. *R News: The Newsletter of the R Project* 5(1). 27–30.
- Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In John R. Hayes (ed.), *Cognition and the development of language*, 279–362. New York: Wiley.
- Brysbaert, Marc & Don C. Mitchell. 1996. Modifier attachment in sentence processing: Evidence from Dutch. *The Quarterly Journal of Experimental Psychology* 49A(3). 664–695. DOI: <https://doi.org/10.1080/713755636>
- Carreiras, Manuel & Charles Clifton, Jr. 1993. Relative clause interpretation preferences in Spanish and English. *Language and Speech* 36. 353–372. DOI: <https://doi.org/10.1177/002383099303600401>
- Chomsky, Noam. 1981. *Lectures on government and binding*. The Netherlands: Dordrecht.
- Chow, Wing-Yee, Shevaun Lewis & Colin Phillips. 2014. Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology* 5. 1–16. DOI: <https://doi.org/10.3389/fpsyg.2014.00630>
- Clifton, Charles, Jr. & Adrian Staub. 2008. Parallelism and competition in syntactic ambiguity resolution. *Language and Linguistics Compass* 2(2). 234–250. DOI: <https://doi.org/10.1111/j.1749-818X.2008.00055.x>
- Clifton, Charles, Jr., Lyn Frazier & Patricia Deevy. 1999. Feature manipulation in sentence comprehension. *Rivista di Linguistica* 11(1). 11–39.
- Clifton, Charles, Jr. & Manuel Carreiras. 1999. Another word on parsing relative clauses: Eyetracking evidence from English and Spanish. *Memory & Cognition* 27(5). 826–833. DOI: <https://doi.org/10.3758/BF03198535>
- Cowles, H. Wind, Matthew Walenski & Robert Kluender. 2007. Linguistic and cognitive prominence in anaphor resolution: Topic, contrastive focus and pronouns. *Topoi* 26. 3–18. DOI: <https://doi.org/10.1007/s11245-006-9004-6>
- Crawley, Rosalind A., Rosemary J. Stevenson & David J. Kleinman. 1990. The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research* 19(4). 245–264. DOI: <https://doi.org/10.1007/BF01077259>
- Cuetos, Fernando & Don C. Mitchell. 1988. Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition* 30(1). 73–105. DOI: [https://doi.org/10.1016/0010-0277\(88\)90004-2](https://doi.org/10.1016/0010-0277(88)90004-2)
- Duffy, Susan A., Robin K. Morris & Keith Rayner. 1988. Lexical ambiguity and fixation times in reading. *Journal of Memory and Language* 27. 429–446. DOI: [https://doi.org/10.1016/0749-596X\(88\)90066-6](https://doi.org/10.1016/0749-596X(88)90066-6)
- Ferreira, Fernanda & Nikole D. Patson. 2007. The ‘Good Enough’ approach to language comprehension. *Language and Linguistics Compass* 1/1-2. 71–83. DOI: <https://doi.org/10.1111/j.1749-818X.2007.00007.x>

- Foraker, Stephani & Brian McElree. 2011. Comprehension of linguistic dependencies: Speed-accuracy tradeoff evidence for direct-access retrieval from memory. *Language and linguistics compass* 5(11). 764–783. DOI: <https://doi.org/10.1111/j.1749-818X.2011.00313.x>
- Frazier, Lyn. 1979. *On comprehending sentences: Syntactic parsing strategies*. Bloomington, IN: Indiana Linguistics Club.
- Frazier, Lyn. 1987. Sentence processing: A tutorial review. In *Attention and performance XII: The psychology of reading*, 559–586. London: Earlbaum.
- Frazier, Lyn & Charles Clifton, Jr. 2005. The syntax-discourse divide: processing ellipsis. *Syntax* 8(2). 121–174. DOI: <https://doi.org/10.1111/j.1467-9612.2005.00077.x>
- Frazier, Lyn & Keith Rayner. 1982. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology* 14(2). 178–210. DOI: [https://doi.org/10.1016/0010-0285\(82\)90008-1](https://doi.org/10.1016/0010-0285(82)90008-1)
- Garnham, Alan, Jane Oakhill & Hannah Cruttenden. 1992. The role of implicit causality and gender cue in the interpretation of pronouns. *Language and Cognitive Processes* 7(3–4). 231–255. DOI: <https://doi.org/10.1080/01690969208409386>
- Garvey, Catherine & Alfonso Caramazza. 1974. Implicit causality in verbs. *Linguistic Inquiry* 5(3). 459–464.
- Gaskell, M. Gareth & William D. Marslen-Wilson. 1997. Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes* 12(5/6). 613–656. DOI: <https://doi.org/10.1080/016909697386646>
- Gernsbacher, Morton Ann. 1989. Mechanisms that improve referential access. *Cognition* 32. 99–156. DOI: [https://doi.org/10.1016/0010-0277\(89\)90001-2](https://doi.org/10.1016/0010-0277(89)90001-2)
- Gilboy, Elizabeth, Josep María Sopena, Charles Clifton, Jr. & Lyn Frazier. 1995. Argument structure and association preferences in Spanish and English complex NPs. *Cognition* 54(2). 131–167. DOI: [https://doi.org/10.1016/0010-0277\(94\)00636-Y](https://doi.org/10.1016/0010-0277(94)00636-Y)
- Green, Georgia M. 1995. Ambiguity resolution and discourse interpretation. In Kees van Deemter & Stanley Peters (eds.), *Semantic ambiguity and underspecification*, chap. 1, 1–26. CSLI Publications.
- Green, Matthew J. & Don C. Mitchell. 2006. Absence of real evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language* 55(1). 1–17. DOI: <https://doi.org/10.1016/j.jml.2006.03.003>
- Greene, Steven B., Gail McKoon & Roger Ratcliff. 1992. Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory and Cognition* 18(2). 266–283. DOI: <https://doi.org/10.1037/0278-7393.18.2.266>
- Hale, John. 2001. A probabilistic earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the north american chapter of the association for computational linguistics*, 1–8. DOI: <https://doi.org/10.3115/1073336.1073357>
- Hemforth, Barbara, Lars Konieczny & Christoph Scheepers. 2000. Syntactic attachment and anaphor resolution: The two sides of relative clause attachment. In *Architectures and mechanisms for language processing*, chap. 11, 259–281. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511527210.012>
- Hemforth, Barbara, Susana Fernandez, Charles Clifton, Jr., Lyn Frazier, Lars Konieczny & Michael Walter. 2015. Relative clause attachment in German, English, Spanish and French: Effects of position and length. *Lingua* 166. 43–64. DOI: <https://doi.org/10.1016/j.lingua.2015.08.010>
- Kaiser, Elsi. 2011. Focusing on s: Consequences of subjecthood, pronominalisation, and contrastive focus. *Language and Cognitive Processes* 26(10). 1625–1666. DOI: <https://doi.org/10.1080/01690965.2010.523082>

- Kaiser, Elsi, Jeffrey T. Runner, Rachel S. Sussman & Michael K. Tanenhaus. 2009. Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition* 112(1). 55–80. DOI: <https://doi.org/10.1016/j.cognition.2009.03.010>
- Karimi, Hossein & Fernanda Ferreira. 2015. Good-enough linguistic representations and online cognitive equilibrium. *The Quarterly Journal of Experimental Psychology* 69(5). 1013–1040. DOI: <https://doi.org/10.1080/17470218.2015.1053951>
- Kehler, Andrew. 2008. Rethinking the SMASH approach to pronoun interpretation. In Jeanette K. Gundel & N. Hedberg (eds.), *Reference: Interdisciplinary perspectives*, 95–122. New York: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780195331639.003.0005>
- Kennison, Shelia M. & Jessie L. Trofe. 2003. Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research* 32(3). 355–378. DOI: <https://doi.org/10.1023/A:1023599719948>
- Konieczny, Lars, Barbara Hemforth, Christoph Scheepers & Gerhard Strube. 1997. The role of lexical heads in parsing: Evidence from German. *Language and Cognitive Processes* 12(2/3). 307–348. DOI: <https://doi.org/10.1080/016909697386871>
- Kuznetsova, A., P. B. Brockhoff & R. H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(12). 1–26. DOI: <https://doi.org/10.18637/jss.v082.i13>
- Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106. 1126–1177. DOI: <https://doi.org/10.1016/j.cognition.2007.05.006>
- Lewis, Richard L. & Shravan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive science* 29(3). 375–419. DOI: https://doi.org/10.1207/s15516709cog0000_25
- Lewis, Richard L., Shravan Vasishth & Julie A. Van Dyke. 2006. Computational principles of working memory in sentence comprehension. *Trends in cognitive sciences* 10(10). 447–454. DOI: <https://doi.org/10.1016/j.tics.2006.08.007>
- Logačev, Pavel & Shravan Vasishth. 2015. A multiple-channel model of taskdependent ambiguity resolution in sentence comprehension. *Cognitive Science*, 1–33. DOI: <https://doi.org/10.1111/cogs.12228>
- Luke, Steven G. 2017. Evaluating significance in linear mixed-effects models. *Behavior Research Methods* 49(4). 1494–1502. DOI: <https://doi.org/10.3758/s13428-016-0809-y>
- MacDonald, Maryellen C. & Brian MacWhinney. 1990. Measuring inhibition and facilitation for pronouns. *Journal of Memory and Language* 29. 469–492. DOI: [https://doi.org/10.1016/0749-596X\(90\)90067-A](https://doi.org/10.1016/0749-596X(90)90067-A)
- MacDonald, Maryellen C., Neal J. Pearlmutter & Mark S. Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101(4). 676–703. <http://lcnl.wisc.edu/publications/archive/7.pdf>. DOI: <https://doi.org/10.1037/0033-295X.101.4.676>
- McElree, Brian. 2006. Accessing recent events. *Psychology of Learning and Motivation* 46. 155–200. DOI: [https://doi.org/10.1016/S0079-7421\(06\)46005-9](https://doi.org/10.1016/S0079-7421(06)46005-9)
- McRae, Ken, Michael J Spivey-Knowlton & Michael K Tanenhaus. 1998. Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language* 38(3). 283–312. DOI: <https://doi.org/10.1006/jmla.1997.2543>
- R Core Team. 2018. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. <http://www.R-project.org>.
- Rigalleau, François & David Caplan. 2000. Effects of gender marking in pronominal coindexation. *The Quarterly Journal of Experimental Psychology* 53A(1). 23–53. DOI: <https://doi.org/10.1080/713755884>

- Rigalleau, François, David Caplan & Vanessa Baudiffier. 2004. New arguments in favour of an automatic gender pronominal process. *The Quarterly Journal of Experimental Psychology* 57A(5). 893–933. DOI: <https://doi.org/10.1080/02724980343000549>
- Rodd, Jennifer, M. Gareth Gaskell & William D. Marslen-Wilson. 2002. Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language* 46. 245–266. DOI: <https://doi.org/10.1006/jmla.2001.2810>
- Spivey, Michael J. & Michael K. Tanenhaus. 1998. Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24(6). 1521. DOI: <https://doi.org/10.1037/0278-7393.24.6.1521>
- Spivey-Knowlton, Michael & Julie C. Sedivy. 1995. Resolving attachment ambiguities with multiple constraints. *Cognition* 55(3). 227–267. DOI: [https://doi.org/10.1016/0010-0277\(94\)00647-4](https://doi.org/10.1016/0010-0277(94)00647-4)
- Staub, Adrian & Keith Rayner. 2007. Eye movements and on-line comprehension processes. In M. Gareth Gaskell (ed.), *The Oxford Handbook of Psycholinguistics*, 327–342. Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780198568971.013.0019>
- Stewart, Andrew J., Judith Holler & Evan Kidd. 2007. Shallow processing of ambiguous pronouns: Evidence for delay. *The Quarterly Journal of Experimental Psychology* 60(2). 1680–1696. DOI: <https://doi.org/10.1080/17470210601160807>
- Sturt, Patrick. 2013. Syntactic constraints on referential processing. In Roger P. G. van Gompel (ed.), *Sentence processing* (Current Issues in the Psychology of Language), chap. 6, 136–159. Psychology Press.
- Swets, Benjamin, Timothy Desmet, Charles Clifton, Jr. & Fernanda Ferreira. 2008. Under-specification of syntactic ambiguities: Evidence from self-paced reading. *Memory & Cognition* 36(1). 201–216. DOI: <https://doi.org/10.3758/MC.36.1.201>
- Swinney, David A. 1979. Lexical access during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior* 18(6). 645–659. DOI: [https://doi.org/10.1016/S0022-5371\(79\)90355-4](https://doi.org/10.1016/S0022-5371(79)90355-4)
- Traxler, Matthew J., Martin J. Pickering & Charles Clifton, Jr. 1998. Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language* 39. 558–592. DOI: <https://doi.org/10.1006/jmla.1998.2600>
- Van Dyke, Julie A. & Clinton L. Johns. 2012. Memory interference as a determinant of language comprehension. *Language and Linguistics Compass* 6(4). 193–211. DOI: <https://doi.org/10.1002/lnc3.330>
- van Gompel, Roger P. G., Martin J. Pickering, Jamie Pearson & Simon P. Liversedge. 2005. Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language* 52. 284–307. DOI: <https://doi.org/10.1016/j.jml.2004.11.003>
- van Gompel, Roger P. G., Martin J. Pickering & Matthew J. Traxler. 2000. Unrestricted race: A new model of syntactic ambiguity resolution. In Alan Kennedy, Ralph Radach, Dieter Heller & Joël Pynte (eds.), *Reading as a perceptual process*, 621–648. Oxford: Elsevier. DOI: <https://doi.org/10.1016/B978-008043642-5/50029-2>
- van Gompel, Roger P. G., Martin J. Pickering & Matthew J. Traxler. 2001. Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language* 45. 225–258. DOI: <https://doi.org/10.1006/jmla.2001.2773>
- van Gompel, Roger P. G. & Simon P. Liversedge. 2003. The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29. 128–139. DOI: <https://doi.org/10.1037/0278-7393.29.1.128>

- Vosse, Theo & Gerard Kempen. 2009. In defense of competition during syntactic ambiguity resolution. *Journal of Psycholinguistic Research* 38(1). 1–9. DOI: [https://doi.org/10.1016/S0010-0277\(00\)00063-9](https://doi.org/10.1016/S0010-0277(00)00063-9)
- Wagers, Matthew & Brian McElree. 2013. Working memory and language processing: Theory, data and directions for future research. In C. Boeckx & K. Grohmann (eds.), *The Cambridge handbook for biolinguistics*, Cambridge University Press.

How to cite this article: Grant, Margaret, Shayne Sloggett and Brian Dillon. 2020. Processing ambiguities in attachment and pronominal reference. *Glossa: a journal of general linguistics* 5(1): 77.1–30. DOI: <https://doi.org/10.5334/gjgl.852>

Submitted: 09 November 2018 **Accepted:** 11 February 2020 **Published:** 28 July 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Glossa: a journal of general linguistics is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS The Open Access logo, which is a stylized 'O' with a person icon inside.