The effect of syntactic constraints on the processing of backwards anaphora

Nina Kazanina [a,*], Ellen F. Lau [b], Moti Lieberman [c],
Masaya Yoshida [b], Colin Phillips [b,d]

[a] *Department of Linguistics, University of Ottawa, 401 Arts Hall, 70 Laurier Avenue East, Ottawa, Ont., KIN 6N5, Canada*
[b] *Department of Linguistics, University of Maryland, College Park, MD 20742, USA*
[c] *Department of Linguistics, McGill University, 1085 Dr. Penfield, Montreal, QC H3A 1A7, Canada*
[d] *Neuroscience and Cognitive Science Program, University of Maryland, College Park, MD 20742, USA*

Abstract

This article presents three studies that investigate when syntactic constraints become available during the processing of long-distance backwards pronominal dependencies (*backwards anaphora* or *cataphora*). Earlier work demonstrated that in such structures the parser initiates an active search for an antecedent for a pronoun, leading to gender mismatch effects in cases where a noun phrase in a potential antecedent position mismatches the gender of the pronoun [Van Gompel, R. P. G. & Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 128–139]. Results from three self-paced reading studies suggest that structural constraints on coreference, in particular Principle C of the Binding Theory [Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht, Foris], exert an influence at an early stage of this search process, such that gender mismatch effects are elicited at grammatically licit antecedent positions, but not at grammatically illicit antecedent positions. The results also show that the distribution of gender mismatch effects is unlikely to be due to differences in the predictability of different potential antecedents. These findings suggest that backwards anaphora dependencies are processed with a grammatically constrained active search mechanism, similar to the mechanism used to process another type of long-distance dependency, the *wh* dependency (e.g., [Stowe, L. (1986). Evidence for online gap creation. *Language and Cognitive Processes, 1,* 227–245; Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: an eye-tracking study. *Journal of Memory and Language*, *35,* 454–475.]). We suggest that the temporal priority for syntactic information observed here reflects the predictability of structural information, rather than the need for an architectural constraint that delays the use of non-syntactic information.
© 2006 Elsevier Inc. All rights reserved.

Keywords: Parsing; Coreference; Binding; Cataphora; Grammatical constraints; Pronouns

---

* Corresponding author. Fax: +1 613 562 5141.
 E-mail address: ninaka@uottawa.ca (N. Kazanina).

Introduction

To understand the architecture of the human language processor it is important to establish the extent to which its behavior is governed by general mechanisms or by highly specific subroutines that are restricted to individual constructions or individual languages. In this article we address the question of the generality of sentence processing mechanisms by examining whether well-known properties of the processing of long-distance *filler-gap dependencies* are also found in the processing of backwards pronominal dependencies (*backwards anaphora* or *cataphora*), as found in sentences like *When she was in Paris, Susan visited the Louvre*. Backwards anaphoric dependencies parallel filler-gap dependencies in terms of the relative ordering of key elements, but differ in a number of other respects, including the obligatoriness and overtness of the elements of the dependency and the syntactic and discourse constraints that they are subject to. An improved understanding of how backwards anaphora is processed may contribute to a fuller understanding of the mechanisms responsible for the construction of long-distance relations in language in general.

Much previous research on filler-gap dependencies such as *wh* interrogatives, topicalizations and scrambling constructions has established two generalizations that form the starting point for the present studies. First, much evidence indicates that the parser constructs filler-gap dependencies *actively*, i.e., after encountering a suitable filler (e.g., a *wh*-phrase) it constructs gap positions without waiting for unambiguous evidence for the position of the gap in the input (e.g., Crain & Fodor, 1985; Frazier & Clifton, 1989; Stowe, 1986). Second, the mechanisms that search for potential gap positions show on-line sensitivity to grammatical constraints on filler-gap dependencies, such that active gap construction is not observed in structural positions that are ruled out by those constraints (e.g., Stowe, 1986; Traxler & Pickering, 1996). In this study we investigate whether these two properties are also found in the processing of backwards anaphora. Previous studies of the processing of backwards anaphora provide some evidence for active dependency formation mechanisms (Van Gompel & Liversedge, 2003) and for the impact of grammatical constraints (Cowart & Cairns, 1987), but in both cases the findings are open to alternative interpretations. In the current study we examine these issues in more detail, with a particular focus on distinguishing the effects of grammatical constraints from the effects of distributional probabilities on active dependency formation.

Properties of long-distance dependencies

The *wh*-constructions in (1) are examples of a long-distance dependency that create a relation between a *wh*-phrase, also known as a filler, and its gap position, as in (1a). Although the filler and the gap may potentially be separated by an indefinite amount of intervening material (1b–c), the occurrence of the filler reliably predicts the occurrence of a gap within the same sentence and if there is no gap, the result is unacceptable (1d). Note that we use the filler-gap terminology here mostly for the purposes of exposition. The long-standing controversy about whether *wh*-dependencies genuinely involve gaps or whether they involve direct associations between verbs and *wh*-phrases (for reviews see Phillips & Wagers, in press; Pickering, 1993) is orthogonal to the processing issues that are the focus of this article.

(1) a. What did John see __?
     b. What did Mary say that John saw __?
     c. What did Bill say that Mary claimed that John saw __?
     d. *What did John see an apple?

The relative positioning of the two elements of a *wh*-dependency is subject to a number of constraints on well-formedness, often known as *island constraints*. The sentences in (2), for example, are unacceptable due to a ban on filler-gap dependencies that cross relative clause boundaries as in (2a) or that span non-complement adverbial clauses as in (2b) (Chomsky, 1973; Huang, 1982; Ross, 1967).

(2) a. *What did the boy that bought __ like Mary?
     b. *What did the boy read a book while his sister was playing with __?

Referential dependencies between a pronoun and an antecedent noun phrase include forwards anaphora, in which the antecedent precedes the pronoun (3a), and backwards anaphora, in which the pronoun precedes the antecedent (3b). Similar to filler-gap dependencies, referential dependencies can span multiple clauses, as in (3a), and are subject to structural restrictions. Most relevant for the current study is a constraint, *Principle C* (Chomsky, 1981), that prohibits referential dependencies in which the pronoun structurally c-commands the antecedent, i.e., where the antecedent is structurally within the scope of the pronoun (4).

(3) a. John$_i$ said that the news report would make the customers think that he$_i$ wanted to sell the company.
     b. Although she$_i$ was sleepy, Susan$_i$ tried to pay attention.

(4) *He$_i$ said that John$_i$ saw the movie.
    intended meaning: 'John said that he
    (=John) saw the movie.'

In research on language processing, the constraints on long-distance filler-gap dependencies have received a good deal of attention, but there has been relatively little research on the processing of constraints on backwards anaphora.

Real time processing of wh-dependencies

Fodor (1978) sketched out two potential mechanisms for parsing filler-gap dependencies, contrasting a 'gap-driven' view whereby dependency formation is initiated only when the parser has encountered all necessary pieces of information in the input, that is, both the filler and the gap, with a 'filler-driven' view, according to which the parser starts predictively building a dependency upon encountering the filler that constitutes the first half of the dependency. Stowe (1986) explored this distinction using a word-by-word self-paced reading technique, by comparing reading times for sentences that contain a licit *wh*-dependency (5a) with controls that lack a *wh*-dependency (5b).

(5) a. My brother wanted to know
        who Ruth will bring us home to __ at
        Christmas.
     b. My brother wanted to know if Ruth
        will bring us home to Mom at Christmas.

Stowe found that readers slowed down at the direct object noun phrase *us* in *wh*-sentences, relative to the corresponding word in the control sentence, and argued that this effect arose because the parser predictively posited a gap in direct object position, which subsequently needed to be retracted upon encountering the overt pronoun *us*. This 'filled-gap effect' is taken to support the 'filler-driven' view and to reflect the parser's willingness to complete a *wh*-dependency before it has unambiguous bottom-up evidence that the verb has a missing argument. Such a mechanism for filler-gap dependency formation was dubbed as *active search*, and has been supported by evidence from similar filled-gap effects in diverse syntactic positions (Aoshima, Phillips, & Weinberg, 2004; Crain & Fodor, 1985; Lee, 2004), and by evidence using different experimental measures, including eye-tracking (Sussman & Sedivy, 2003; Traxler & Pickering, 1996), event-related potentials (Garnsey, Tanenhaus, & Chapman, 1989; Kaan, Harris, Gibson, & Holcomb, 2000), and cross-modal lexical priming (Nicol, Fodor, & Swinney, 1994).

The evidence for active *wh*-dependency formation leads to a further question about the grammatical accuracy of the parser. Whereas a gap-driven parsing mechanism could be expected to posit gaps only in grammatically acceptable positions, by virtue of its dependence on direct evidence in the language input, this is not guaranteed for an active dependency formation mechanism, since it is predictive and less closely tied to evidence in the bottom-up input. If the process of active dependency formation immediately adheres to grammatical constraints, the parser should not attempt to posit a gap inside an island, despite its preference to complete the dependency as soon as possible. In her second experiment, Stowe (1986) found that no filled-gap effect was observed inside a syntactic island, suggesting that the parser respects constraints on *wh*-movement even at the earliest stages of filler-gap dependency formation (see also Bourdages, 1992; Phillips, 2006; Traxler & Pickering, 1996 for a similar conclusion and Pickering, Barton, & Shillcock, 1994 for discussion of potentially discrepant findings).

Referential dependencies

Backward anaphoric dependencies parallel filler-gap dependencies in the respect that they may span long-distances, and in the respect that a dependent element (*wh*-phrase, pronoun) precedes a controlling element (verb or gap, antecedent noun phrase). This raises the possibility that the parser may analyze backward anaphoric dependencies in the same fashion as filler-gap dependencies, searching for an antecedent for an unanchored pronoun in a manner that is active yet sensitive to constraints on referential dependencies. However, unlike *wh*-fillers that require a gap position, finding an overt antecedent is not a grammatical requirement for a pronoun, which can be taken to refer to an unspecified discourse referent. For this reason, an active search mechanism may not be as useful for backwards anaphora as for filler-gap dependencies. Furthermore, if common mechanisms are used for parsing forwards and backwards referential dependencies, then this is another reason why active search might not be deployed, since forward anaphoric dependencies cannot be recognized until both the antecedent and the pronoun have been encountered.

If pronouns do trigger an active search for an antecedent, this leads to the question of whether binding principles immediately constrain active search for antecedents. If binding principles, like island constraints on filler-gap dependencies, constrain the earliest stages of structure generation, we should expect that during an active search for an antecedent triggered by a cataphoric pronoun the parser should never consider positions that would violate the binding constraints. Note that our use

of the term 'antecedent' in referring to a noun phrase does not necessarily entail that that noun phrase is a licit licensor for the dependent element in question. Instead it will be used as a shortcut for 'a candidate antecedent' which *could* become a licit antecedent provided that (i) it occurs in a licit structural position, and (ii) it matches the pronoun in its morphological features (gender, number).

Active search in backwards anaphora processing

Previous evidence for active search for antecedents in the processing of backwards anaphora comes from a study by Van Gompel and Liversedge (2003) that used eye-tracking to examine reading times for bi-clausal sentences as in (6). The gender of the main subject was manipulated such that it either matched (6a) or mismatched (6b) with the preceding pronoun. In all sentences an antecedent for the pronoun appeared at some point in the main clause.

(6) Stimuli from Experiment 1, Van Gompel and Liversedge (2003).
   a. *gender match*
      When <u>he</u> was at the party, <u>the boy</u> cruelly teased the girl during the party games.
   b. *gender mismatch*
      When <u>he</u> was at the party, <u>the girl</u> cruelly teased the boy during the party games.
   c. *control*
      When <u>I</u> was at the party, <u>the boy</u> cruelly teased the girl during the party games.

Van Gompel and Liversedge found a mismatch effect in early eye-tracking measures at the region immediately following the main clause subject noun: first-pass reading times were slower at *cruelly* in (6a) than in (6b). Furthermore, inclusion of the control condition in (6c) made it possible to rule out the possibility that the longer reading times following the second subject in (6b) were due to the introduction of a new discourse entity, since reading times in the critical region in the control condition (6c) did not differ from (6a). Van Gompel and Liversege instead argue that this *gender mismatch effect* reflects formation of a referential dependency between the pronoun and the second subject noun phrase before relevant bottom-up information about the antecedent has been taken into consideration, which leads to processing difficulty when the antecedent noun phrase is recognized as semantically incompatible with the pronoun. This evidence parallels the filled-gap effect found in filler-gap dependency processing, and may be considered as evidence for active formation of backwards anaphoric dependencies

(for similar findings from Japanese see Aoshima, 2003; Aoshima, Yoshida, & Phillips, submitted for publication).

Note that a number of different processing mechanisms for backward anaphora may be considered as 'active', in the respect that dependencies are constructed before all bottom-up information has been analyzed. Van Gompel and Liversedge (2003) assume that the syntactic category of the antecedent must be processed before a referential dependency can be formed, whereas computation of its morphological properties 'is delayed until after the computation of coreference relations' (p. 128), and thus argue that syntactic category information has an architectural priority in parsing (e.g., Cowart & Cairns, 1987; Fodor, 1983; Frazier, 1987). However, Van Gompel and Liversedge's findings are also compatible with an even more 'active' processing mechanism that constructs referential dependencies as soon as an antecedent position can be reliably predicted, potentially before any of the specific features of the antecedent are encountered in the input. Under this approach, the parser may recognize that a sentence-initial *while*-clause, as in (6), must be followed by a main clause that contains a subject noun phrase, and may use this knowledge to construct a dependency between the pronoun and the main clause subject even before the end of the *while*-clause. If this is the case, then the gender mismatch effect may reflect the fact that a subject position may be reliably projected ahead of time as a structurally licit antecedent, whereas its gender and other features may only become known once the subject noun phrase is encountered bottom-up. This view would obviate the need to impose an architectural constraint that delays the availability of morphological information, as Van Gompel and Liversedge propose.

The gender mismatch paradigm therefore provides a useful method for tracking active dependency formation in backwards anaphora, but leaves open the question of how 'active' the search for pronoun antecedents is, and also how sensitive it is to constraints on referential dependencies.

Processing of constraints on forwards anaphora

A number of previous studies explored the time-course of the application of binding constraints on forwards anaphora constructions in which the pronoun follows the antecedent, in particular the requirement that reflexives find a clausemate antecedent (7) and the requirement that pronouns not have a clausemate antecedent (8), known as *Principle A* and *Principle B* respectively (Chomsky, 1981; Reinhart & Reuland, 1993).

(7)   a. Susan heard that Bill_i had painted himself_i.
      b. *Bill_i heard that Susan had painted himself_i.
(8)   a. *Susan heard that Bill_i had painted him_i.
      b. Bill_i heard that Susan had painted him_i.

Evidence has been found for both early/immediate and late/delayed application of these constraints. Nicol and Swinney (1989) examined the activation of candidate antecedents at a pronoun or reflexive using a cross-modal priming task and found that candidates excluded by Conditions A and B were not considered (see also Clifton, Kennison, & Albrecht, 1997). In contrast, Badecker and Straub (2002) showed in a self-paced reading task that reading times for a pronoun were affected by the gender congruency of a grammatically inaccessible antecedent, and concluded from this that the application of grammatical constraints on binding was delayed relative to early parsing processes (see also Kennison, 2003). More recently, Sturt (2003) found in eye-tracking studies on the processing of reflexives that early eye-tracking measures (first-fixation and first-pass) were unaffected by the gender congruency of grammatically inaccessible antecedents, although some later eye-tracking measures (second-pass, regression path) did show effects of those antecedents. Sturt's findings raise the possibility that discrepancies among earlier findings might have been due to differences in the stage of processing that different tasks tapped into. This parallels findings on the processing of filler-gap dependencies, where evidence for the construction of illicit dependencies comes primarily from studies that have used whole sentence judgment tasks.

A difficulty in interpreting effects of grammatical constraints on the processing of forwards anaphora arises from the fact that the pronoun or reflexive cannot, in general, be anticipated, and therefore the search for an antecedent cannot be initiated until the end of the dependency is reached. Consequently, the search for an antecedent involves a search in memory for candidate antecedents. If there are multiple candidate antecedents, they may need to be evaluated in parallel. Studies of binding constraints in processing have asked whether the parser selectively searches structural positions where antecedents are allowed, but this question overlaps with the question of what type of memory representation is searched. The pieces of a sentence may be encoded in echoic memory, in a syntactic parse, or in a discourse model, and may to some degree be stored in all of these simultaneously. Therefore, evidence for effects of grammatically inaccessible antecedents on the processing of pronouns or reflexives may reflect processes that search illegal antecedent positions in a syntactic structure, or they may reflect processes that access a non-syntactic encoding of potential antecedents. Furthermore, in cases where the effect of structurally inaccessible antecedents is realized as a processing disruption due to the presence of two candidate antecedents (e.g., Badecker & Straub, 2002; Sturt, 2003), this may reflect interference effects at a lexical or conceptual level that make a grammatically accessible antecedent harder to retrieve from memory when it is similar to another noun phrase in the sentence. These difficulties do not arise in the processing of backwards anaphora, since the dependency can be anticipated after processing of its first element, i.e., the pronoun, and therefore each potential antecedent can be evaluated in succession. It is therefore possible to investigate, for a series of different structural positions, whether the search for an antecedent for the pronoun is sensitive to grammatical constraints on backwards anaphora. This may be viewed as a particularly strong test of the impact of binding constraints in parsing, since successful constraint application requires that the parser ignore a candidate antecedent at a point where no other antecedent is yet available.

Principle C in grammar and processing

Although backwards anaphora is productive and fully acceptable in English (see van Hoek, 1997 for multiple naturalistic examples), it is excluded in configurations where a pronoun c-commands its antecedent, as in the examples in (9), where the noun phrases *he* and *John* cannot be understood as coreferential. This constraint, known as Principle C (Chomsky, 1981), is highly robust across languages (e.g., Baker, 1991; Jelinek, 1984; but cf. Speas, 1990; Bruening, 2001) and appears early in language development (Crain & McKee, 1985; Kazanina & Phillips, 2001).

(9)   a. *He_i likes John_i.
      b. *He_i said that John_i likes wine.
      c. *He_i drank beer while John_i watched a
         soccer game.

It should be noted that there are at least two well-defined classes of apparent exceptions to Principle C, which must be taken into consideration when designing studies on this constraint. The first type of exception can be seen in a sentence like *He then did what John always did in such situations*, where coreference between *he* and *John* is acceptable. Such cases are generally understood to involve comparisons of multiple mental representations or 'guises' of the same individual (e.g., Heim, 1992; Reinhart, 1983), and minimal changes to the situations under comparison can lead to failure of coreference, as in *He then did what John had done half an hour earlier*. Rather than undermine the structural account of Principle C, such contrasts suggest that we should understand coreference in terms

of elements in a mental model, which may stand in a many-to-one relation with entities in the world. A second type of exception involves sentences like *He was threatening to leave when Billy noticed that the computer had died* (Harris & Bates, 2002). Such cases are clearly acceptable, and are most effective in descriptions of scenarios where the embedded clause event interrupts the main clause event. Again, minimal changes make coreference unacceptable, as in *He decided to leave when Billy noticed that the computer had died*. Syntactic tests suggest, however, that the *when*-clause in such exceptional examples is a sentential modifier, rather than a verb phrase-subordinator, which takes away the problematic c-command relation between the pronoun and the embedded subject and obviates the Principle C violation. In sum, both apparent exceptions to Principle C may not ultimately be counterexamples to the structural generalization (see Kazanina, 2005 for more detail). Yet, in all of the studies reported below care was taken to avoid such exceptions in constructing the experimental materials and the acceptability of all contrasts under investigation was verified by off-line questionnaires.

Previous psycholinguistic investigation of Principle C in sentence processing has been limited. In a whole-sentence self-paced reading Hirst and Brill (1980) looked for effects of an inaccessible antecedent in constructions that contained a potential Principle C violation using two sentence sequences such as (10), that contained both a grammatically accessible antecedent for the pronoun (*John*) and a grammatically inaccessible one (*Henry*).

(10) <u>John</u> stood watching. <u>He</u> ran for a doctor after <u>Henry</u> fell down some stairs.

Hirst and Brill found that reading times for the second sentence differed as a function of the plausibility of the inaccessible antecedent (*Henry*) as the referent for the pronoun. They concluded that during pronoun resolution speakers temporarily consider grammatically illicit referents. However, the coarse temporal resolution of the task makes it difficult to draw inferences regarding the time-course and the source of the effect. Specifically, the effect could be due to a later stage of processing at which information from multiple sources is combined and the sentence is evaluated for overall plausibility.

Cowart and Cairns (1987) used sentence fragments such as (11)–(12) to investigate the processing of backwards anaphora and the application of Principle C.

(11) a. Whenever they lecture during the procedure, charming babies is...
　　 b. Whenever you lecture during the procedure, charming babies is...

(12) a. If they want to believe that visiting uncles is...
　　 b. If you want to believe that visiting uncles is...

The first phrase of the 2nd clause (*charming babies, visiting uncles*), which is ambiguous between a noun phrase and a gerund, is disambiguated by the number marking on the following auxiliary in favor of the gerund parse. Cowart and Cairns investigated whether this ambiguity was affected by a pronoun subject in a preceding clause. If the pronoun initiates a search for an antecedent then this may increase the likelihood that the ambiguous phrase should be analyzed as a noun phrase in (11a)–(12a), relative to (11b)–(12b), which contain an indexical pronoun. In either condition there was an additional restriction against treating the ambiguous phrase as the antecedent of the pronoun: in (11a) it is a semantically implausible antecedent for the pronoun; in (12a) it is a grammatically inaccessible antecedent, due to Principle C. In a naming task, Cowart and Cairns found longer response latencies for *is* in (11a) relative to (11b), whereas no difference was found in (12a) vs. (12b). They concluded that the parser was more biased towards the noun phrase analysis when it needed to link the pronoun with an antecedent, but only if that antecedent was in a grammatically accessible position. Further, they interpreted the contrast between (11) and (12) to suggest that semantic restrictions, unlike syntactic restrictions, do not have the immediate effect of excluding an incongruent noun phrase from the set of candidate antecedents for the pronoun.

The findings in Cowart and Cairns (1987) suggest two important conclusions. First, the cataphoric pronoun *they* initiates an active search for an antecedent in the main subject position. In fact, these results may provide stronger evidence for active search than Van Gompel and Liversedge's (2003) study: if the presence of the pronoun biases the parser to analyze the ambiguous phrase as a noun phrase, this suggests that dependency formation does not wait until after grammatical categories have been identified based on bottom-up analysis of the input. Secondly, their results suggest that the parser does not search for an antecedent in positions that are subject to Principle C. However, some objections can be raised with respect to the design of the study and this interpretation of the results. First, the naming task that was used did not provide information on positions beyond the probe position, which may have concealed important effects at earlier or later words. Second, as the authors themselves point out, the results were only significant in the participants analysis but not in the items analysis, which raises the possibility that the results are dependent on the specific

properties of individual items. Finally, the argument for immediate effects of Principle C is based on the fact that the pronoun-type effect is present in the 'semantic' pair (11) but not in the 'syntactic' pair (12). This argument requires the implicit assumption that different semantic and syntactic violations should produce an effect of a comparable magnitude at the same point in time. In light of this, the conclusions drawn from these findings should be more cautious, perhaps that the syntactic considerations are taken into account by the parser earlier than the semantic considerations. However, the claim that the parser never violates a syntactic constraint needs stronger evidence, e.g., using structures as controls that are not subject to any constraints. In the experiments that we present below, we used methods and materials designed to resolve these difficulties.

Experiment 1

Experiment 1 consisted of an acceptability rating task and an on-line self-paced reading task that used a gender mismatch paradigm to test for the impact of Principle C on the search for pronoun antecedents.

Participants

Participants were 60 native speakers of English from the University of Maryland undergraduate population with normal or corrected-to-normal vision and no history of language disorders. All participants in this and subsequent studies gave informed consent and were paid $10/ h for their participation.

Materials and design

The materials for the on-line task are described first, since the materials for the acceptability rating study were

derived from them. The experiment had five conditions, four of which were organized in a 2×2 design with the factors *constraint* (Principle C vs. no-constraint) and *gender congruency* (match vs. mismatch between the pronoun and the subject of the second clause). A full set of items is shown in Table 1. The gender of the cataphoric pronoun was balanced across stimulus sets: half of the sets were built on the basis of the masculine pronoun *he*, and the other half on the basis of the feminine pronoun *she*. The gender-match and gender-mismatch sentences differed only in the gender of the subject of the second clause, which was always a gender-unambiguous proper name, matched for the number of letters and syllables in the match and mismatch variants within each set of items. To guard against a semantic bias for or against a coreferential interpretation of the pronoun and the second subject, the two clauses were carefully selected such that the events in each clause could plausibly be performed either by the same agent or by different agents.

In addition, to ensure that the cataphoric pronoun received a grammatical antecedent in every case, the target structures were embedded in a further sentence introduced by the conjunctions *although* or *since*. The gender of the third clause subject was chosen such that each sentence had a unique grammatical antecedent for the pronoun. Thus, in the Principle C conditions the subject of the third clause always matched the gender of the pronoun and served as a grammatical antecedent. In the no-constraint conditions the gender of the third clause subject mismatched the pronoun in the gender-match condition, due to the possibility of coreference between the pronoun and the second clause subject, but the third clause subject matched the gender of the pronoun in the gender-mismatch condition.

Following Van Gompel and Liversedge (2003), we added a fifth 'name' condition to each set. This condition was identical to the no-constraint/mismatch condition in the number of referents introduced in each clause. This was done to ensure that any observed

Table 1
Sample set of experimental items for on-line Experiment 1

Principle C/match	Because last semester she$_i$ was taking classes full-time while <u>Kathryn</u> was working two jobs to pay the bills, Erica$_i$ felt guilty
Principle C/mismatch	Because last semester she$_i$ was taking classes full-time while <u>Russell</u> was working two jobs to pay the bills, Erica$_i$ felt guilty
No constraint/match	Because last semester while she$_i$ was taking classes full-time <u>Kathryn$_i$</u> was working two jobs to pay the bills, Russell never got to see her
No constraint/mismatch	Because last semester while she$_i$ was taking classes full-time <u>Russell</u> was working two jobs to pay the bills, Erica$_i$ promised to work part-time in the future
No constraint/name	Because last semester while Erica$_i$ was taking classes full-time <u>Russell</u> was working two jobs to pay the bills, she$_i$ promised to work part-time in the future

The underlined name indicates the critical second subject noun phrase. Subscript indices indicate intended backward anaphoric dependencies.

mismatch effect at the second subject in the no-constraint pair could not merely be due to the introduction of a new discourse referent. If the number of referents is indeed the reason for a mismatch effect in the no-constraint pair, we should expect a similar increase in reading times in the name condition.

Thirty sets of five conditions were distributed among 5 lists in a Latin Square design, and combined with 90 filler sentences. To mask experimental sentences, the fillers bore a number of similarities with the target items, including length and average clause number and were designed in several subgroups, each built around a salient feature of the targets, such as the use of proper names and pronouns, or a subordinator followed by a temporal modifier at the beginning of the sentence. There were no instances of unresolved anaphora in the filler sentences. Thus, we ensured that throughout the experiment pronouns always found intra-sentential antecedents. This was appropriate, since the interest of the study is not in whether readers search for an antecedent but, rather, in where they search.

Acceptability rating task (off-line)

We conducted an off-line rating task using a similar methodology to Gordon and Hendrick (1997). In each sentence a pronoun and a noun phrase were highlighted in bold and participants were instructed 'to determine how plausible it is that the pronoun in bold and the noun in bold refer to the same person' on a scale from 1 (impossible) to 5 (absolutely natural). We sought to confirm that, provided matching gender/number values, participants indeed accept coreference between the cataphoric pronoun and the second subject in the no-constraint environments, but reject it in the Principle C environments. 12 sets of items were chosen from the materials for the online task of Experiment 1 and then simplified to create the two conditions shown in (13). Two different lists were constructed based on these 12 sets using a Latin Square design, and each list had two different stimulus order randomizations. 40 native English speakers from the University of Maryland undergraduate population were recruited specifically for the purposes of the rating study and completed the questionnaire, which contained 6 instances of each condition, interspersed with 20 additional filler sentences, twelve of which tested the configurations used in Experiment 2.

(13) Stimuli from off-line rating task, Experiment 1.
      a. * *Principle C*
         Because last semester she$_i$ was taking classes full-time while Kathryn was working two jobs to pay the bills, Erica$_i$ felt guilty.

      b. *No-constraint*
         Because last semester while she$_i$ was taking classes full-time Kathryn$_i$ was working two jobs to pay the bills, Russell never got to see her.

The mean rating score for the Principle C condition (mean = 1.4, standard error = .12) was significantly lower than the mean score in the no-constraint condition (mean = 4.1, standard error = .13) both in the participants analysis and items analysis (two-tailed paired *t*-test, both $ps < .001$, $t1(39) = 12.2$, $t2(11) = 38.2$). These results fully agree with the predictions of the Principle C constraint: coreference between the pronoun in the main subject position and the name in the second subject position is acceptable only when the pronoun does not c-command the name, i.e., in the no-constraint condition but not in the Principle C condition.

Procedure

Participants were tested using a desktop PC running the *Linger* software (Doug Rohde, MIT) in a standard self-paced word-by-word moving window paradigm (Just, Carpenter, & Woolley, 1982). Each trial started with a blank screen. Upon pressing the space bar, a sentence masked by dashes appeared on the screen. The masks extended to all letters and punctuation marks, but left spaces unmarked. As the participant pressed the spacebar, a new word appeared on the screen as the previous one was re-masked by dashes. A comprehension question appeared after the end of each sentence all at once (e.g. *Was Kathryn/Russell working two jobs?*). Participants were instructed to read sentences at a natural pace and to respond to the comprehension questions as accurately as possible. To answer the question the subject pressed the *f*-key for 'yes' and the *j*-key for 'no.' If the question was answered incorrectly the word 'Incorrect' appeared briefly in the center of the screen. Each participant was randomly assigned to one of the lists, and the order of the stimuli within the presentation list was randomized for each participant.

Analysis

Only trials for which the corresponding comprehension question was answered correctly were included in the analysis. Reading times that exceeded a threshold of 2.5 standard deviations above a participant's mean reading rate for each region were replaced by the threshold value. This winsorizing procedure affected 2.4% of the data (range 1.9–2.6% for individual conditions).

The regions used for the data analysis corresponded to single words, except for regions corresponding to the end of the clause, for which several words were combined due to variation in the clause length between items (see the legend in Figs. 1 and 2 for regions). The critical second subject position in which the gender manipulation occurred corresponded to region 8 in all conditions.

The data from the first four conditions were entered into a 2×2 repeated-measures ANOVA with the factors *constraint* (Principle C, no constraint) and *congruency* (match, mismatch). Reading times from the name condition were compared pairwise to each of the no-constraint conditions in a one-way ANOVA. ANOVAs were computed on the participant mean raw reading times collapsing over items (F1), and on item means

collapsing over participants (F2). Furthermore, *min F'* statistics (Clark, 1973) were computed, and 95% confidence intervals (CIs) based on the mean squared errors of the respective effect from the participants analysis were calculated for comparisons between means of conditions (Masson & Loftus, 2003).

Results

Comprehension question accuracy

Two of 60 participants showed comprehension question accuracy that was more than 2.5 standard deviations below the mean accuracy. Consequently, the lowest scoring participant from each list was removed from the analysis to balance the number of participants
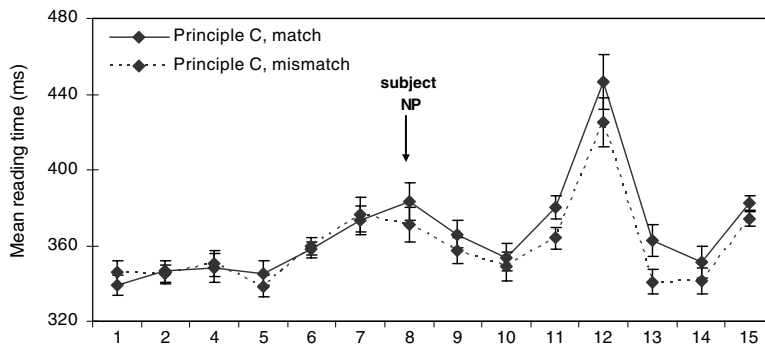
Because$_1$ / last semester $_2$ / she$_4$ / was taking$_5$ /classes full-time$_6$ / while$_7$ / Kathryn (Russell)$_8$ / was working$_9$ / two$_{10}$ / jobs to pay the bills,$_{11}$ / Erica$_{12}$ / felt$_{13}$ /guilty$_{14}$.

Fig. 1. Mean reading times in milliseconds for the Principle C conditions from Experiment 1. The error bars represent one standard error above/below the participant mean at each region. The arrow marks the critical second subject noun phrase.
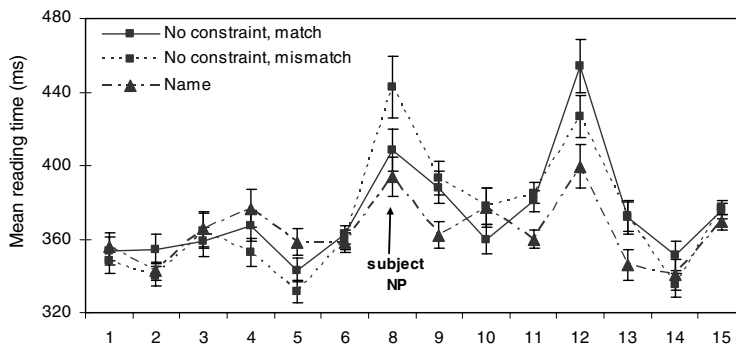
Because$_1$ / last semester$_2$ / while$_3$ / she (Erica)$_4$ /was taking$_5$ / classes full-time$_6$ / Kathryn (Russell)$_8$ / was working$_9$ / two$_{10}$ /jobs to pay the bills,$_{11}$ / Russell (Erica) (she)$_{12}$ / never$_{13}$ / got$_{14}$ / to see her$_{15}$.

Fig. 2. Mean reading times in milliseconds for the no-constraint and the name conditions from Experiment 1. The error bars represent one standard error above/below the participant mean at each region. The arrow marks the critical second subject noun phrase.

on each list. For the remaining 55 participants, the mean accuracy was 91%, with a range of 90–93% for individual conditions. There were no reliable differences in accuracy between individual conditions.

Self-paced reading

We first present the results of the 2×2 ANOVA on the Principle C and no-constraint conditions (Table 2), and then discuss the name condition. The results from the

Table 2
Results of 2×2 ANOVAs and *min F'* values for Experiment 1

	By participants				By items				Min F'		
	df	MSE	F1	p	df	F2	p	df	min F'	p
Region 4 (1st subject)										
Constraint	1,54	2357	2.0	0.160	1,29	1.6	0.216	1,70	<1	0.348
Congruency	1,54	2509	1.2	0.276	1,29	3.1	0.091	1,81	<1	0.354
Constraint × congruency	1,54	1622	2.9	0.094	1,29	1.0	0.319	1,50	<1	0.388
Region 5 (was taking)										
Constraint	1,54	2163	<1	0.622	1,29	<1	0.721	1,59	<1	0.772
Congruency	1,54	1699	2.0	0.162	1,29	1.4	0.244	1,66	<1	0.366
Constraint × congruency	1,54	2017	<1	0.768	1,29	<1	0.558	1,76	<1	0.792
Region 6 (classes full-time)										
Constraint	1,54	1785	<1	0.910	1,29	<1	0.799	1,72	<1	0.917
Congruency	1,54	1227	<1	0.443	1,29	1.0	0.320	1,83	<1	0.541
Constraint × congruency	1,54	2300	<1	0.946	1,29	<1	0.952	1,70	<1	0.964
Region 8 (2nd subject)										
Constraint	1,54	12602	9.2	0.004	1,29	10.9	0.003	1,79	5.0	0.028
Congruency	1,54	8026	<1	0.385	1,29	<1	0.419	1,72	<1	0.551
Constraint × congruency	1,54	6982	4.8	0.033	1,29	2.8	0.104	1,62	1.8	0.188
Region 9 (was working)										
Constraint	1,54	4571	8.6	0.005	1,29	11.4	0.002	1,81	4.9	0.030
Congruency	1,54	3471	<1	0.834	1,29	<1	0.725	1,80	<1	0.857
Constraint × congruency	1,54	2895	1.0	0.313	1,29	<1	0.357	1,71	<1	0.493
Region 10 (two)										
Constraint	1,54	2309	6.2	0.016	1,29	2.3	0.142	1,50	1.7	0.202
Congruency	1,54	2742	<1	0.347	1,29	<1	0.414	1,69	<1	0.535
Constraint × congruency	1,54	2495	2.6	0.112	1,29	2.3	0.141	1,72	1.2	0.273
Region 11 (jobs to pay the bills)										
Constraint	1,54	2883	1.4	0.236	1,29	<1	0.579	1,42	<1	0.614
Congruency	1,54	2234	1.1	0.295	1,29	1.8	0.196	1,82	<1	0.411
Constraint × congruency	1,54	2183	1.4	0.249	1,29	<1	0.545	1,45	<1	0.590
Region 12 (3rd subject)										
Constraint	1,54	6854	<1	0.820	1,29	<1	0.754	1,83	<1	0.853
Congruency	1,54	9304	4.2	0.046	1,29	3.1	0.087	1,68	1.8	0.185
Constraint × congruency	1,54	8821	<1	0.970	1,29	<1	0.762	1,56	<1	0.971
Region 13 (never)										
Constraint	1,54	3266	5.9	0.019	1,29	4.3	0.048	1,67	2.5	0.121
Congruency	1,54	3381	2.2	0.140	1,29	1.6	0.216	1,67	<1	0.337
Constraint × congruency	1,54	3623	2.7	0.109	1,29	3.4	0.075	1,80	1.5	0.226
Region 14 (got)										
Constraint	1,54	1706	<1	0.517	1,29	<1	0.588	1,67	<1	0.676
Congruency	1,54	2514	3.6	0.062	1,29	3.2	0.083	1,73	1.7	0.195
Constraint × congruency	1,54	1958	<1	0.698	1,29	<1	0.817	1,50	<1	0.842
Region 15 (to see her)										
Constraint	1,54	1339	<1	0.920	1,29	<1	0.707	1,61	0.0	0.922
Congruency	1,54	1089	1.9	0.178	1,29	1.0	0.329	1,59	0.6	0.425
Constraint × congruency	1,54	1513	<1	0.983	1,29	<1	0.897	1,57	0.0	0.983

Principle C conditions are presented in Fig. 1 and the results from the no-constraint and name conditions in Fig. 2.

Principle C and no-constraint conditions. There were no reliable main effects or interactions in any region in the first clause.

In the main ANOVA at the critical second subject in region 8 there was a main effect of constraint due to shorter mean reading times in the Principle C conditions than in the no-constraint conditions (mean for the no-constraint conditions = 425.5 ms, mean for the Principle C conditions = 377.3 ms, 95% CI = 30.3 ms). This effect can be attributed to the differences in the content of the preceding region between the two pairs of conditions. Most importantly, there also was a significant constraint × congruency interaction at the second subject in the participants analysis. Pairwise comparisons within each level of the constraint factor showed no effect of gender congruency for the Principle C conditions (gender-match—383.4 ms, gender-mismatch—371.1 ms, 95% CI = 23.8 ms) and a marginally significant effect of gender congruency for the no-constraint conditions (gender-match—408.4 ms, gender-mismatch—442.6 ms, 95% CI = 40.3 ms) due to a slowdown in average reading times in the gender-mismatch condition (i.e., a mismatch effect). The main effect of constraint persisted to the region immediately following the second subject (mean for the no-constraint conditions = 390.7 ms, mean for the Principle C conditions = 361.7 ms, 95% CI = 18.3 ms). However, neither the main effect of congruency nor the constraint × congruency interaction were significant and no significant differences were observed in the pairwise comparisons within each level of the factor constraint (Principle C conditions: gender-match—366.1 ms, gender-mismatch—357.3 ms, 95% CI = 16.1 ms; no-constraint conditions: gender-match—388.0 ms, gender-mismatch—393.5 ms, 95% CI = 26.1 ms). The main effect of constraint was still significant in the second post-subject region in the participants analysis, although not in the items analysis (mean for the no-constraint conditions = 351.4 ms, mean for the Principle C conditions = 369.0 ms, 95% CI = 13.0 ms).

Other reliable effects included a main effect of congruency at the third subject noun phrase in region 12 that was significant in the participants analysis and marginally significant in the items analysis, due to longer reading times for the third subject in the match conditions than in the mismatch conditions (mean for the gender-match conditions = 450.2 ms, mean for the gender-mismatch conditions = 426.0 ms, 95% CI = 26.1 ms). Recall that the congruency factor manipulated the (mis)match in gender between the pronoun and the 2nd subject, rather than the 3rd subject. There was also a main effect of the constraint factor at the following word (mean for

the no-constraint conditions = 351.8 ms, mean for the Principle C conditions = 371.9 ms, 95% CI = 15.4 ms) due to slower reading times in the Principle C conditions. Pairwise comparisons in the same region showed a significant effect of congruency for the Principle C pair (gender-match—362.7 ms, gender-mismatch—340.9 ms, 95% CI = 18.6 ms) due to longer reading times in the match condition, but no corresponding effect of congruency in the no-constraint pair (gender-match—371 5 ms, gender-mismatch—372.3 ms, 95% CI = 26.4 ms). No other differences were statistically reliable. The effects in regions 12 and 13 do not receive a straightforward explanation, e.g. they cannot be explained by the parser's anticipation of an antecedent noun phrase in conditions where the anaphoric dependency has not yet been resolved (i.e. both Principle C conditions and the gender-mismatch condition of the no-constraint pair).

Name condition. In the first clause there was a significant difference in reading times between the name and the mismatch condition of the no-constraint pair at the subject noun phrase in region 4 (mean for the name condition = 377.4 ms, mean for the no-constraint mismatch condition = 351.0 ms, 95% CI = 19.5 ms) and at the following region (mean for the name condition = 356.7 ms, mean for the no-constraint mismatch condition = 331.9 ms, 95% CI = 16.7 ms), due to longer reading times in the name condition. These effects are expected in light of the length differences in the first subject noun phrase, which was a personal pronoun in the no-constraint conditions and a proper name in the name condition.

Reading times at the critical second subject noun phrase were significantly longer in the mismatch condition of the no-constraint pair than in the name condition (mean for the name condition = 393.9 ms, mean for the no-constraint mismatch condition = 442.6 ms, 95% CI = 41.0 ms). However, reading times at the same region did not differ between the name and the match condition (mean for the name condition = 393.9 ms, mean for the no-constraint match condition = 408.4 ms, 95% CI = 22.6 ms). The fact that reading times at the second subject were longer in the mismatch condition than in the other two conditions is unlikely to be a spill-over from an earlier region, in light of the almost identical reading times across all three conditions at the immediately preceding region (i.e. region 6). There were additional reading-time differences between the name condition and either of the no-constraint conditions at later regions. These differences are expected in light of the parser's attempt to resolve coreference at the second subject noun phrase in the no-constraint conditions but not in the name condition and are not reported here as they are not germane to the goals of this study.

Discussion

The off-line ratings showed that participants judged the second subject noun phrase to be an acceptable antecedent for the preceding pronoun in the no-constraint condition, but not in the Principle C condition, confirming that the contrast between the conditions is real, as predicted by the Principle C constraint. Notwithstanding the fact that the effect remained marginally significant, our results from the self-paced reading task replicated Van Gompel and Liversedge's (2003) gender-mismatch effect in the no-constraint conditions: there was a slowdown in the second subject position if that noun phrase mismatched in gender with the preceding pronoun. In contrast, there was no effect of gender congruency at the corresponding position in the Principle C conditions, or in any other region in the second clause. The absence of a mismatch effect in the Principle C condition suggests that in this condition the parser never considered the subject of the second clause as a potential antecedent, thus making the gender-congruency between that subject and the preceding pronoun irrelevant. We can also rule out the possibility that the gender-mismatch effect in the no-constraint condition was due to the cost of adding a new discourse referent, since reading times at the second subject in the name condition patterned with the gender-match condition rather than with the gender-mismatch condition. The name condition was no harder to process than the gender match condition in any of the regions, so the difference between the gender match and mismatch conditions cannot be due to the introduction of a new discourse referent in the mismatch condition.

The presence of the gender mismatch effect in the no-constraint conditions indicates that an anaphoric dependency is formed at the second subject noun phrase before gender information is taken into account, but it does not show exactly when the dependency is formed. This could occur immediately following the identification of the category of the subject based on bottom-up information, as in Van Gompel & Liversedge's account. Alternatively, dependency formation could occur before any bottom-up information about the second subject noun phrase appears, due to the fact that a sentence-initial adverbial clause reliably predicts a higher clause that contains a subject noun phrase. Both of these accounts may be viewed as evidence for 'active' search for the antecedent for a pronoun, but uncertainty about the role of prediction leads to a problem in interpreting the Experiment 1 results. The argument for the early on-line effects of Principle C involves the absence of a mismatch effect, and therefore relies on the assumption that there is no other difference between the conditions that could have led to the absence of a mismatch effect in the Principle C conditions. However, there is an inherent asymmetry between the Principle C and the no-constraint conditions related

to the parser's ability to predict the second subject position. Recall that the sentences used in the experiment were 3-clause structures, schematized in (14).

(14) Schematic representation of conditions from Experiment 1.
No-constraint:
[Because [while *pronoun* [2nd subject……]], 3rd subject…]
Principle C:
[Because [*pronoun*….[while 2nd subject…..]], 3rd subject…]

In the no-constraint conditions the parser could reliably predict the second subject position immediately after encountering the subordinator *while* in the first clause. In contrast, the second subject position cannot be reliably predicted during the first clause in the Principle C conditions. Furthermore, in these conditions the third subject position, which corresponds to the main clause subject, can be reliably predicted during the first clause. Therefore, the reason for the absence of a gender mismatch effect in the Principle C conditions could be due to the unpredictability of the second subject position, rather than to the effects of the grammatical constraint. Furthermore, the lack of expectation for the second subject position could strongly reduce any effect of the gender of the noun phrase in that position.

To separate the effects of grammatical constraints and predictability, Experiments 2 and 3 tested the effects of the Principle C constraint in configurations where the critical noun phrase position is equally predictable across all conditions.

Experiment 2

The goal of Experiment 2 was to test the effect of Principle C on active search for antecedents, while avoiding confounds due to differences in the predictability of the critical subject noun phrase. In Experiment 2 both the Principle C and the no-constraint conditions were identical with respect to the predictability of the second clause at all times. Thus, a contrast between the results for the no-constraint and Principle C conditions in this study would provide stronger evidence for the impact of grammatical constraints on the construction of anaphoric dependencies.

Participants

Participants were 60 native speakers of English from the University of Maryland undergraduate population with normal or corrected-to-normal vision and no history of language disorders.

Materials and design

The materials for the on-line task are described first, since the materials for the rating study were derived from them. The experiment had four conditions in a 2×2 design with the within-subjects factors *constraint* (Principle C vs. no-constraint) and *gender congruency* (match vs. mismatch between the pronoun and the subject of the second clause). A full set of items is shown in Table 3.

All target sentences started with a clause with an impersonal subject and a non-agentive predicate such as *it seemed* or *it was surprising*. We reasoned that a sentence fragment like *It seemed worrisome to . . .* would create a strong expectation for a complement clause and hence for an additional subject position. Consequently, in all conditions the second subject position could be anticipated when the cataphoric pronoun was first encountered, thereby eliminating the potential confound that made it more difficult to interpret the distribution of gender mismatch effects in Experiment 1. In the Principle C conditions the cataphoric pronoun was the complement of the dative preposition *to* or *for* and referred to the experiencer of the main clause predicate. In this configuration coreference between the pronoun and the subject of an embedded clause is unacceptable. In the no-constraint conditions the pronoun was embedded as the possessor inside the complement of *to* or *for*. In this configuration the pronoun may corefer with the subject of an embedded clause. Note that this contrast in acceptability follows from the syntactic formulation of Principle C only if prepositions are assumed to be irrelevant for calculation of c-command relations. This assumption is consistent with results from other diagnostics of c-command, such as negative polarity item licensing and scope assignment, and the reader is referred to Reinhart (1983) and Brody (1994) for further discussion. As in Experiment 1, the materials were designed such that a cataphoric pronoun could always find a grammatically accessible intra-sentential antecedent, even in the Principle C conditions. This was

achieved by adding a *but*-clause to the end of the sentence, the subject of which was a licit antecedent for the pronoun. The gender of the third subject was chosen such that there was a unique antecedent for the pronoun in each condition. The stimuli were designed such that plausibility considerations should not favor coreference between the pronoun and the embedded subject more strongly in either the Principle C condition or the no-constraint condition (irrespective of grammatical acceptability). For example, in the set of materials in Table 3 both John and his family could plausibly be worried about something negative involving John such as gaining weight.

Thirty two sets of target items were distributed among four presentation lists in a Latin Square design. Each list contained 32 experimental sentences (8 per condition) and 90 filler sentences of varying length and complexity, thereby maintaining target-to-filler ratio of approximately 1:3. Filler sentences were identical across all four lists, and were designed with the same considerations used in Experiment 1. Each participant was randomly assigned to one of the lists, and the order of the stimuli within the presentation list was randomized for each participant. The full list of materials is given in Appendix B.

Acceptability rating task (off-line)

An off-line rating task was used to confirm that, provided matching gender/number values, participants accept coreference between the cataphoric pronoun and the second subject in the no-constraint environment, but reject it in the Principle C environments. The acceptability rating task for Experiment 2 was run together with the corresponding task for Experiment 1, and so was completed by the same 40 participants described there, using the same design and procedure. Participants were asked to rate the acceptability of coreference on a scale from 1 (impossible) to 5 (absolutely natural). The test items were 12 pairs of gender-match conditions chosen from the materials for the online task, distributed

Table 3
Sample set of experimental items for on-line Experiment 2

Principle C/match	It seemed worrisome to him$_i$ that <u>John</u> was gaining so much weight, but Matt$_i$ didn't have the nerve to comment on it
Principle C/mismatch	It seemed worrisome to him$_i$ that <u>Ruth</u> was gaining so much weight, but Matt$_i$ didn't have the nerve to comment on it
No constraint/match	It seemed worrisome to his$_i$ family that <u>John</u>$_i$ was gaining so much weight, but Ruth thought it was just a result of aging
No constraint/mismatch	It seemed worrisome to his$_i$ family that <u>Ruth</u> was gaining so much weight, but Matt$_i$ thought it was just a result of aging

The underlined name indicates the critical second subject noun phrase. Subscript indices indicate intended backward anaphoric dependencies.

across two presentation lists using a Latin Square design and interspersed with 48 filler items (additional to the materials from Experiment 1).

The mean rating score for the Principle C condition (mean = 1.5, standard error = .12) was significantly lower than in the no-constraint condition (mean = 4.2, standard error = .13) both in the participants and items analyses (two-tailed paired *t*-test, both *p*s < .001, $t1(39) = 12.9$, $t2(11) = 31.2$). These results strongly suggest that participants recognize the constraint on coreference for the Principle C structures in (14).

Procedure and analysis

The procedure was a self-paced word-by-word moving window task, identical to Experiment 1. Each sentence was followed by a yes-no comprehension question.

Data from all 60 participants was included in the analysis, which followed the same steps as in Experiment 1. Sentences for which the comprehension question was answered incorrectly were excluded from the analysis, and reading times that exceeded a threshold of 2.5 standard deviations above a participant's mean reading rate for each region were replaced by the threshold value. This winsorizing procedure affected 2.0% of trials (1.9–2.2% for individual conditions).

The regions used for the data analysis corresponded to single words, except for regions corresponding to the end of the clause, for which several words were averaged together due to variation in the clause length between items (see the legend for Fig. 3 for region encoding). Raw reading times for each region were entered into a 2×2 repeated-measures ANOVA with the factors *constraint* and *gender congruency*.

Results

Comprehension question accuracy

The mean question answering accuracy was 95.7% and ranged between 94.4% and 97.3% for individual conditions. There were no reliable differences in accuracy between individual conditions (*F*s < 1).

Self-paced reading

The results from Experiment 2 are plotted in Fig. 3. The results of the 2×2 ANOVA analysis are reported in Table 4.

At the impersonal subject *it* in region 1 the general ANOVA showed an interaction of constraint and congruency that was significant in the participants analysis and marginally significant in the items analysis. However, this interaction was not supported in the pairwise comparisons within each level of the *constraint* factor. There were no other significant effects or interactions anywhere in the first clause.

At the complementizer *that* in region 7 there was a main effect of the constraint factor, due to longer reading times for the no-constraint conditions than for the Principle C conditions (mean for the no-constraint conditions = 321.7 ms, mean for the Principle C conditions = 302.4 ms, 95% CI = 7.7 ms). This difference was likely due to the lexical difference between conditions at the preceding regions. A main effect of the constraint factor in the same direction was also found at the subsequent region, the critical subject of the second clause (region 8: mean for the no-constraint conditions = 328.5 ms, mean for the Principle C conditions = 317.7 ms, 95% CI = 10.3 ms). However, in region 8 the constraint × congruency interaction did not reach significance, and planned pairwise comparisons revealed no
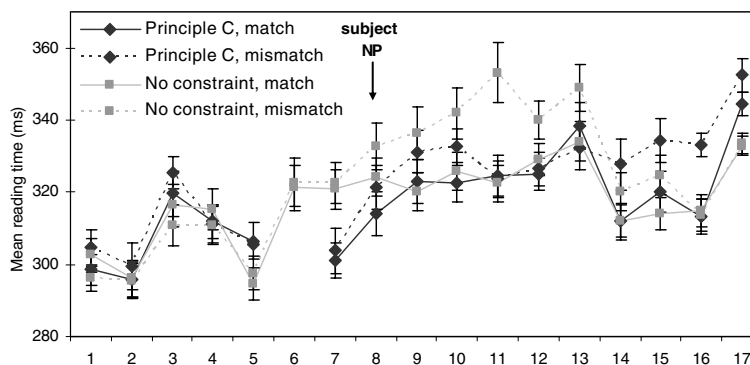
It$_1$ seemed$_2$ worrisome$_3$ to$_4$ {him$_5$/ his$_5$ family$_6$} that$_7$ John/Ruth$_8$ was$_9$ gaining$_{10}$ so$_{11}$ (much weight)$_{12}$, but$_{13}$ Matt$_{14}$ didn't$_{15}$ have$_{16}$ (the nerve to comment on it)$_{17}$.

Fig. 3. Mean reading times in milliseconds from Experiment 2. The error bars represent one standard error above/below the participant mean at each region. The arrow marks the critical second subject noun phrase.

Table 4
Results of 2×2 ANOVAs and *Min F'* values for Experiment 2

	By participants				By items			*Min F'*		
	df	MSE	F1	p	df	F2	p	df	min F'	p
Region 4 (to)										
Constraint	1,59	1279.1	<1	0.626	1,31	<1	0.754	1,31	<1	0.792
Congruency	1,59	943.0	<1	0.564	1,31	<1	0.690	1,31	<1	0.743
constraint × congruency	1,59	880.6	<1	0.571	1,31	<1	0.750	1,31	<1	0.781
Region 5 (him/his)										
Constraint	1,59	955.4	5.4	0.024	1,31	3.2	0.082	1,31	2.0	0.165
Congruency	1,59	944.8	<1	0.821	1,31	<1	0.754	1,31	<1	0.855
Constraint × congruency	1,59	820.4	<1	0.743	1,31	<1	0.701	1,31	<1	0.803
Region 7 (that)										
Constraint	1,59	1448.3	16.3	<0.001	1,31	10.6	0.003	1,31	6.4	0.016
Congruency	1,59	997.3	<1	0.486	1,31	<1	0.654	1,31	<1	0.707
Constraint × congruency	1,59	1401.9	<1	0.912	1,31	<1	0.966	1,31	<1	0.968
Region 8 (2nd subject)										
Constraint	1,59	1589.5	4.7	0.034	1,31	4.3	0.048	1,31	2.2	0.145
Congruency	1,59	1187.6	3.2	0.077	1,31	2.7	0.112	1,31	1.5	0.235
Constraint × congruency	1,59	1624.0	<1	0.927	1,31	<1	0.895	1,31	<1	0.940
Region 9 (was)										
Constraint	1,59	1925.5	<1	0.880	1,31	<1	0.708	1,31	<1	0.889
Congruency	1,59	2223.8	3.7	0.059	1,31	3.9	0.058	1,31	1.9	0.179
Constraint × congruency	1,59	1681.0	<1	0.544	1,31	<1	0.406	1,31	<1	0.625
Region 10 (gaining)										
Constraint	1,59	1422.0	1.6	0.209	1,31	1.3	0.260	1,31	<1	0.401
Congruency	1,59	1055.7	8.8	0.004	1,31	5.0	0.033	1,31	3.2	0.085
Constraint × congruency	1,59	1587.0	<1	0.601	1,31	<1	0.438	1,31	<1	0.665
Region 11 (so)										
Constraint	1,59	1339.5	8.1	0.006	1,31	5.2	0.030	1,31	3.2	0.085
Congruency	1,59	1909.9	7.4	0.009	1,31	5.7	0.023	1,31	3.2	0.083
constraint × congruency	1,59	2056.6	6.7	0.012	1,31	9.0	0.005	1,31	3.8	0.059
Region 12 (much weight)										
Constraint	1,59	3863.8	4.0	0.049	1,31	3.1	0.090	1,31	1.7	0.196
Congruency	1,59	5224.2	1.9	0.175	1,31	5.0	0.033	1,31	1.4	0.250
Constraint × congruency	1,59	2737.8	1.1	0.297	1,31	1.9	0.183	1,31	<1	0.411
Region 13 (but)										
Constraint	1,59	1464.2	1.6	0.210	1,31	1.0	0.326	1,31	<1	0.439
Congruency	1,59	2053.6	<1	0.450	1,31	1.3	0.256	1,31	<1	0.530
Constraint × congruency	1,59	2752.1	2.4	0.126	1,31	5.9	0.021	1,31	1.7	0.200
Region 14 (3rd subject)										
Constraint	1,59	1106.5	<1	0.480	1,31	<1	0.578	1,31	<1	0.662
Congruency	1,59	1562.1	5.8	0.019	1,31	8.5	0.007	1,31	3.5	0.073
Constraint × congruency	1,59	1231.9	<1	0.383	1,31	<1	0.367	1,31	<1	0.530
Region 15 (didn't)										
Constraint	1,59	1399.0	2.5	0.122	1,31	1.4	0.240	1,31	<1	0.349
Congruency	1,59	1101.5	8.8	0.004	1,31	3.0	0.091	1,31	2.3	0.143
Constraint × congruency	1,59	1840.9	<1	0.789	1,31	<1	0.764	1,31	<1	0.842
Region 16 (have)										
Constraint	1,59	1134.4	3.8	0.058	1,31	1.0	0.332	1,31	<1	0.387
Congruency	1,59	1309.7	4.0	0.050	1,31	4.5	0.042	1,31	2.1	0.156
Constraint × congruency	1,59	1176.0	5.5	0.022	1,31	5.6	0.024	1,31	2.8	0.105

effect of congruency within either level of the constraint factor.

The ANOVAs at each of the three regions immediately following the subject (regions 9–11 in Fig. 1) showed a significant or a marginally significant main effect of congruency, due to slower reading times in mismatching conditions than matching conditions (region 9: mean for the gender-match conditions = 321.6 ms, mean for the gender-mismatch conditions = 333.8 ms, 95% CI = 12.2 ms; region 10: mean for the gender-match conditions = 324.3 ms, mean for the gender-mismatch conditions = 337.2 ms, 95% CI = 8.4 ms; region 11: mean for the gender-match conditions = 323.4 ms, mean for the gender-mismatch conditions = 338.5 ms, 95% CI = 11.3 ms). In region 11 there also was a significant main effect of constraint (mean for the no-constraint conditions = 337.8 ms, mean for the Principle C conditions = 324.1 ms, 95% CI = 11.7 ms) and a significant constraint × congruency interaction due to longer reading times in the no-constraint mismatch condition (no-constraint conditions: gender-match—322.4 ms, gender-mismatch—353.1 ms; Principle C conditions: gender-match—324.4 ms, gender-mismatch—323.9 ms, 95% CI = 23.4 ms). Planned pairwise comparisons at regions 9–11 showed a significant or marginally significant effect of gender congruency (i.e., gender-mismatch effect) in the no-constraint conditions at all three regions (region 9: mean for the gender-match condition = 320.2 ms, mean for the gender-mismatch condition = 336.4 ms, 95% CI = 16.8 ms; region 10: 325.8, 341.9 and 14.9 ms; region 11: 322.4, 353.1 and 19.2 ms, respectively). The effect of gender congruency did not reach significance in the Principle C conditions in any region in the second clause, although there was a numerical trend towards slower reading times in the incongruent conditions.

In the third clause there was a significant main effect of congruency in the overall ANOVA at the subject noun phrase in region 14, due to longer reading times in the mismatch conditions than in the match conditions (mean for the gender-match conditions = 311.9 ms, mean for the gender-mismatch conditions = 323.9 ms, 95% CI = 10.2 ms). Pairwise comparisons showed that the effect of gender congruency was significant in the Principle C conditions (region 14: mean for the gender-match condition = 311.9 ms, mean for the gender-mismatch condition = 327.7 ms, 95% CI = 12.7 ms), but not in the no-constraint conditions (region 14: mean for the gender-match condition = 312.0 ms, mean for the gender-mismatch condition = 320.2 ms, 95% CI = 14.6 ms). This was the first region where an accessible antecedent for the pronoun could be found in all but the no-constraint/match condition. This pattern of reading-time contrasts was unexpected, especially in light of the persistency of the effect that continued to hold in regions 15 and 16.

Discussion

The goal of Experiment 2 was to test the claim that Principle C immediately restricts dependency formation during online processing of backwards anaphora, using materials that match the predictability of potential antecedent positions across conditions. As in Experiment 1, this experiment elicited a gender mismatch effect in the no-constraint conditions, and no corresponding effect in the Principle C conditions. These findings support the claim that the parser does not search for an antecedent for a pronoun in grammatically inaccessible positions, and do not straightforwardly lend themselves to an explanation in terms of predictability, unlike Experiment 1.

However, it is clear that the mismatch effect in the no-constraint conditions in Experiment 2 was both delayed and diminished in comparison to Experiment 1. The gender mismatch effect showed a significant or marginally significant mismatch effect at the each of the three words following the critical region, but the overall constraint × congruency interaction did not reach significance until the third region after the critical noun phrase.

One possible reason for the smaller and delayed gender-mismatch effect in this experiment relative to Experiment 1 involves the structural necessity of the clause that contains the potential antecedent noun phrase. Experiment 1 introduced pronouns in adverbial clauses introduced by subordinators like *while*, which are obligatorily followed by a main clause. The target items in Experiment 2 introduced pronouns inside a main clause with a beginning like *It seemed clear...*, which were assumed to create a strong bias for a clausal continuation, based on intuition and on informal corpus searches (*Linguist's Search Engine*: Resnik & Elkiss, 2003). However, the lead-ins used in Experiment 2 are compatible with a grammatical (although pragmatically unlikely) parse as mono-clausal sentences with referential subject pronouns. To test whether the weaker gender mismatch effect in this experiment might have been due to a weaker expectation for a following clause, we conducted an offline sentence completion task to determine the strength of the prediction for a complement clause.

Thirteen undergraduate students from the University of Maryland completed a pencil-and-paper completion task for monetary compensation. The questionnaire contained sentence-initial fragments such as (15), which participants were asked to complete in a natural fashion. The last word of the fragment was a possessive pronoun, e.g. *his*. The pronoun was included to assess what proportion of the completions that contained an embedded clause also contained a referent for the pronoun as part of the embedded clause. All target fragments contained the masculine 3rd person singular pronoun *his* or plural *their*. The feminine pronoun *her* was not used in the

target sentences due to the homophony of the accusative and possessive feminine forms. The questionnaire consisted of 12 target fragments, interspersed with 60 fragments of other types.

  (15) It seemed clear to his....

The results from the completion task suggest that a sentence fragment such as (15) created a strong but not absolute expectation for an upcoming embedded clause. Participants provided a clausal continuation in 69% (99/143) of cases (71% (51/72) in the 'his' condition, 67% (48/71) in the 'their' condition), but in the remaining 29% (44/143) of cases the completed sentence was a simple monoclausal sentence. Such monoclausal completions are grammatical, since the sentence-initial *it* can be understood as a 3rd person neuter pronoun with an unspecified referent, but they are pragmatically infelicitous in the absence of a supporting context. The complement clause contained a potential antecedent for the cataphoric pronoun *his* or *their* in 70% (69/99) of the completions that included a second clause, suggesting that when the embedded clause was expected, a referent for the pronoun was also often expected. In 87% (60/69) of the completions that provided a potential antecedent, the potential referent appeared in subject position.

The results from the fragment completion questionnaire lend support to the notion that speakers normally expect a clausal continuation following a lead-in like *It seemed clear to his ...*, and also suggest that in most cases speakers also anticipate an antecedent for the pronoun. However, the 30% of completions that did not contain an embedded clause still represent a significant difference from the materials in Experiment 1, where a completion to a fragment like *While he ...* would necessarily include a second clause. Thus, it is possible that the difference in strength of prediction represented by this 30% contributed to the weakened gender mismatch effect in Experiment 2.

Therefore, Experiment 2 provided further evidence that active search for antecedents for pronouns is restricted by grammatical constraints on coreference, while leaving open certain questions about the strength and immediacy of gender-mismatch effects and what they show about the time-course of anaphoric dependency formation. In particular, the results of Experiment 2 show that the distribution of gender-mismatch effects cannot be explained entirely in terms of predictability, but do not indicate whether the predictability of a potential antecedent position is a prerequisite for eliciting a gender-mismatch effect. The results of Experiments 1 and 2 are consistent with a view in which active dependency formation mechanisms for cataphora are activated only at the time when the pronoun is encountered. Under this view, which we refer to as *instantaneous resolution*, the parser actively constructs anaphoric depen-

dencies only in positions that can be reliably predicted at the time when the pronoun is encountered. Instantaneous resolution predicts that a dependency between the anaphoric pronoun (*his*) and the second subject position (*John*) would be actively constructed in a structure like 'While his mother slept, John sang', in which the second subject position can be anticipated even before the pronoun is encountered, due to the early appearance of the subordinator *while*. On the other hand, instantaneous resolution predicts that the parser would not attempt a similar dependency in 'His mother slept, while John sang', as the second subject position cannot be reliably predicted until after the pronoun is encountered. An alternative to the instantaneous resolution view is that the parser's search for candidate antecedent positions remains active beyond the point when the pronoun is encountered and takes account of any new piece of structure predicted by bottom-up information following the pronoun. This view, which we refer to as *continuous resolution*, predicts that the parser would actively consider a dependency between a cataphoric pronoun and the second subject even in 'His mother slept, while John sang', despite the fact that a reliable prediction for the second subject position becomes available only with the subordinator *while* and after the anaphoric pronoun is encountered. This latter view is a closer parallel to the properties of active search for gaps in processing filler-gap dependencies, where active gap creation effects have been found in positions that could not have been anticipated at the point when the filler was first encountered (e.g., Aoshima et al., 2004; Frazier & Clifton, 1989).

Importantly, whereas the two views disagree on the specific configuration that must obtain to initiate an active search, they both agree that a cataphoric dependency can be formed based on a prediction for an antecedent position and before the antecedent becomes available bottom-up. We assume that speakers can encode the fact that a pronoun is linked to another syntactic position that serves as its antecedent, independent of encoding what the antecedent refers to. We refer to this process as 'linking', and suggest that it may occur as soon as an antecedent position is predicted. We assume that a separate process of 'valuation' occurs when the properties of the antecedent are processed bottom-up, and the pronoun automatically inherits the referential properties of the antecedent.

Experiment 3 attempts to adjudicate between the instantaneous resolution and the continuous resolution proposals by testing a configuration in which no syntactic prediction for a potential antecedent position is available at the time when the cataphoric pronoun is encountered. If a gender mismatch effect is found in such a configuration, it would be evidence that the search mechanism does not search for candidate antecedents just once, but repeatedly, as more information comes online. Therefore, the results of this third experiment

will also speak further to the extent of the parallelism between active search mechanisms in the two construction types.

Experiment 3

The goals of Experiment 3 were twofold. First, to test the instantaneous vs. continuous resolution accounts, i.e., whether active search for pronoun antecedents occurs in structural positions that could not be reliably predicted at the point when the pronoun is first encountered. Second, as in Experiments 1 and 2 we sought to test whether active search for pronoun antecedents is restricted by grammatical constraints, specifically Principle C.

Participants

Participants were 60 native speakers of English from the University of Maryland undergraduate population with normal or corrected-to-normal vision and no history of language disorders. They gave informed consent and were financially compensated for their participation.

Materials and design

The target items for Experiment 3 consisted of 24 sets of 4 items organized in a 2×2 design with the factors *constraint* (no-constraint vs. Principle C) and *congruency* (match vs. mismatch) as within-subjects factors. One full set of items is shown in Table 5 and a complete set of materials can be found in Appendix C.

All conditions began with a main clause that was followed by an adverbial clause headed by a subordinator such as *while*. The pronoun appeared as the subject of the main clause in the Principle C conditions, or as the possessor of the main clause subject in the no-constraint conditions. The critical potential antecedent position was the subject noun phrase of the embedded adverbial clause, which could not be reliably predicted at the point when the pronoun was first encountered. The first possible cue for such a position appeared later in the sentence when the subordinator *while* was encountered. Importantly, the subordinator *while* did not immediately precede the critical head noun of the second subject noun phrase, but instead was separated from the head noun by the determiner *the* and two adjectives. This ensured that the parser had ample opportunity to process the information associated with the subordinator before encountering the critical noun, i.e., (i) that there is a new clause and, hence, an upcoming subject position, and (ii) that in the Principle C condition this subject position is c-commanded by the cataphoric pronoun and thus excluded as a potential antecedent due to Principle C.

Gender congruent and incongruent sentences were formed by manipulating the gender of the sentence-initial pronoun such that it either matched or mismatched the head noun of the second subject noun phrase. The nouns used were either lexically (e.g. *king*, *queen*) or conventionally (e.g. *quarterback*) strongly gender-specific. To make it impossible to reliably anticipate the gender of the noun, the adjectives preceding the noun did not include gender-biased adjectives (e.g., *pregnant* or *handsome*).

As in the previous experiments, to ensure the possibility of intra-sentential coreference, the sentences included an additional clause that contained a licit antecedent for the cataphoric pronoun whenever no suitable antecedent was available in the second clause. The twenty-four sets of experimental stimuli were distributed across four lists in a Latin Square design. Each list contained 24 experimental sentences and 72 filler sentences. Filler sentences were identical across all four lists. Each participant was randomly assigned to one of the lists, and the order of the stimuli within the presentation list was randomized for each participant.

Table 5
Sample set of experimental items for on-line Experiment 3

Principle C/match	He$_i$ chatted amiably with some fans while the talented, young quarterback signed autographs for the kids, but Steve$_i$ wished the children's charity event would end soon so he could go home
Principle C/mismatch	She$_i$ chatted amiably with some fans while the talented, young quarterback signed autographs for the kids, but Carol$_i$ wished the children's charity event would end soon so she could go home
No constraint/match	His$_i$ managers chatted amiably with some fans while the talented, young quarterback$_i$ signed autographs for the kids, but Carol wished the children's charity event would end soon so she could go home
No constraint/mismatch	Her$_i$ managers chatted amiably with some fans while the talented, young quarterback signed autographs for the kids, but Carol$_i$ wished the children's charity event would end soon so she could go home

The underlined noun indicates the head of the critical second subject noun phrase. Subscript indices indicate intended backward anaphoric dependencies.

Acceptability rating task (off-line)

As in Experiments 1 and 2, an off-line acceptability rating task was used to confirm that the configurations in the no-constraint and Principle C conditions effectively varied the acceptability of coreference between the pronoun and the subject of the second clause. Participants were the same 60 native speakers of English who participated in the on-line portion of Experiment 3. To avoid bias from the acceptability rating task on the on-line results, participants completed the rating questionnaire after the on-line experiment. The task was identical to that used in Experiments 1 and 2, i.e., participants rated the acceptability of coreference on a scale from 1 (impossible) to 5 (absolutely natural). Twenty-four sets of stimuli were constructed based on 24 target stimulus sets from on-line Experiment 3. The *Principle C* and the *no-constraint* conditions were derived from the gender-match conditions of the on-line study by removing the sentence-final *but*-clause. The *forwards anaphora* condition was created by exchanging the pronoun and the second subject noun phrase in the Principle C condition. This condition was included to address the possible concern that lower acceptability ratings in the Principle C condition might reflect the implausibility of having the main and embedded clause events be simultaneously performed by the same agent. The *but*-condition was identical to the Principle C/match condition in the online study and was included to test the effectiveness of our effort to provide a licit intra-sentential antecedent for the pronoun through the subject of an additional *but*-clause.

Four experimental lists were constructed using a Latin Square design, such that each participant saw only one condition from each set. In addition to the 24 experimental sentences, each questionnaire also contained 36 filler sentences that were identical across the four lists. The mean coreference rating scores from Experiment 3 are summarized in Table 6.

The Principle C condition received a mean rating score of 1.7 that was significantly lower than the rating score in the other three conditions (2-tailed paired t-test, all ps < .01). The coreference rating score in the no-constraint and the *but*-conditions was significantly lower than in the forwards anaphora condition, but this is expected given that forwards anaphora is the preferred way of expressing coreference in these contexts.

Table 6
Mean rating scores from Experiment 3

Condition	Mean rating (*Standard error*)
Principle C	1.7 (.09)
No-constraint	3.4 (.13)
Forwards anaphora	4.3 (.08)
But-condition	3.9 (.09)

Procedure and analysis

The procedure was a self-paced word-by-word moving window task, identical to Experiments 1 and 2. Each sentence was followed by a yes-no comprehension question.

The data from 2 of the 60 participants could not be included due to technical problems. To balance the number of participants across each list, one participant was excluded from each list, such that analysis was performed on 56 participants, distributed equally among lists. As in Experiment 1, trials on which the comprehension question was answered incorrectly were excluded, and reading-times that exceeded a threshold of 2.5 standard deviations above the mean reading rate for each region and participant were replaced by that threshold value. This procedure affected 2.2% of trials (2.1–2.3% for individual conditions). The data were entered into a 2×2 repeated-measures ANOVA.

Results

The mean accuracy on comprehension question was 92.6% (91.6-93.8% for individual conditions). There were no reliable differences in accuracy between individual conditions (Fs < 1).

The results from all conditions in Experiment 3 are shown in Fig. 4. The results of the 2×2 ANOVAs are summarized in Table 7.

In the first clause, there was a significant main effect of the constraint factor at the verb-adverb sequence in regions 3 and 4 due to longer reading times in the no-constraint conditions than in the Principle C conditions (region 3: mean for the no-constraint conditions = 406.3 ms, mean for the Principle C conditions = 338.6 ms, 95% CI = 18.3 ms; region 4: 381.2, 345.1 and 14.6 ms respectively). In region 5 there was a significant main effect both of constraint (mean for the no-constraint conditions = 358.0 ms, mean for the Principle C conditions = 338.7 ms, 95% CI = 15.1 ms) and of congruency (mean for the gender-match conditions = 340.9 ms, mean for the gender-mismatch conditions = 355.8 ms, 95% CI = 8.8 ms), but no significant interaction of the factors. These effects were partly due to lexical differences between conditions at the preceding regions and partly unexpected. At the conjunction *while* in region 6 there was a significant main effect of constraint (mean for the no-constraint conditions = 362.3 ms, mean for the Principle C conditions = 337.8 ms, 95% CI = 12.8 ms) and a significant constraint × congruency interaction. Pairwise comparisons revealed that the interaction was caused by an unexpected effect of congruency in the Principle C conditions that contained identical lexical material until that point (region 6: mean for the gender-match condition = 328.3 ms, mean for the gender-mismatch condition =
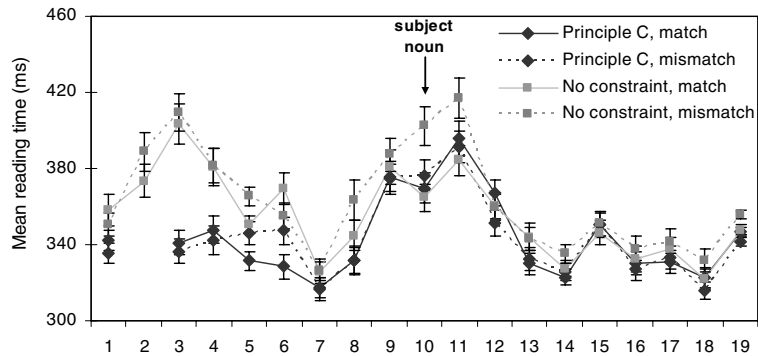
He (she)$_1$ / [Principle C] *or* His (her)$_1$ / managers$_2$ / [no-constraint]
/ chatted$_3$ / amiably$_4$ / with some fans$_5$ / while$_6$ / the$_7$ / talented,$_8$ / young$_9$ / quarterback$_{10}$
/ signed$_{11}$ / autographs$_{12}$ /for$_{13}$ /the kids,$_{14}$ / but$_{15}$ / Steve (Carol)$_{16}$ / wished$_{17}$ / the$_{18}$ / children's
charity event would end soon so he could go home$_{19}$.

Fig. 4. Mean reading times for the Principle C and no-constraint conditions from Experiment 3. The arrow marks the position of the critical gender-marked head-noun. The error bars represent one standard error above/below the participant mean at each region. The arrow marks the critical second subject noun phrase.

347.4 ms, 95% CI = 17.0 ms) that was not observed in the no-constraint conditions (mean for the gender-match condition = 369.6 ms, mean for the gender-mismatch condition = 354.9 ms, 95% CI = 17.2 ms). The effect of constraint was marginally significant at the determiner in region 7 (mean for the no-constraint conditions = 326.0 ms, mean for the Principle C conditions = 316.9 ms, 95% CI = 9.6 ms), and was significant at the first adjective in region 8 (mean for the no-constraint conditions = 354.1 ms, mean for the Principle C conditions = 331.6 ms, 95% CI = 16.6 ms). Importantly, there were no effects of congruency in the two regions preceding the critical noun and similarly, no effects of the constraint factor in the region immediately preceding the critical noun.

At the critical noun in region 10 there was a main effect of congruency (mean for the gender-match conditions = 367.1 ms, mean for the gender-mismatch conditions = 389.4 ms, 95% CI = 14.1 ms) and a significant constraint × congruency interaction. Separate pairwise comparisons of the Principle C and no-constraint conditions revealed a strong effect of congruency in the no-constraint pair in the predicted direction (region 10: mean for the gender-match condition = 364.6, mean for the gender-mismatch condition = 402.5, 95% CI = 18.1 ms). No corresponding effect in the Principle C pair (region 10: mean for the gender-match condition = 369.6, mean for the gender-mismatch condition = 376.4 ms, 95% CI = 19.9 ms).

The constraint × congruency interaction was also significant at the word following the subject noun in region 11. Once again, pairwise comparisons within each level of the constraint factor showed that the interaction was due to the presence of a significant effect of congru-

ency in the no-constraint conditions (mean for the gender-match condition = 384.4, mean for the gender-mismatch condition = 417.0, 95% CI = 24.9 ms), and no corresponding effect in the Principle C conditions (mean for the gender-match condition = 395.6, mean for the gender-mismatch condition = 391.3, 95% CI = 22.4 ms). There were no other significant effects in the remainder of the second clause or anywhere in the final clause.

Discussion

The main finding of Experiment 3 was the presence of a gender mismatch effect in the no-constraint condition at the critical gender-marked head noun. The immediacy and robustness of this effect in Experiment 3 is important in light of the fact that the corresponding effect was marginally significant in Experiment 1 and delayed in Experiment 2. The gender mismatch effect was present despite the fact that there was no independent prediction for an antecedent position at the time when the cataphoric pronoun was encountered. This suggests that the parser's active search for pronoun antecedents is not a strictly instantaneous process that occurs when the pronoun is first encountered, but is rather a continuous process that applies whenever it becomes possible to anticipate a new potential antecedent position. In this respect, therefore, active search for pronoun antecedents resembles active search for gaps in filler-gap dependencies.

In contrast, the gender mismatch effect was not found at the critical second subject or in any other region in the second clause in the Principle C conditions. These results corroborate the conclusion drawn from Experiments 1

Table 7
Results of 2×2 ANOVAs and *min F'* values for Experiment 3

	By participants				By items				*Min F'*		
	df	MSE	F1	p	df	F2	p	df	min F'	p
Region 6 (while)										
Constraint	1,55	2283.0	13.7	0.001	1,24	14.9	0.001	1,69	7.1	0.009
Congruency	1,55	2230.7	<1	0.825	1,24	<1	0.832	1,65	<1	0.878
Constraint times congruency	1,55	1843.1	7.6	0.008	1,24	11.2	0.003	1,75	4.5	0.037
Region 7 (the)										
Constraint	1,55	1293.0	3.6	0.062	1,24	4.2	0.051	1,71	2.0	0.167
Congruency	1,55	1205.2	<1	0.936	1,24	<1	0.851	1,71	<1	0.941
Constraint times congruency	1,55	1500.1	<1	0.652	1,24	<1	0.896	1,28	<1	0.900
Region 8 (talented)										
Constraint	1,55	3848.1	6.2	0.016	1,24	11.4	0.003	1,78	4.0	0.049
Congruency	1,55	2863.8	1.3	0.260	1,24	1.5	0.229	1,71	<1	0.405
Constraint times congruency	1,55	3604.4	1.4	0.250	1,24	1.1	0.316	1,60	<1	0.445
Region 9 (young)										
Constraint	1,55	3726.0	1.0	0.326	1,24	1.2	0.280	1,72	<1	0.463
Congruency	1,55	1963.0	<1	0.779	1,24	<1	0.928	1,29	<1	0.931
Constraint times congruency	1,55	3876.2	<1	0.655	1,24	<1	0.576	1,77	<1	0.725
Region 10 (quarterback)										
Constraint	1,55	2772.2	2.5	0.118	1,24	3.0	0.097	1,71	1.4	0.246
Congruency	1,55	3077.2	8.7	0.005	1,24	4.9	0.037	1,52	3.1	0.082
Constraint times congruency	1,55	1949.5	8.1	0.006	1,24	6.2	0.021	1,60	3.5	0.066
Region 11 (signed)										
Constraint	1,55	5072.1	<1	0.468	1,24	<1	0.431	1,71	<1	0.591
Congruency	1,55	3125.3	2.9	0.095	1,24	2.8	0.110	1,66	1.4	0.239
Constraint times congruency	1,55	4710.2	4.2	0.045	1,24	5.7	0.026	1,74	2.4	0.124
Region 12 (authographs)										
Constraint	1,55	2112.0	<1	0.980	1,24	<1	0.935	1,65	<1	0.981
Congruency	1,55	1405.5	3.9	0.052	1,24	<1	0.408	1,33	<1	0.444
Constraint times congruency	1,55	2298.9	2.0	0.161	1,24	1.3	0.258	1,56	<1	0.373
Region 13										
Constraint	1,55	1642.1	4.4	0.041	1,24	2.9	0.103	1,56	1.7	0.193
Congruency	1,55	1175.3	<1	0.866	1,24	<1	0.845	1,74	<1	0.898
Constraint times congruency	1,55	1593.0	<1	0.886	1,24	<1	0.606	1,63	<1	0.890
Region 14										
Constraint	1,55	2114.9	1.3	0.259	1,24	<1	0.407	1,51	<1	0.499
Congruency	1,55	2606.5	4.3	0.042	1,24	1.8	0.192	1,45	1.3	0.263
Constraint times congruency	1,55	3352.8	<1	0.562	1,24	<1	0.871	1,28	<1	0.875
Region 15 (but)										
Constraint	1,55	1847.7	<1	0.748	1,24	<1	0.617	1,79	<1	0.786
Congruency	1,55	1536.3	<1	0.740	1,24	<1	0.742	1,67	<1	0.814
Constraint times congruency	1,55	1672.7	<1	0.460	1,24	<1	0.794	1,30	<1	0.805
Region 16 (3rd subject)										
Constraint	1,55	1530.2	1.5	0.224	1,24	1.9	0.183	1,72	<1	0.363
Congruency	1,55	1227.0	<1	0.987	1,24	<1	0.865	1,56	<1	0.987
Constraint times congruency	1,55	1604.4	<1	0.398	1,24	<1	0.493	1,56	<1	0.592
Region 17										
Constraint	1,55	1992.0	1.6	0.205	1,24	1.8	0.190	1,69	0.9	0.355
Congruency	1,55	1251.7	0.1	0.704	1,24	0.2	0.661	1,74	0.1	0.773
Constraint times congruency	1,55	1799.6	0.2	0.664	1,24	0.0	0.911	1,27	0.0	0.914

and 2 that the initial set of antecedents for a cataphoric pronoun does not contain noun phrase candidates that violate Principle C. Note that this does not necessarily exclude the possibility that the 'illicit' noun phrases might be considered at some later point, before they are ultimately rejected, although we find no evidence of such 'late' effects (see Sturt, 2003 for evidence of late effects involving forwards anaphora).

General discussion

Our primary concern in this study was to investigate the extent to which mechanisms for processing backwards anaphoric dependencies match the mechanisms that have been previously reported for the processing of filler-gap dependencies, with a focus on mechanisms of active search for pronoun antecedents and on the impact of grammatical constraints on coreference relations. Results from our three off-line and three on-line experiments together yield a fairly consistent picture.

The three off-line rating studies all showed that judgments of coreference are substantially degraded when a pronoun c-commands its antecedent, as predicted by the Principle C constraint. These results show that Principle C captures genuine acceptability contrasts, but leave open the question of whether the constraint impacts the real-time construction of coreference relations. The three on-line experiments all used the gender mismatch paradigm (Aoshima et al., submitted for publication; Van Gompel & Liversedge, 2003) to test where active dependency formation occurs. All three experiments found a gender mismatch effect in the no-constraint conditions and no gender mismatch effect in the Principle C conditions. These findings are therefore consistent with the hypothesis that upon encountering a cataphoric pronoun the parser actively searches for an antecedent for the pronoun, except in positions that are excluded by grammatical constraints. This in turn suggests that constraints on coreference apply during the earliest stages of sentence processing.

The results of Experiment 1 showed the contrast between the no-constraint conditions and the Principle C conditions, and a fifth control condition demonstrated that the gender mismatch effect in the no-constraint conditions could not simply be an effect of additional discourse complexity. However, in this experiment an additional difference in structure predictability was confounded with the difference in the relevance of Principle C. The lack of a gender mismatch effect in the Principle C conditions could have been due to the effects of Principle C, or it may have reflected the fact that the critical antecedent position could not be reliably predicted far in advance. Experiments 2 and 3 were designed to match the conditions in terms of predictability. Experiment 2 contained structures that shared equally strong indicators (prior to

the pronoun) of the upcoming critical subject position, while in Experiment 3 the critical antecedent position was equally unpredictable at the pronoun across the two pairs of conditions. In both experiments we obtained essentially the same results as in Experiment 1: a significant reading time slowdown associated with gender mismatch in the no-constraint conditions, but no corresponding slowdown in the Principle C conditions. These findings suggest that the parser considered the critical subject noun phrase as a potential antecedent in advance of bottom-up morphological information, but only if the subject position was not ruled out by a grammatical constraint on coreference.

These findings clearly parallel results from previous work on filler-gap dependencies that has shown that those dependencies are parsed using an active search mechanism that posits gaps in advance of bottom-up confirmation, but does not posit gaps in positions that are excluded by island constraints on long-distance dependencies (Phillips, 2006; Stowe, 1986; Traxler & Pickering, 1996). Backwards anaphora constructions have the same dependent-element-first property as filler-gap dependencies, but differ from filler-gap dependencies in a number of respects, most notably the fact that potential antecedent noun phrases are overt and hence easier to identify based on bottom-up information than phonologically null gaps. The similarity in the processing mechanisms for these two different-looking types of constructions raises the possibility that the parser might draw on a single mechanism for all linguistic dependencies that share this same basic dependent-element-first structure. It also suggests that the active search mechanism is not a direct consequence of the fact that gaps are difficult to identify using bottom-up information alone, since the same mechanism is found in the search for antecedents of pronouns.

In light of the evidence from Experiment 3 for continuous (rather than instantaneous) resolution of cataphoric pronouns, the difference in the amplitude and timing of the gender mismatch effect in Experiment 2 relative to the other two experiments cannot be accounted in terms of the weaker predictability of the second subject position in that experiment. Instead, the strength of the mismatch effect may be dependent on the distance of the critical subject position from its structural predictor. This idea was supported by the stronger mismatch effect observed in Experiment 3. Although the critical subject position in that experiment was entirely unpredictable at the point when the pronoun was first processed, the subsequent subordinator (e.g., *while*) made it possible to predict the head of the subject noun phrase three words before it was confirmed by bottom-up information. This increased distance may have enabled the parser to fully process the subordinator and the accompanying structural prediction for a subject position. Although it is possible that the parser did not initiate

processing of the pronoun-antecedent relation until bottom-up information about the antecedent was encountered, as assumed by Van Gompel and Liversedge (2003), this would entail that the consequences of grammatical constraints such as Principle C are computed faster than the gender of the antecedent can be identified, and would also make it more difficult to account for the delayed gender mismatch effect in Experiment 2.

Generality of active search mechanism

To understand how the parser actively constructs long-distance dependencies in advance of bottom-up information it is important to know under what circumstances dependency formation can be initiated. In the domain of *wh*-dependencies there is good reason for the parser to initiate the search for a gap as soon as a *wh*-phrase is encountered, since the *wh*-phrase itself provides sufficient evidence that a gap must occur somewhere in the sentence. Matters are less straightforward in the case of backwards anaphora, since the occurrence of a pronoun early in a sentence does not normally guarantee that an antecedent for the pronoun will appear later in the sentence. Furthermore, the occurrence of a pronoun does not normally guarantee that even a potential antecedent position will appear. However, the results of Experiment 3 indicate that active search can be initiated even when evidence for an upcoming potential antecedent position does not occur until a few words after the pronoun. Of course, we cannot at present exclude the possibility that the generality of active search in our studies was related to the fact that the target and filler items in our studies were designed to ensure that all pronouns found an intra-sentential antecedent. This aspect of the design was well-motivated, based on our interest in the absence of gender mismatch effects in the Principle C conditions, but it may have exaggerated the evidence for active search.

A related question about the generality of active search involves the range of positions where the parser considers potential antecedents to occur. In the case of *wh*-constructions previous evidence indicates that gaps are actively posited in object positions (Stowe, 1986) and subject positions (Lee, 2004) of the clause that is local to the *wh*-phrase, and also in embedded clauses that could not have been anticipated at the point when the *wh*-phrase was first encountered (Aoshima et al., 2004; Frazier & Clifton, 1989; Phillips, 2006; Phillips, Kazanina, & Abada, 2005). It should be noted that the studies reported here share with previous studies on backwards anaphora processing (Aoshima et al., submitted for publication; Van Gompel & Liversedge, 2003) the limitation that they show active construction of anaphoric dependencies only in subject positions. This limitation does not undermine our conclusions about the effects of Principle C, but it leaves open the

question of whether the active search for pronoun antecedents is as general as the active search for gaps.

Temporal priority for structural information

As discussed in the Introduction, the existence of gender mismatch effects implies that the parser constructs anaphoric dependencies before gender information about a potential antecedent noun phrase is fully processed. Our current results add to this the finding that structural constraints on coreference such as Principle C are implemented before gender information is processed, providing further evidence for a temporal priority for structural information in the processing of backwards anaphora. Van Gompel and Liversedge (2003) argue that this timing advantage for syntactic information provides evidence against 'strongly interactive' processing models and favors 'modular' models that give priority to syntactic evidence (p. 138). We would argue, however, that this conclusion is not required, since the temporal priority for syntactic information in processing backwards anaphora can be explained without the need to appeal to architectural constraints that specifically delay the use of certain types of information. We suggest that structural information has priority in this domain because it is often the only type of information about a potential antecedent that is reliably derivable before bottom-up information about the noun phrase is encountered. If the parser encounters structural evidence for an upcoming clause, then it can immediately and reliably predict that the clause will have a subject noun phrase, and it could also immediately evaluate whether that subject noun phrase is in a position that makes it a structurally licit or illicit antecedent for a previously encountered pronoun. This information could all be computed before any bottom-up information about the subject noun phrase is encountered. In contrast, other information about the subject noun phrase such as its morphological features and its semantic match to a previously encountered pronoun cannot be evaluated except via bottom-up information, because these properties are not structurally predictable. Thus, under this view syntactic information plays a crucial role because it enables the parser to make predictions about upcoming material earlier than any other type of information, and hence there is no need to impose architectural constraints that force certain information types to have priority. Note that this line of reasoning might apply differently in languages that display richer morphological agreement than English, such that it may be possible to reliably predict morphological properties of an upcoming noun in advance of the noun itself.

Studies on the processing of forwards anaphora have also explored the issue of priority for structural information, and the discussion has focused on the question of

whether structural constraints on coreference immediately restrict the parser's selection of candidate antecedents for pronouns and reflexives. Some studies have found immediate effects of structural constraints in processing forwards anaphora (e.g., Nicol & Swinney, 1989; Sturt, 2003), while others have reported effects of ungrammatical antecedents on the processing of pronouns (e.g., Badecker & Straub, 2002; Kennison, 2003; see also Runner, Sussman, & Tanenhaus, 2003). Sturt (2003) proposes a 'binding-as-defeasible-filter' model that attributes the differing results across studies to different processing stages. Sturt suggests that syntactic constraints on coreference are respected at an initial phase of processing, but may be violated at later stages. He observes time-course differences in his own studies in the effects of structurally accessible and inaccessible antecedents. Further, Sturt's account may capture the fact that in previous studies even highly salient structurally inaccessible antecedents have failed to elicit a gender mismatch effect, a measure of early anaphor resolution (for more detailed review see Kazanina, 2005). In the respect that we observed no effects of structurally inaccessible antecedents in our studies, our results are compatible with Sturt's account. Note, however, that care should be taken in comparing the resolution of forwards and backwards anaphora, due to differences in their temporal dynamics and due to differences in the representations that must be searched for an antecedent. Forwards anaphora involves a retrospective search for an antecedent in memory, and multiple candidate antecedents may be considered simultaneously. Backwards anaphora, in contrast, involves a prospective search for an antecedent. Consequently, potential antecedents do not need to be retrieved from memory, and may be evaluated in succession. In contrast to some studies of forwards anaphora, our studies found no delayed effect of structurally inaccessible antecedents. This difference may reflect either the temporal dynamics of backwards anaphora resolution, or the different representations that are searched for an antecedent.

Analogous effects of giving priority to one type of information over others during anaphora processing have also been discussed by Sanford and Garrod (1989) for cross-sentential forwards anaphor resolution. Their findings suggest that bottom-up morphological information about a noun phrase and a pronoun can set in motion a process they call 'bonding': the formation of a tentative (and ultimately incorrect) dependency in cases like "Harry was sailing to Ireland. It sank without trace" on the basis of the shared singular morphology of *Ireland* and *it* (originally reported in Sanford, Garrod, Lucas, & Henderson, 1984). These findings are similar to ours in the sense that one type of information has temporal priority over others in the processing of anaphoric dependency. However, whereas the Sanford and Garrod's effect relies on bottom-up morpho-

logical information to begin the resolution process, our cases crucially depend on the combined use of information from the pronoun and structural expectations to create a dependency in advance of the bottom-up information, although the full details are filled in only after the bottom-up information about the antecedent is encountered. Therefore, we think that these two types of effects are probably driven by distinct mechanisms.

Implications for ungrammatical representations

Finally, our results are also relevant to questions about the nature of grammatical constraints and how they are implemented. In some linguistic frameworks there is a clear distinction between the range of structures that the grammar generates and the range of structures that satisfy all grammatical constraints or 'filters.' This overgenerate-and-filter property is explicitly advocated in such frameworks as Optimality Theory (Prince & Smolensky, 1993/2004) and Government and Binding Theory (Chomsky, 1981). A different perspective on illicit representations is that ill-formedness is the result of an inability to generate a structure, such that there is a closer match between the set of well-formed structures and the set of representable structures. This is more characteristic of analyses in such approaches as Head-Driven Phrase Structure Grammar (HPSG: Pollard & Sag, 1994) and the Minimalist Program (Chomsky, 1995). To the extent that parsing is seen as reflective of the linguistic knowledge modeled by these theories, our results are consistent with the latter view, at least for Principle C. Speakers do not appear to generate representations that violate Principle C and then reject them based on application of a later filter. Instead, it appears that Principle C violations simply do not occur to speakers of English. Note that this finding contrasts with the conclusion from an interesting but controversial set of findings by Freedman and Forster (1985), who claimed based on the results of a series of sentence-matching studies that the parser considers illicit filler-gap dependencies that violate syntactic island constraints (for alternative interpretations see Crain & Fodor, 1987; Stowe, 1992). Note, also, that our results cannot exclude the possibility that the parser does temporarily construct representations that violate Principle C, but successfully filters these representations before any bottom-up information about an antecedent is presented.

Conclusion

We have shown that upon encountering a cataphoric pronoun the parser initiates an active search for an antecedent in the following material. Importantly, during this search the parser does not consider positions that

are inaccessible antecedent positions due to a structural constraint on coreference, Principle C. These results support a view whereby syntactic constraints immediately restrict active search processes, whether the search is for a thematic position, as in *wh*-dependencies, or for a referent, as in backwards anaphora. More generally, our results suggest that the mechanisms used for processing *wh*-dependencies and referential dependencies like backwards anaphora might be largely identical.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jml.2006.09.003.

References

Aoshima, S. (2003). *The grammar and parsing of wh-dependencies*. PhD dissertation, University of Maryland, College Park.

Aoshima, S., Phillips, C., & Weinberg, A. S. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language, 51*, 23–54.

Aoshima, S., Yoshida, M., & Phillips, C. (submitted for publication). Incremental processing of coreference and binding in Japanese.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphora. *Journal of Experimental Psychology: Learning Memory and Cognition, 28*, 748–769.

Baker, M. A. (1991). On some subject/object non-asymmetries in Mohawk. *Natural Language and Linguistic Theory, 9*, 537–576.

Bourdages, J. S. (1992). Parsing complex NPs in French. In H. Goodluck & M. S. Rochemont (Eds.), *Island constraints: Theory, acquisition and processing* (pp. 61–87). Netherlands: Kluwer Academic Publishers.

Brody, M. (1994). *Phrase structure and dependence*. London: Ms. University College.

Bruening, B. (2001). *Syntax at the edge: cross-clausal phenomena and the syntax of Passamaquoddy*. PhD dissertation, MIT.

Chomsky, N. (1973). Conditions on transformations. In S. Anderson & P. Kiparsky (Eds.), *A festschrift for morris halle* (pp. 232–286). New York: Holt, Rinehart & Winston.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. (1995). *The minimalist program*. Cambridge, Mass: MIT Press.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335–359.

Clifton, C., Kennison, S. M., & Albrecht, J. E. (1997). Reading the words *him* and *her*: implications for parsing principles based on frequency and on structure. *Journal of Memory and Language, 36*, 276–292.

Cowart, W., & Cairns, H. S. (1987). Evidence for an anaphoric mechanism within syntactic processing: some reference relations defy semantic and pragmatic constraints. *Memory & Cognition, 15*, 318–331.

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers. In D. R. Dowty, L. Karttunen, & A. M. Zicky (Eds.), *Natural language parsing: psychological, computational, and theoretical perspectives* (pp. 94–128). Cambridge: Cambridge University Press.

Crain, S., & Fodor, J. D. (1987). Sentence matching and overgeneration. *Cognition, 26*, 123–169.

Crain, S., & McKee, C. (1985). The acquisition of structural restrictions on anaphora. In S. Berman, J. Choe, & J. McDonough (Eds.), *Proceedings of NELS15*. Amherst, MA: GLSA Publications.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry, 9*, 427–473.

Frazier, L. (1987). Sentence processing: a tutorial review. In M. Coltheart (Ed.), *Attention and performance XII*. Hillsdale, N.J.: Erlbaum.

Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes, 4*, 93–126.

Freedman, S. E., & Forster, K. I. (1985). The psychological status of overgenerated sentences. *Cognition, 19*, 101–131.

Garnsey, S. M., Tanenhaus, M. K., & Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research, 18*, 51–60.

Gordon, P. C., & Hendrick, R. (1997). Intuitive knowledge of linguistic co-reference. *Cognition, 62*, 325–370.

Harris, C. L., & Bates, E. A. (2002). Clausal backgrounding and pronominal reference: a functionalist approach to c-command. *Language and Cognitive Processes, 17*, 237–269.

Heim, I. (1992). Anaphora and semantic interpretation: a reinterpretation of Reinhart's approach. In U. Sauerland & O. Percus (Eds.), *The interpretive tract* (pp. 205–246). Cambridge, MA: MIT Working Papers in Linguistics.

Hirst, W., & Brill, G. A. (1980). Contextual aspects of pronoun assignment. *Journal of Verbal Learning and Verbal Behavior, 19*, 168–175.

Huang, C.-T. J. (1982). Move-wh in a language without wh-movement. *The Linguistic Review, 1*, 369–416.

Jelinek, E. (1984). Empty categories, case, and configurationality. *Natural Language and Linguistic Theory, 2*, 39–76.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes and in reading comprehension. *Journal of Experimental Psychology: General, 3*, 228–238.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*, 159–201.

Kazanina, N. (2005). *The acquisition and processing of backwards anaphora*. Ph.D. dissertation, University of Maryland, College Park.

Kazanina, N., & Phillips, C. (2001). Coreference in child russian: distinguishing syntactic and discourse constraints. In A. Do, L. Domiínguez, & A. Johansen (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 413–424). Somerville MA: Cascadilla Press.

Kennison, S. M. (2003). Comprehending the pronouns *her*, *him*, and *his*: Implications for theories of referential processing. *Journal of Memory and Language, 49*, 335–352.

Lee, M. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research, 33*, 51–73.

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57*, 203–220.

Nicol, J. L., Fodor, J. D., & Swinney, D. (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory & Cognition, 20*, 1229–1238.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18*, 5–19.

Phillips, C. (2006). The real-time status of island phenomena. *Language, 82*, 4.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research, 22*, 407–428.

Phillips, C. & Wagers, M. (in press). Relating structure and time in linguistics and psycholinguistics. In: G. Gaskell (ed.), *Oxford handbook of psycholinguistics*. Oxford, UK: Oxford University Press.

Pickering, M. J. (1993). Direct association and sentence processing: a reply to Gorrell and to Gibson and Hickok. *Language and Cognitive Processes, 8*, 163–196.

Pickering, M. J., Barton, S., & Shillcock, R. (1994). Unbounded dependencies, island constraints and processing complexity. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 199–224). London: Lawrence Erlbaum.

Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Prince, A., & Smolensky, P. (1993/2004). *Optimality theory: Constraint interaction in generative grammar*. Technical Report, Rutgers University and University of Colorado at Boulder, 1993. Revised version published by Blackwell, 2004. Rutgers Optimality Archive 537.

Reinhart, T. (1983). *Anaphora and semantic interpretation*. London: Croom Helm.

Reinhart, T., & Reuland, E. (1993). Reflexivity. *Linguistic Inquiry, 24*, 657–720.

Resnik, P. & Elkiss, A. (2003). *The linguist's search engine: Getting started guide*. Technical Report LAMP-TR-108, University of Maryland.

Ross, J. R. (1967). *Constraints on variables in syntax*. Ph.D. dissertation, MIT.

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2003). Assignment of reference to reflexives and pronouns in picture noun phrases: evidence from eye movements. *Cognition, 89*, B1–B13.

Sanford, A. J., Garrod, S. C., Lucas, A., & Henderson, R. (1984). Pronouns without antecedents? *Journal of Semantics, 2*, 303–318.

Sanford, A. J., & Garrod, S. C. (1989). What, when and how? Questions of immediacy in anaphoric reference resolution. *Language and Cognitive Processes, 4*, 235–262.

Speas, M. (1990). *Phrase structure in natural language*. Dordrecht: Kluwer Academic Publishers.

Stowe, L. (1986). Evidence for online gap creation. *Language and Cognitive Processes, 1*, 227–245.

Stowe, L. A. (1992). The processing implementation of syntactic constraints: the sentence matching debate. In H. Goodluck & M. Rochemont (Eds.), *Island constraints: Theory, acquisition and processing* (pp. 419–443). Dordrecht: Kluwer Academic Publishers.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language, 48*, 542–562.

Sussman, R. S., & Sedivy, J. C. (2003). The time-course of processing syntactic dependencies: evidence from eye-movements during spoken wh-questions. *Language and Cognitive Processes, 18*, 143–163.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: an eye-tracking study. *Journal of Memory and Language, 35*, 454–475.

Van Gompel, R. P. G., & Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 128–139.

van Hoek, K. (1997). *Anaphora and conceptual structure*. Chicago: University of Chicago Press.