



## **Discours**

Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics

**12 | 2013**

**Varia**

---

# Discourse Coherence in Pronoun Resolution

Clare Patterson

---



### **Electronic version**

URL: <http://journals.openedition.org/discours/8820>

DOI: 10.4000/discours.8820

ISSN: 1963-1723

### **Publisher:**

Laboratoire LATTICE, Presses universitaires de Caen

### **Electronic reference**

Clare Patterson, « Discourse Coherence in Pronoun Resolution », *Discours* [Online], 12 | 2013, Online since 10 July 2013, connection on 30 April 2019. URL : <http://journals.openedition.org/discours/8820> ; DOI : 10.4000/discours.8820

---



*Discours* est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.





Revue de linguistique, psycholinguistique et informatique

<http://discours.revues.org/>

## Discourse Coherence in Pronoun Resolution

---

Clare Patterson

Potsdam Research Institute for Multilingualism  
University of Potsdam, Germany

.....  
Clare Patterson, « Discourse Coherence in Pronoun Resolution », *Discours* [En ligne], 12 | 2013, mis en ligne le 10 juillet 2013.

.....  
URL : <http://discours.revues.org/8820>

.....  
Titre du numéro : *Varia*  
Coordination : Saveria Colonna et Sarah Schimke

**revues.org**  
CENTRE POUR L'ÉDITION ÉLECTRONIQUE OUVERTE  
CENTRE FOR OPEN ELECTRONIC PUBLISHING



 Presses  
universitaires  
de Caen



# Discourse Coherence in Pronoun Resolution

---

Clare Patterson

Potsdam Research Institute for Multilingualism  
University of Potsdam, Germany

.....

According to “Centering Theory”, an entity that links to the prior discourse could receive a boost in the current prominence ranking. Such a boost will affect pronoun processing because pronouns tend to seek out the most prominent entity for reference. The current paper presents a reanalysis of data from a study that investigated how pronoun processing was affected by the prior discourse. In an eye-movement monitoring experiment, English speakers were presented with texts that manipulated (1) the pronoun’s gender match with possible antecedents and (2) the appearance of the antecedents in prior discourse. The results showed that prior discourse affected eye-movement measures from the pronoun region onwards, but the antecedents were differently affected by this manipulation. The results can be accounted for if the overall discourse coherence is taken into account, especially the number of switches in attention that a reader/hearer has to make. More attention switches in a relatively short discourse may have created uncertainty about the local discourse topic, making pronoun processing more problematic. This study shows that the prior discourse can influence the processing of pronouns, but processing is also affected by coherence.

**Keywords:** pronouns, reference, coherence, processing, salience, eye-movements

## 1. Introduction

- 1 The current paper presents a new analysis of data from a study that investigated how pronoun processing is affected by the prior discourse (reported in Patterson, 2012). The current paper benefits both from insights from related experiments (reported in Cunnings et al., submitted) and also from a more robust analysis technique which reveals patterns that were not apparent under the previous analysis. This has led to a further understanding of the experimental results, and for this reason the paper has been entirely rewritten to incorporate these aspects. A summary of the previous findings from Patterson (2012) is given at the start of the discussion section and differences in the analysis are pointed out in the analysis section.
- 2 The processing of pronouns has received a considerable amount of attention in psycholinguistic research. In particular, third-person subject pronouns (such as *he/she/they* in English) have been the target of research, since they refer to understood or explicit referents within the discourse. It is generally agreed that these subject pronouns refer to the most prominent entity in the preceding discourse (see Garnham, 2001 for a review). That being said, a large number of cues and signals to discourse prominence have been suggested, such as subjecthood (Crawley et al., 1990; Järvikivi et al., 2005; Kaiser & Trueswell, 2008), first-mention (Gernsbacher & Hargreaves, 1988; Arnold, 2001; Arnold et al., 2007), and topichood (Cowles et al., 2007; Wubs et al., 2009; Kaiser, 2011; Van Rij, 2012). Most researchers agree that these cues interact, but no comprehensive model of how this takes place has been presented.

3 The notion of topichood is interesting because its role in pronoun processing appears to operate on a sentence level and also on the level of the wider discourse. A precise definition of the notion of topic is still under debate. In terms of the sentence, topics can be thought of as “a sentence-initial constituent which is predicated about” (Schwabe & Winkler, 2007: 2). In English topics are often, but not always, subjects. They are thought of as cognitively prominent entities, and as such they are good antecedents for pronouns. The cognitive prominence of discourse topics and sentence topics was tested by Cowles et al. (2007) in a cross-modal priming study, who found that both discourse and sentence topics were activated at the point of a subsequent pronoun. They defined discourse topic as a referent that a set of propositions is about, and more specifically for their own study, as “a referent that has been a sentential topic for more than one sentence without any intervening sentence topics” (Cowles et al., 2007: 6). This implies that pronouns seeking a prominent antecedent will take into consideration prominence not just within a sentence but also within the preceding discourse. Note that there are some limits on what can be considered a topic. Quantified phrases, for example, are generally thought not to be topics (see Hinterwimmer et al., 2009, for a discussion), and entities in embedded clauses are unlikely to be thought of as sentence topics because they do not appear in the matrix clause.

4 Topicality is an important element of “Centering Theory” (CT) (Grosz et al., 1995). CT is a framework within which multiple factors affecting salience are accounted for. CT was not originally formulated specifically as a model for pronoun processing, but as an account of textual coherence. However, the framework has received much attention in pronoun processing because it proposes that links between discourse entities are the key to discourse coherence, and reference and prominence are as such very important in this account. CT uses the factors of subjecthood and first-mention, as well as previous discourse status, to provide a ranking of the discourse entities at different points throughout a discourse. Thus, if an entity is the subject, or is first-mentioned, or has the status of “backward-looking center” (Cb) (that is, the entity that links the current utterance to the last one, usually by appearing in some form in the current and previous utterance) then it receives a boost in its prominence ranking which is represented as the set of “forward-looking centers” (Cf)<sup>1</sup>. The top-ranked entity in the Cf of an utterance is assumed to be the most prominent one, and this is commonly assumed to be the topic. CT does not specify how subjecthood, first-mention and previous discourse status interact with one another or whether they compete. It does however capture particular observations about discourse referents and provides a rough metric by which they can be assessed. One claim of CT is that a topic shift is more costly in processing terms than maintaining the same topic. Topic shift occurs when an entity that was not highly ranked in a previous utterance is the most highly ranked entity in the current utterance. One further observation

1. See, for example, the discussion in Gordon et al. (1993) about Cb status and prominence in the Cf.

that is frequently described in terms of CT is the repeated-name penalty (Gordon et al., 1993). CT makes the assumption that the top-ranked entity of a previous utterance should preferentially be realized as a pronoun (if anything in the current utterance is realized as a pronoun). This was tested by Gordon et al. (1993) who presented utterances in which the highest-ranked entity from the previous utterance was realized as a repeated name in the current utterance rather than a pronoun. This led to a disruption in reading times on the repeated name, in comparison to utterances that contained a pronoun<sup>2</sup>.

5 While many individual factors have been identified as contributing to discourse prominence, a precise understanding of how these factors interact, and what the underlying processes are that drive the observed effects, has not yet been reached. The current study aims to explore the influence of the previous discourse on pronoun resolution. As far as possible this will be described in terms of CT, but the application of CT to the current materials is not unproblematic as is discussed below. The study can also be described in more theory neutral terms as exploring the influence of discourse topichood on pronoun processing. Interestingly, while many studies have looked at pronouns with two (or more) potential antecedents, not many have looked at two sentence-internal antecedents that are both subjects. Two recent eye-movement monitoring experiments, reported in Cunnings et al. (submitted) (henceforth Cunnings et al.), used such materials and found evidence of a recency preference while investigating the role of variable-binding and coreference in pronoun resolution in English. The Cunnings et al. study will be described and discussed below because the results are informative for the current study, and the materials for the current study are derived from their materials.

6 In Cunnings et al., native speakers of English read texts that consisted of a neutral introductory sentence, a critical sentence containing two potential antecedents and a third-person singular subject pronoun, and a neutral wrap-up sentence. In Experiment 1 both the potential antecedents had the grammatical role of subject in the critical sentence (see example [1]) and in Experiment 2 the first antecedent was in a cleft structure and the second was a grammatical subject (see example [2]). A match/mismatch paradigm was used, such that the pronoun matched in gender with both, neither, or only one of the antecedents presented. The antecedents were either quantifier phrases (QPs) or proper names. QPs were used because the study was investigating a timing hypothesis about variable-binding and coreference (see

---

2. Gordon et al. (1993) claim that the repeated-name penalty is applicable to all definite descriptions, not just names. It should be noted that much experimental work in this area has focused on names (e.g., Gordon et al., 1993; Gordon & Scarce, 1995; Rose, 2007; Fukumara & Van Gompel, 2010; Gelormini-Lezama & Almor, 2011). This leaves open the question of whether the repeated-name penalty would be seen if definite descriptions were repeated. Secondly, Gordon and Hendrick (1998) claim that in complex constructions, the repeated-name penalty is reduced. This is certainly a logical expectation. If the reader is processing a complex discourse, processing resources for maintaining the topic of an utterance will be more limited. Therefore, reminding the reader/hearer more explicitly about the topic of an utterance by repeating a name might be more helpful than disruptive.

below). The QPs contained role names such as soldier and babysitter<sup>3</sup>. The gender of the proper name and the pronoun was manipulated to create four conditions: double match, QP match, name match and no match.

- [1]   Cunnings et al. Experiment 1 materials  
       The squadron paraded through town. Every soldier who knew that {James/Helen} was watching was convinced that {he/she} should wave as the parade passed. The entire town was extremely proud that day.
  
- [2]   Cunnings et al. Experiment 2 materials  
       The squadron paraded through town. It looked to {James/Helen} that every soldier was completely convinced that {he/she} should wave as the parade passed. The entire town was extremely proud that day.

7     In Experiment 1, there were effects of the gender match of the proper name (James/Helen) to the pronoun in early and late measures in the pronoun region and later measures in the spillover region, as well as some marginal effects in prefinal and final regions. When the proper name mismatched in gender with the pronoun, longer reading times were seen in all these measures and regions. There were no effects of the QP in any of the regions or measures, indicating that readers were not affected if the QP mismatched in gender with the pronoun. Results from the first experiment indicated clearly that the readers were considering the proper name as a potential antecedent but they were not considering the QP. There could be several reasons for such a result. It could, firstly, be because of properties of the antecedents themselves. For example, there could be a general dispreference for linking pronouns with QPs, especially when another potential antecedent (that is not a QP) is also present in the discourse. Burkhardt's (2005) model and results from Carminati et al. (2002) indicate that this could indeed play a role because of the complexity of the QP's representation. It is also the case that QPs, not having specific discourse referent (making coreference impossible and thereby acting as variable binders) may lack strong semantic cues, for example with regard to number, that the proper name referent has. Alternatively, it is possible that the preference for the proper name does not come from properties of the antecedent nouns but from their position relative to the pronoun. Previous studies have shown that the cost associated with resolving a pronoun increases as the distance between pronoun and antecedent increases (Clark & Sengul, 1979; Streb et al., 2004). In order to understand the observed preference further, and attempt to distinguish between positional properties and properties of the antecedents themselves, Experiment 2 was carried out (see example [2]). In this experiment, the QP was nearest the pronoun, and the proper name was in a cleft structure at the beginning of the critical sentence. The same

---

3. The role names had been rated in a pretest according to whether they were typically male or typically female. Only the role names that scored highly (typically male or typically female) were included. Previous studies (Sturt, 2003; Kennison & Trofe, 2003) have shown that readers experience disruption in reading times when presented with nouns with stereotypical gender followed by a mismatching pronoun.



match/mismatch paradigm was used as in Experiment 1. This time, there were effects of the QP at first-pass times, rereading times (marginal) and total viewing times in the pronoun region, and rereading times in the spillover region. Reading times were longer for the QP mismatch conditions than the QP match conditions. In this experiment, then, readers were disrupted when the QP mismatched the pronoun in gender. Overall, the results from Cunnings et al. suggest that it was not properties of the antecedents themselves (or, indeed, the exact mechanism, variable-binding or coreference, by which they were linked to the pronoun) that determined the consideration of the antecedent when the pronoun was encountered. Instead there was a recency preference, regardless of the type of antecedent presented. The most recent antecedent was therefore more prominent than the earlier antecedent.

- 8 The findings of the Cunnings et al. experiments are important for this study not just because of the similarity of materials but also because they show that antecedents in a variable-binding relation to the pronoun and antecedents that are co-referential with the pronoun can both be affected by a recency preference. Whether the recency preference is driven by memory considerations or by discourse properties that prioritise the information in the relative clause, QPs and names were both affected by it during processing. When processing the pronoun, the reader's attention was drawn to the most recent antecedent, which suggests that the different mechanisms of variable-binding and coreference can both be affected by the same factors that control prominence.

## 2. The current study

- 9 The current study (also described in Patterson, 2012) uses materials that are based on the materials from Cunnings et al. Experiment 1, with some modifications, and are described below. The current study investigates how pronoun processing is affected by topichood/Cb status from the prior discourse. Can a discourse referent receive facilitation (in terms of salience) from previous discourse status in the text, as suggested by CT? In theory-neutral terms, how does discourse topichood affect pronoun processing in a subsequent sentence? And how does the previous discourse status interact with other existing preferences: for example, can a boost in prominence from previous mention overcome the recency preference observed in Cunnings et al.<sup>4</sup>
- 10 The experimental materials were comprised of 24 short scenarios of three sentences each. The first sentence was the context sentence, the second was the critical sentence and the third was the wrap-up sentence (see example [3] below).

---

4. The latter question that is addressed in the current study has been changed since the Patterson (2012) paper because of the recency preference that was revealed in the Cunnings et al. experiments. Instead of investigating "pronoun resolution preferences with two potential subject antecedents" (Patterson, 2012: 101), the question has been reframed as the recency preference interacting with cues from the prior discourse.

Each scenario contained a pronoun and two potential antecedents (a QP and a determiner phrase [DP]). Unlike Cunnings et al., the context sentence mentions one of the two potential antecedents: this was systematically varied. The critical sentence contained the pronoun and both potential antecedents. The first potential antecedent in the critical sentence was a QP (e.g., *every soldier*), and the second was a DP (e.g., *the queen*). All the nouns in the QPs were stereotypical role names as in Cunnings et al. All the DPs were nouns with definitional gender. The overall set of QPs did not differ in length or frequency from the set of DPs, although it was not possible to match length and frequency in each individual item. To avoid biases, the texts were composed so that either antecedent could plausibly be the referent of the pronoun, but the gender cues determined the correct antecedent.

[3] Materials for the current study

DP ANTECEDENT, CONTEXT MATCH

The queen felt really quite uneasy about the squadron parade. Every soldier who knew that the queen was watching intently was absolutely convinced that she should wave as the parade passed. [wrap-up sentence]

DP ANTECEDENT, CONTEXT MISMATCH

The soldiers felt really quite uneasy about the squadron parade. Every soldier who knew that the queen was watching intently was absolutely convinced that she should wave as the parade passed. [wrap-up sentence]

QP ANTECEDENT, CONTEXT MATCH

The soldiers felt really quite uneasy about the squadron parade. Every soldier who knew that the queen was watching intently was absolutely convinced that he should wave as the parade passed. [wrap-up sentence]

QP ANTECEDENT, CONTEXT MISMATCH

The queen felt really quite uneasy about the squadron parade. Every soldier who knew that the queen was watching intently was absolutely convinced that he should wave as the parade passed. [wrap-up sentence]

- 11 Two factors were systematically varied to create the four experimental conditions shown above. These factors were *antecedent* (DP or QP) and *context* (match or mismatch). *Antecedent* refers to whether the pronoun matches in gender with the QP antecedent or the DP antecedent. *Context* refers to whether the context sentence matches or mismatches with the antecedent. We have already seen (from Cunnings et al. Experiment 1) that the most recent antecedent in these sentence types (in this case the DP) should be preferred, but that was in the absence of a prior context that established “the soldiers” or “the queen” as a sentence topic. CT suggests that previous discourse status is important in determining the prominence of entities in the current utterance. This is because prominence in the Cf is affected not only by first-mention and subjecthood but also by Cb status. Since the top-ranked entity in Cf will be the preferred referent for a pronoun in a subsequent utterance, the context match condition should have a processing advantage.

### 2.1. Applying CT assumptions to the current materials

12 There are some challenges in applying CT assumptions directly to the materials for the current experiment. The first challenge is the use of QPs such as *every soldier* which c-command the pronoun. As such, the link between the pronoun and the QP is established via variable-binding as opposed to coreference. The link between the DP and the pronoun can only be established via coreference, since the DP does not c-command the pronoun. One of the reasons why two different mechanisms for pronoun resolution have been proposed is that QPs are not referring expressions, and as such they cannot establish or refer to an individual in the discourse representation. In CT this makes it unclear what kind of representation of the QP appears in the ranking of Cf, since discourse referents are normally assumed here. The prominence hierarchy in CT models the state of the hearer's/reader's attention (Grosz et al., 1995). Whether we accept that the mental representation is separable from the semantic representation (e.g., mental models theory, Johnson-Laird, 1983; as applied to anaphora, Garnham & Oakhill, 1992; Garnham, 2001), or whether mental representations are propositional in nature, such as the representations in many semantic theories, there is general agreement that a QP such as *every soldier* does not invoke an individual discourse referent. What must be assumed is that a set of soldiers is invoked or introduced (or sets of sets according to traditional semantic notation) which may be represented along with its verbal argument, possibly in a tripartite structure. Can such representations be ranked as more or less salient in the same way as a normal discourse referent? I propose that the Cunnings et al. study described above provides evidence that a QP antecedent can indeed be influenced by factors contributing to prominence in the same way as a DP antecedent. As such, an attempt should be made to incorporate QPs into CT-based explanations, despite the fact that so far they have not been explicitly included. For the purpose of this study I will assume that a set, with or without its verbal argument, is subject to certain prominence factors and can therefore be represented and ranked in the Cf. I do not claim that QPs are equivalent in all aspects to DPs in terms of their representation. They are clearly different syntactically, semantically and even in their information structural status. I rather claim that, as far as QP antecedents can be influenced by recency, which may in turn be driven by discourse or memory considerations or both, they can be considered to be affected by certain prominence factors and thus should be considered under a CT approach. Another question is whether the DP forms part of the QP set because it is inside the relative clause. Again, because of DP's ability to be prominent enough to create a recency preference in Cunnings et al., I will consider that it should be represented separately in the discourse as a normal entity.

13 The second challenge in applying CT to the current materials is that there is no consensus about what constitutes an utterance. While much debate and empirical testing has addressed the notion of utterances as sentences versus utterances as clauses (e.g., Kameyama, 1998), intervening relative clauses such as those in the

current materials have not been fully addressed (Poesio et al., 2004). I will therefore proceed as follows: the context sentence in the current experiment is a single clause and can be straightforwardly assigned to Utterance 1 (U<sub>1</sub>). The second sentence is long and somewhat complex and will be split into two utterances (U<sub>2</sub> and U<sub>3</sub>). I will keep the QP, the intervening relative clause and the main verb together in U<sub>2</sub>, and start U<sub>3</sub> at the CP following the main verb. This splits a complex sentence into coherent parts and crucially keeps the QP together with the verb, while allowing the complement of the main verb (which is a finite clause) to be separated. The CT analysis from which I derive my CT predictions is as follows:

DP antecedent, context match:

U<sub>1</sub>: The queen felt really quite uneasy about the squadron parade.

No Cb. Cf = {queen}

U<sub>2</sub>: Every soldier who knew that the queen was watching intently was absolutely convinced

Cb = queen. Cf = {queen, soldiers}

U<sub>3</sub>: that she should wave as the parade passed (she = queen)

Cb = queen.

DP antecedent, context mismatch:

U<sub>1</sub>: The soldiers felt really quite uneasy about the squadron parade.

No Cb. Cf = {soldiers}

U<sub>2</sub>: Every soldier who knew that the queen was watching intently was absolutely convinced

Cb = soldiers. Cf = {soldiers, queen}

U<sub>3</sub>: that she should wave as the parade passed (she = queen)

Cb = queen.

QP antecedent, context match:

U<sub>1</sub>: The soldiers felt really quite uneasy about the squadron parade.

No Cb. Cf = {soldiers}

U<sub>2</sub>: Every soldier who knew that the queen was watching intently was absolutely convinced

Cb = soldiers. Cf = {soldiers, queen}

U<sub>3</sub>: that he should wave as the parade passed (he = soldiers)

Cb = soldiers.

QP antecedent, context mismatch:

U<sub>1</sub>: The queen felt really quite uneasy about the squadron parade.

No Cb. Cf = {queen}

U<sub>2</sub>: Every soldier who knew that the queen was watching intently was absolutely convinced

Cb = queen. Cf = {queen, soldiers}

U<sub>3</sub>: that he should wave as the parade passed (he = soldiers)

Cb = soldiers.

- 14 This analysis makes the assumption that the Cb status in U<sub>2</sub> is enough to bring that entity to the top of the Cf rankings. This assumes that the importance of Cb status outweighs the prominence factors that drive the recency preference, and also that it outweighs first-mention, which is often cited as important in the Cf rankings. I propose that first-mention can be discounted in these materials because, as the results of Cunnings et al. showed, it did not influence the pronoun preferences in almost identical materials. Subjecthood can also be discounted because there are two subjects in U<sub>2</sub>. The question remains as to whether recency or Cb status will be more important in the rankings of U<sub>2</sub>. Specific predictions are made in the section below.
- 15 Outside of CT, the materials can be thought of in terms of topichood. In the context sentence, either the soldiers or the queen are the sentence topics. Under the definition of Cowles et al. (2007), whether they are also discourse topics depends on whether an intervening topic is found in the following sentence. In the critical sentence of the materials, which is the sentence containing the pronoun, there is a QP in the matrix clause and a DP in the relative clause. As observed above, QPs are thought not to serve as topics. It is also unlikely that the DP inside the relative clause can be thought of as a sentence topic here, because it is not in the matrix clause. Under this analysis, there is no intervening topic before the pronoun is encountered. If the pronoun is searching for a cognitively prominent discourse or indeed sentence topic, it should find the subject of the context sentence. This leads to the same predictions as the CT analysis above.
- 16 Note, however, that both of these proposals assume that the preferences which were found in the Cunnings et al. study are overridden by the effects of the context sentence. It could be that the very local preferences take priority, or that local recency preferences interact with the prominence from the context sentence. Note too that introducing the context sentence may in fact disrupt the coherence of the texts by introducing competing prominence cues for the entities. For simplicity, I will restrict my predictions to those where the prior discourse overrides recency considerations, and those where it does not.

### 3. Predictions

- 17 Under a CT analysis, if Cb status boosts the prominence of an entity in the Cf of U<sub>2</sub>, the processing of the pronoun will be easier when it matches the top-ranked entity as opposed to when it mismatches. This should lead to disruption in reading measures in the mismatching context conditions.
- 18 The same outcome is expected if pronouns search for a cognitively prominent discourse topic, and very local preferences are outweighed by this.
- 19 If Cb status/discourse topichood does not override the recency preference, there should be an effect of antecedent such that the DP conditions have shorter reading times than the QP conditions, because the DP is the most recent antecedent.

#### 4. Method

20 Scenarios were distributed among four lists in a Latin-square design so that each participant saw 6 items in each condition, and they did not see the same item in more than one condition. The lists were pseudo-randomized with fillers and pseudo-fillers<sup>5</sup> so that items did not appear close together. The item order of the four lists was reversed to create four further lists, so that there were eight lists in total. Participants were 35 native speakers of English who were recruited from the University of Essex and the surrounding community. All were paid £7 for their participation. All participants confirmed that they had not been diagnosed with any language difficulty/disorder. Some participants wore contact lenses or glasses during the experiment, and all reported that they could read from the screen successfully.

21 Eye movement data was recorded using the head-mounted EyeLink II system (by SR Research, Canada). Participants were seated in front of a PC monitor at a distance of 80 cm. They wore a headset carrying two eye cameras and a head camera, with a rubber cap underneath the headset to prevent it from slipping during the experiment. The head camera compensated for small head movements that might be made during recording via the tracking of LEDs at the four corners of the display monitor. The position of the eye cameras was adjusted for each participant. Data from the right eye was recorded, with the pupil being tracked at a 500 Hz sampling rate. There was a drift correction procedure before each trial in which a target appeared in the top left of the screen (above the position where the text would start in the next trial), and participants had to press a button while fixating on the target. After set up, participants were presented with a screen containing instructions, followed by seven practice trials, three of which had practice comprehension questions. Responses to the comprehension questions and to the drift correction screen were made using a USB gamepad connected to the host PC. Participants were instructed to sit still for the duration of the experiment.

22 Following two thirds of the trials there were comprehension questions. This was to ensure that the participants were reading and understanding the trials that were presented. Comprehension questions did not explicitly probe the referent of the pronoun, in order not to draw attention to the purpose of the experiment or encourage strategic reading behaviour. After the eye-tracking part of the

---

5. 36 real fillers and 12 pseudo-fillers were used. The pseudo-fillers were materials for a different experiment (not described here), they contained a pronoun and two potential antecedents, but were of a different structure to the experimental items. The real fillers were of a different structure to the experimental items. Half of the fillers were sentences without pronouns but of a comparable length to the experimental items (they contained names and occupations, like the experimental items); there were also 4 items containing reflexives, and 14 containing pronouns of a different type to those in the experimental items.

experiment was complete, participants took a working memory test and finally they did a questionnaire experiment. The experiment session took around 1 hour in total.

## 5. Data preparation

- 23 Data preparation remains the same as reported in Patterson (2012). Individual fixations shorter than 80 ms and within 1 degree of visual angle of a neighbouring fixation were merged with the neighbouring fixation. Other fixations shorter than 80 ms, and longer than 1,000 ms, were removed. Data from three participants were removed due to excessive track loss, leaving data from 32 participants. Three individual trials were removed from the analysis due to excessive track loss, representing 0.5% of the data. Within a trial, regions that were initially skipped were removed from the analysis.

## 6. Data analysis

- 24 The experimental sentences were divided into regions for analysis. The regions of analysis were divided in the same way as reported in Patterson (2012), but additional regions are reported on here (prefinal and final), and the precritical region is examined in more detail. The regions of interest are the precritical region (one word before the critical region), the critical region (*that he* or *that she*), the spillover region (two words following the critical region), the prefinal region (two words following the spillover region) and the final region (two words following the prefinal region, always the end of the critical sentence). Five eye-tracking measures were analyzed: first-fixation durations, first-pass times, regression-path times, rereading times and total viewing times.
- 25 For the new analysis, the dataset was transformed using a logarithmic transformation ( $\log_e(x + 1)$ )<sup>6</sup>. Untransformed and transformed data was plotted against a normal distribution line and visually inspected. Where the transformed data better fitted the normal distribution, it was used in the analysis in place of the untransformed data. This was the case for all regions and measures except the rereading times (all regions). Untransformed data was used in the rereading times. The data transformation is a crucial difference in the analysis between Patterson (2012) and the new analysis presented here. Log transformation can ensure that outlying data points do not exert an undue influence on the data analysis. The data was analyzed with linear mixed effects models using the `lmer` function in the `lme4` package in R (Bates et al., 2012). This is the other crucial difference from

6.  $\log_e(x + 1)$ : 1 was added to each point in the dataset before transformation so that no datapoint would equal zero. Zero cannot be log-transformed.

the previous analysis, which used ANOVAs. There are several advantages to using mixed effects models for analyzing data from linguistic experiments (see Quené & Van den Bergh, 2004; Baayen, 2008; Baayen et al., 2008). Mixed effects models use the raw data as opposed to means, and the random factors of participants and items can be accounted for in the same analysis. They also take better account of the variance that comes about through sampling participants and items.

- 26 The fixed factors in the analysis were *antecedent* (DP or QP), *context* (match or mismatch), and an interaction (*antecedent* x *context*). The inclusion of each fixed factor in the model was tested by comparing a model containing the factor of interest to a model without the factor of interest, using a Likelihood ratio test. The first test compared the model with all fixed factors to a model with only *antecedent* and *context*, without the interaction. If the additional interaction failed to improve the model (determined by  $p < 0.10$ ), *antecedent* and *context* were tested one by one for inclusion. If neither factor on its own improved the model, the empty model was considered the best model, i.e., the factors in question did not have an effect on the results. For the random part of the model, the full random effects structure was preferred (Barr et al., 2013). However, it was not always possible to use the full random effects structure because of non-convergence. Within each region/measure, effects were removed from the random part of the model one by one, starting with the interaction, until a converging model was found (as suggested by Barr et al., 2013). Likelihood ratio tests for the fixed factors were always conducted using models that had the same random effects structure, to avoid additional differences in degrees of freedom between the models being compared. Where the interaction was found to be significant, the factor *antecedent* was relevelled and the model run again to obtain the coefficient for the effect of context at this level. P-values for the coefficients in the best model per region/measure were derived from the upper bound of the t-statistic (Baayen, 2008).

## 7. Results

### 7.1. Comprehension questions

- 27 Three questions were removed from the analysis of the comprehension question scores because they were ambiguous and received low response accuracy rates. For the remaining questions, the overall correct response rate was 84%. No individual participant scored below 70%.

### 7.2. Eye-movement data

- 28 The skipping rates in the analyzed regions were as follows: precritical, 22%; critical, 21%; spillover, 8%; prefinal, 15%; final, 11%.
- 29 The means and standard deviations for all measures in all analyzed regions are shown in Table 1, and the outcomes of the statistical analysis are shown in Table 2.



	First-fixation mean	First-pass mean	Regression-path mean	Rereading mean	Total mean
PRECITICAL REGION					
DP, context match	211 (68)	244 (124)	312 (253)	126 (251)	370 (298)
DP, context mismatch	233 (87)	262 (126)	359 (424)	113 (226)	375 (254)
QP, context match	222 (73)	240 (90)	379 (424)	124 (259)	364 (273)
QP, context mismatch	210 (62)	231 (96)	330 (315)	128 (260)	359 (273)
CRITICAL REGION					
DP, context match	199 (53)	258 (150)	358 (378)	163 (212)	421 (248)
DP, context mismatch	212 (80)	265 (148)	329 (277)	153 (279)	418 (311)
QP, context match	206 (60)	273 (124)	404 (430)	203 (362)	476 (388)
QP, context mismatch	213 (75)	250 (118)	320 (316)	188 (296)	438 (327)
SPILLOVER REGION					
DP, context match	204 (80)	292 (188)	430 (459)	243 (336)	536 (380)
DP, context mismatch	195 (58)	273 (172)	340 (270)	184 (300)	457 (324)
QP, context match	198 (71)	291 (184)	447 (535)	214 (350)	505 (434)
QP, context mismatch	204 (68)	288 (157)	436 (471)	232 (317)	520 (347)
PREFINAL REGION					
DP, context match	211 (61)	281 (144)	438 (371)	182 (289)	464 (314)
DP, context mismatch	211 (64)	273 (138)	492 (721)	173 (278)	446 (306)
QP, context match	206 (60)	285 (152)	417 (458)	140 (227)	425 (259)
QP, context mismatch	217 (72)	311 (161)	448 (340)	205 (335)	516 (377)
FINAL REGION					
DP, context match	224 (79)	312 (180)	991 (1573)	129 (220)	441 (271)
DP, context mismatch	202 (67)	310 (177)	769 (1019)	126 (280)	436 (345)
QP, context match	203 (73)	297 (165)	825 (1331)	140 (256)	437 (314)
QP, context mismatch	213 (75)	312 (199)	812 (986)	134 (237)	446 (329)

Table 1. Means for all conditions in milliseconds (with standard deviations shown in brackets) for first-fixation durations, first-pass times, regression-path times, rereading times and total viewing times in all regions from precritical to final

	First-fixation durations		First-pass times		Regression-path times		Rereading times		Total viewing times	
	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value
PRECRITICAL REGION										
Antecedent	0.05 (0.04)	1.27	0.02 (0.04)	0.55	0.14 (0.08)	1.72(*)	-	-	-	-
Context	0.10 (0.04)	2.41*	0.10 (0.05)	1.81(*)	0.12 (0.08)	1.52	-	-	-	-
Antecedent x Context	-0.14 (0.05)	-3.03**	-0.13 (0.06)	-2.43*	-0.20 (0.08)	-2.39*	-	-	-	-
Context (antecedent = QP)	-0.04	-0.98	-0.04	-0.66	-0.09	-1.13	-	-	-	-
CRITICAL REGION										
Antecedent	-	-	-	-	-	-	-	-	-	-
Context	-	-	-	-	-0.10 (0.07)	-2.17*	-	-	-	-
Antecedent x Context	-	-	-	-	-	-	-	-	-	-
SPILLOVER REGION										
Antecedent	-	-	-	-	-0.01 (0.07)	-0.07	-35.26 (34.60)	-1.02	-0.12 (0.07)	-1.70(*)
Context	-	-	-	-	-0.14 (0.07)	-1.86(*)	-64.28 (32.04)	-2.01*	-0.14 (0.06)	-2.29*

Antecedent x Context	-	-	-	0.17 (0.08)	1.95(*)	78.57 (43.94)	1.79(*)	0.24 (0.08)	2.87**
Context (antecedent = QP)				0.03	0.42	14.28	0.45	0.09	1.42
PREFINAL REGION									
Antecedent	-	-	0.01 (0.05)	0.24	-	-	-	-0.05 (0.06)	-0.86
Context	-	-	< -0.01 (0.05)	-0.10	-	-	-	-0.04 (0.06)	-0.61
Antecedent x Context	-	-	0.11 (0.06)	1.79(*)	-	-	-	0.19 (0.10)	1.79(*)
Context (antecedent = QP)			0.11	2.24*				0.15	2.22*
FINAL REGION									
Antecedent	-0.10 (0.03)	-2.82**	-	-	-	-	-	-	-
Context	-0.10 (0.03)	-2.83**	-	-	-	-	-	-	-
Antecedent x Context	0.14 (0.05)	2.90**	-	-	-	-	-	-	-
Context (antecedent = QP)	0.04	1.27							

Table 2. Outcome of the modelling process for each region and measure. The estimate for the adjustment for each factor in the model is given in the “estimate” column, and the standard error of the estimate is shown in brackets. The *t*-value for each estimate is given with the significance level indicated: (\*) =  $p < .10$ , \* =  $p < .05$ , \*\* =  $p < .001$ . The baseline for the *antecedent* was DP and the baseline for the *context* is match. Coefficient estimates and *t*-values are not given (indicated by “-”) when the outcome of the modelling process was that the empty model was the best model

30 In all models the baseline for the factor antecedent is DP and the baseline for the factor context is match. When there was an interaction between the main effects, the factor antecedent was relevelled (so that the baseline was QP) and the model run again to obtain the coefficient for the effect of context at this level. Note that when there is a significant interaction, any main effects are interpreted in relation to the interaction. The models are estimated using two categorical factors of two levels each, so that in the presence of an interaction the estimates all need to be interpreted in relation to the baseline. The positive or negative value of the estimate and corresponding  $t$ -value indicates the direction of the effect. In the presence of an interaction, a significant main effect of the factor antecedent reveals a difference between the QP and DP conditions when context = match (i.e., a difference between the two match conditions). A significant main effect of the factor context reveals a difference between the match and the mismatch conditions when antecedent = DP. A main effect of context in the relevelled model reveals a difference between match and mismatch conditions when antecedent = QP. A significant interaction itself reveals a difference between the two mismatch conditions. When there is no interaction, significant main effects can be interpreted in the usual way, e.g., a main effect of context reveals an overall difference between the match and the mismatch conditions.

31 In the precritical region, first-fixation durations, first-pass times and regression-path times show a significant interaction between antecedent and context (first-fixation durations  $\beta = -0.14$ ,  $t = -3.03$ ; first-pass times  $\beta = -0.13$ ,  $t = -2.43$ ; regression-path times  $\beta = -0.20$ ,  $t = -2.49$ ). The mismatch condition has shorter first-fixation durations, first-pass times and regression-path times when the antecedent is a QP compared to when it is a DP. In first-fixation durations (see Figure 1), there is also a significant main effect of context ( $\beta = 0.10$ ,  $t = 2.41$ ) in the presence of the interaction. This shows that for the DP antecedent, the mismatching context has significantly higher first-fixation durations than the matching context. The difference between match and mismatch in the QP conditions was non-significant. The main effect of antecedent was also non-significant. A very similar pattern is seen in the first-pass times of the precritical region (see Figure 2). There is a significant interaction between antecedent and context, and a marginally significant effect of context ( $\beta = 0.10$ ,  $t = 1.81$ ). The main effect of antecedent is non-significant as is the effect of context in the QP conditions. In regression-path times (see Figure 3), in addition to the significant interaction, there is a marginally significant effect of antecedent ( $\beta = 0.14$ ,  $t = 1.72$ ) showing that the match condition has marginally longer regression-path times in the QP compared to the DP condition. The effect of context is not significant in the DP conditions or the QP conditions. For the rereading times and the total viewing times in this region, no effects or interactions significantly improved model fit.

32 In the critical region, the analysis showed no significant effects in the early measures (first-fixation durations and first-pass times). For the regression-path

times, the best model was one in which only context was included as a fixed factor. There was a significant main effect of context ( $\beta = -0.10$ ,  $t = -2.17$ ); this shows that overall the match conditions had longer regression-path times than the mismatch conditions, as can be seen in Figure 4. This pattern of results can be seen as a numerical trend in the rereading times in the critical region but the analysis revealed no significant effects here or in the total viewing times.

33 In the spillover region, the analysis showed no significant effects in the early measures (first-fixation durations and first-pass times), although numerically the first-pass times follow the same pattern as the later measures. The total viewing times (an aggregate measure) show a significant interaction of *antecedent* and *context* ( $\beta = 0.24$ ,  $t = 2.87$ ). This is reflected in both regression-path times and rereading times as a marginal interaction (regression-path times  $\beta = 0.17$ ,  $t = 1.95$ ; rereading times  $\beta = 78.57$ ,  $t = 1.79$ )<sup>7</sup>. This shows that the QP mismatch condition has longer total viewing times (and marginally longer regression and rereading times) than the DP mismatch condition. In regression-path times (see Figure 5) there is a marginally significant main effect of *context* ( $\beta = -0.14$ ,  $t = -1.86$ ), showing that the DP match condition had marginally longer regression-path times than the DP mismatch condition. There was no effect of context in the QP conditions and no main effect of antecedent. A very similar pattern is seen in the rereading times (see Figure 6). In the presence of a marginally significant interaction between *antecedent* and *context*, there is a significant main effect of *context* ( $\beta = -64.28$ ,  $t = -2.01$ ), showing longer rereading times in the DP match condition compared to the DP mismatch condition. The main effect of *antecedent* is non-significant as is the effect of context in the QP conditions. In total viewing times (see Figure 7), in addition to the significant interaction there is a marginally significant effect of *antecedent* ( $\beta = -0.12$ ,  $t = -1.70$ ) showing that there are marginally longer total viewing times in the DP match condition compared to the QP match condition. There is a significant effect of *context* in the DP conditions ( $\beta = -0.14$ ,  $t = -2.29$ ) with longer total viewing times for DP match compared to DP mismatch. There is no effect of *context* in the QP conditions.

34 In the prefinal region, the total viewing times showed a marginally significant interaction between *antecedent* and *context* ( $\beta = 0.19$ ,  $t = 1.79$ ) (see Figure 8); this was also seen in the first-pass times ( $\beta = 0.11$ ,  $t = 1.79$ ) (see Figure 9). This shows marginally longer total viewing and first-pass times in the QP mismatch condition compared to the DP mismatch condition. There is no main effect of context in either measure in the DP conditions but there is a main effect of context in the QP conditions (first-pass times  $\beta = 0.11$ ,  $t = 2.24$ ; total viewing times  $\beta = 0.15$ ,  $t = 2.22$ ), with longer first-pass and total viewing times in the

7. Recall that the data for the rereading times was not log-transformed before analysis, unlike the data from the other measures. This is why the coefficient appears much larger for the rereading times in comparison to other measures.

mismatch condition compared to the match condition. There is also no effect of *antecedent* in either measure showing that there is no difference between the QP match and the DP match conditions. There were no significant effects or interactions in the first-fixation durations, regression-path times or rereading times, although the numerical trends are mainly in the same direction as the effects (little difference between match and mismatch for the DP conditions, but a numerically bigger difference between them in QP conditions, where mismatch produces higher reading times).

35 In the final region, it is only the first-fixation durations that show any effects (see Figure 10). Here, there is a significant interaction between *antecedent* and *context* ( $\beta = 0.14$ ,  $t = 2.90$ ), as well as main effects of *antecedent* ( $\beta = -0.10$ ,  $t = -2.82$ ) and *context* in the DP conditions ( $\beta = -0.10$ ,  $t = -2.83$ ), but not in the QP conditions. The interaction shows that the QP mismatch condition has longer first-fixation durations compared to the DP mismatch condition. In the presence of the interaction, the main effect of *antecedent* shows that first-fixation durations are longer for the DP match condition than for the QP match condition. The effect of *context* shows that the DP match condition has significantly longer first-fixation durations than the DP mismatch condition. In the QP conditions, there is no effect of context.

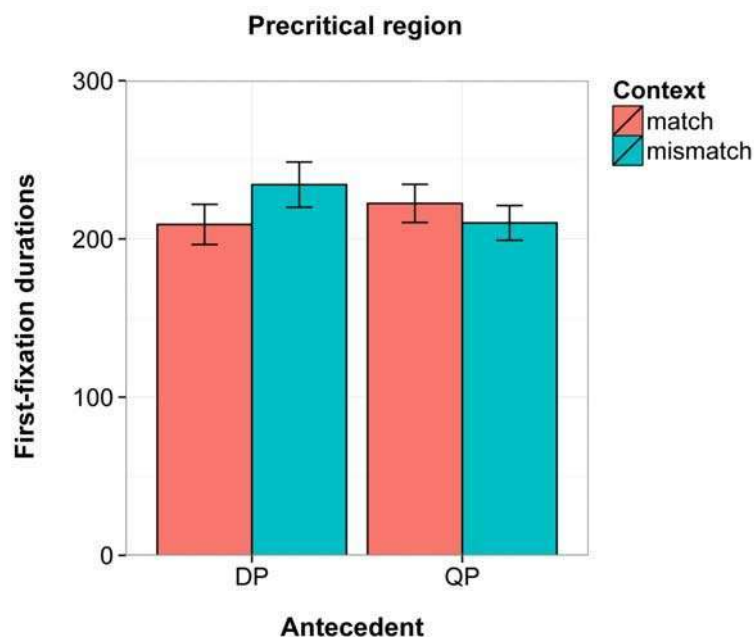


Figure 1. Mean first-fixation durations in the precritical region, per condition. Error bars are 95% confidence intervals

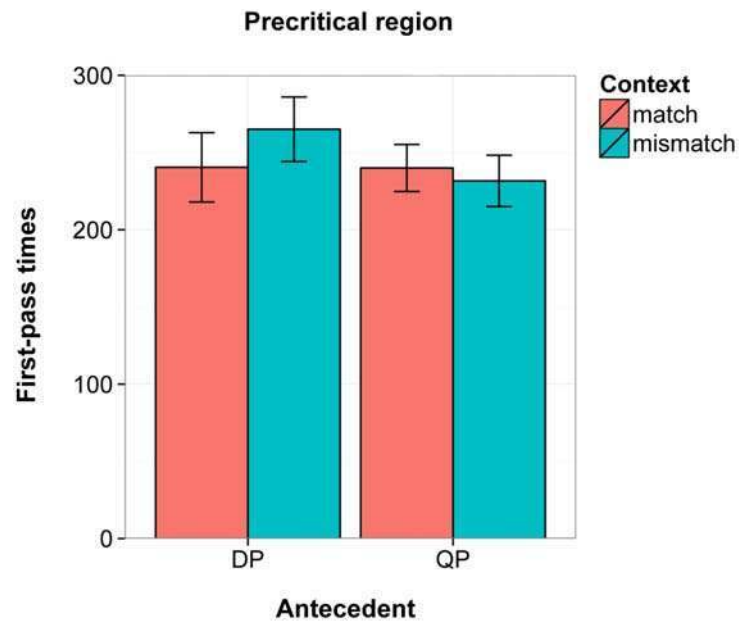


Figure 2. Mean first-pass times in the precritical region, per condition. Error bars are 95% confidence intervals

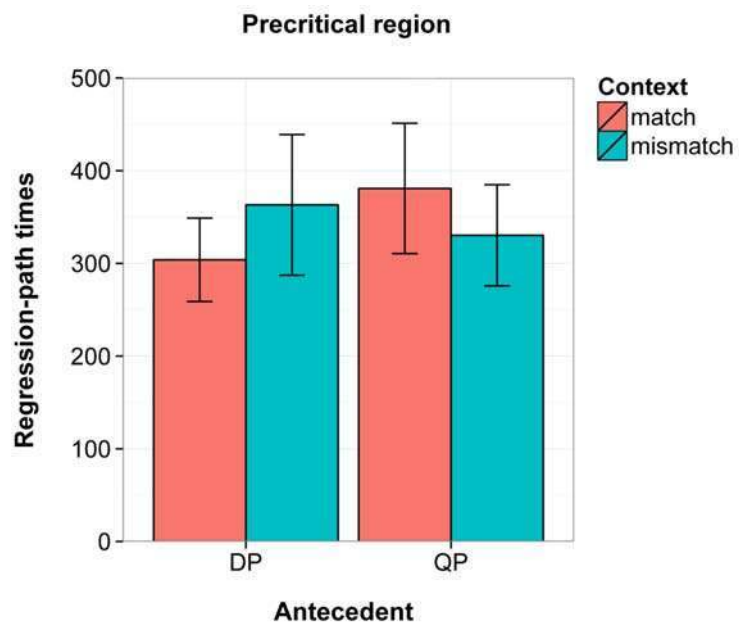


Figure 3. Mean regression-path times in the precritical region, per condition. Error bars are 95% confidence intervals

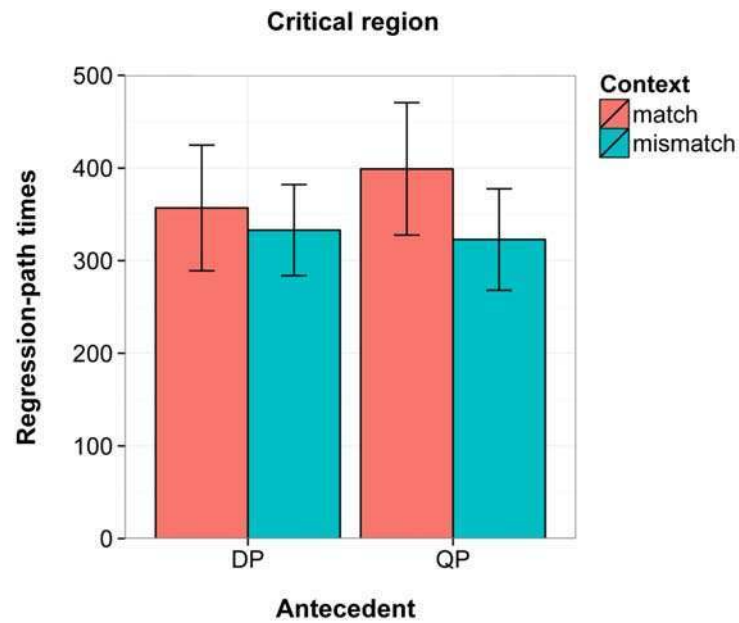


Figure 4. Mean regression-path times in the critical region, per condition. Error bars are 95% confidence intervals

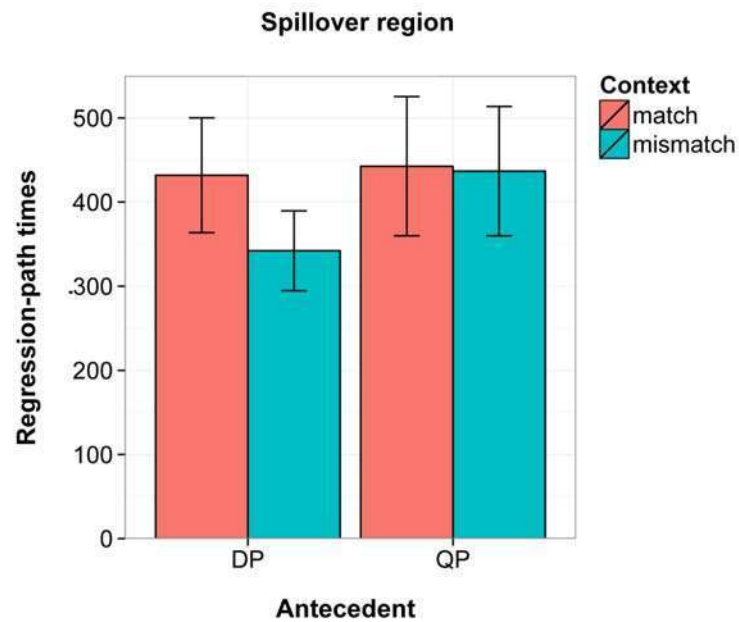


Figure 5. Mean regression-path times in the spillover region, per condition. Error bars are 95% confidence intervals



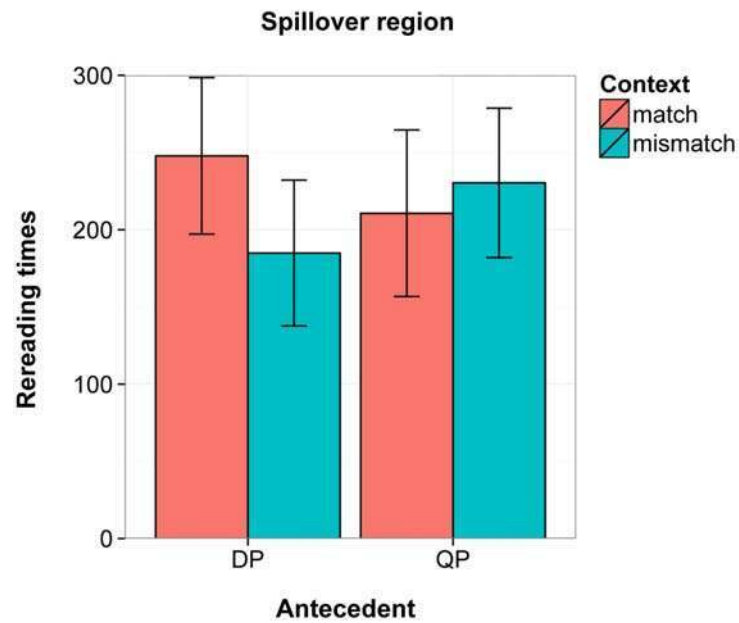


Figure 6. Mean rereading times in the spillover region, per condition. Error bars are 95% confidence intervals

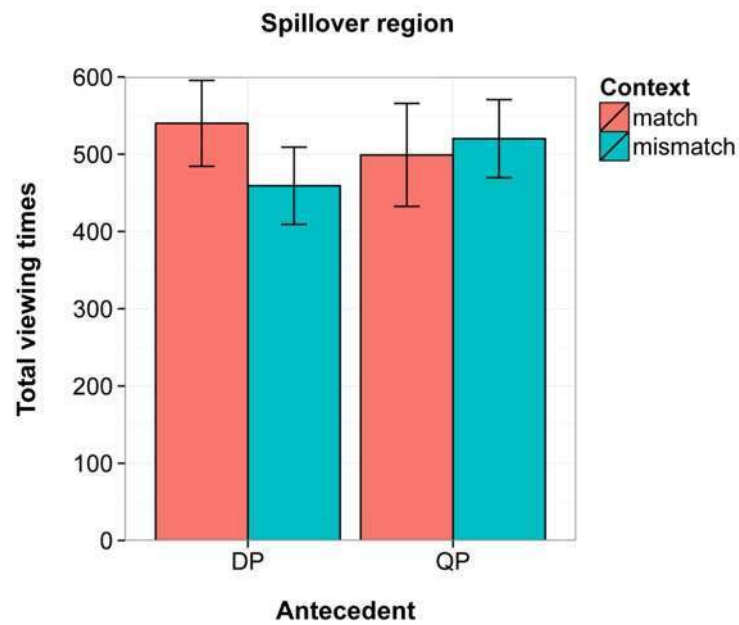


Figure 7. Mean total viewing times in the spillover region, per condition. Error bars are 95% confidence intervals

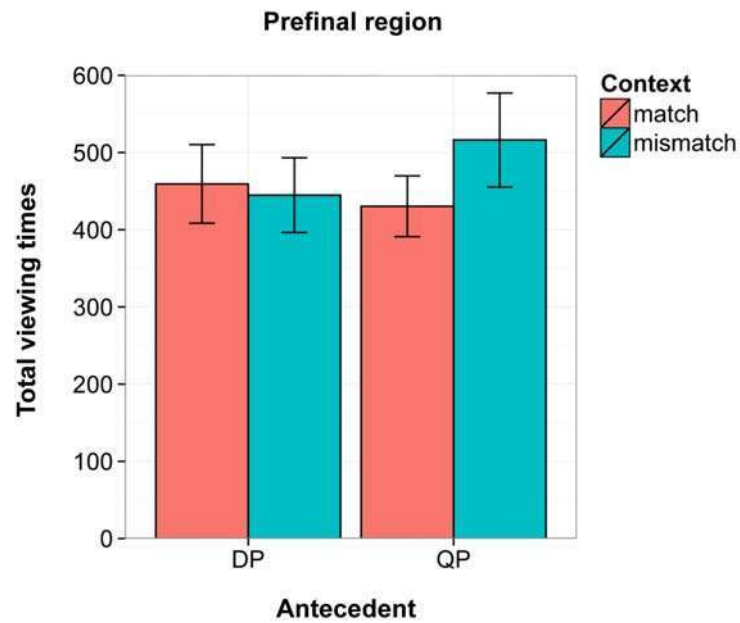


Figure 8. Mean total viewing times in the prefinal region, per condition. Error bars are 95% confidence intervals

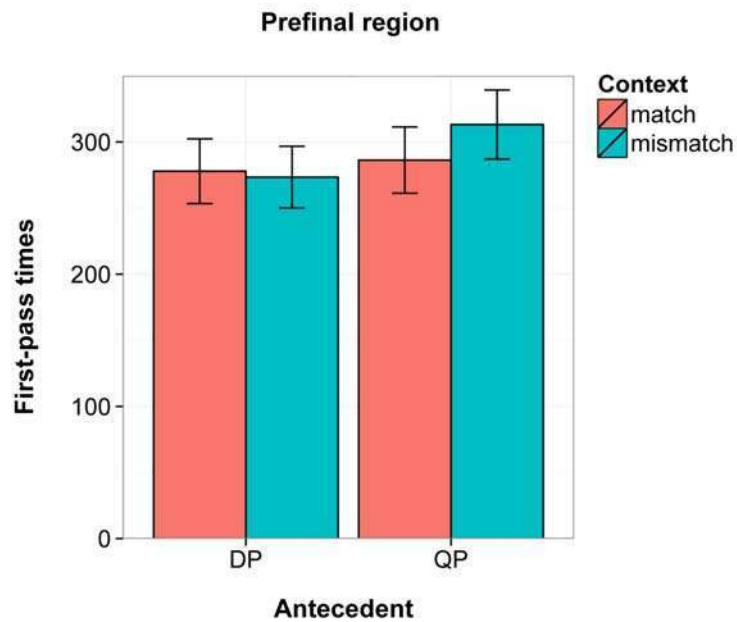


Figure 9. Mean total first-pass times in the prefinal region, per condition. Error bars are 95% confidence intervals

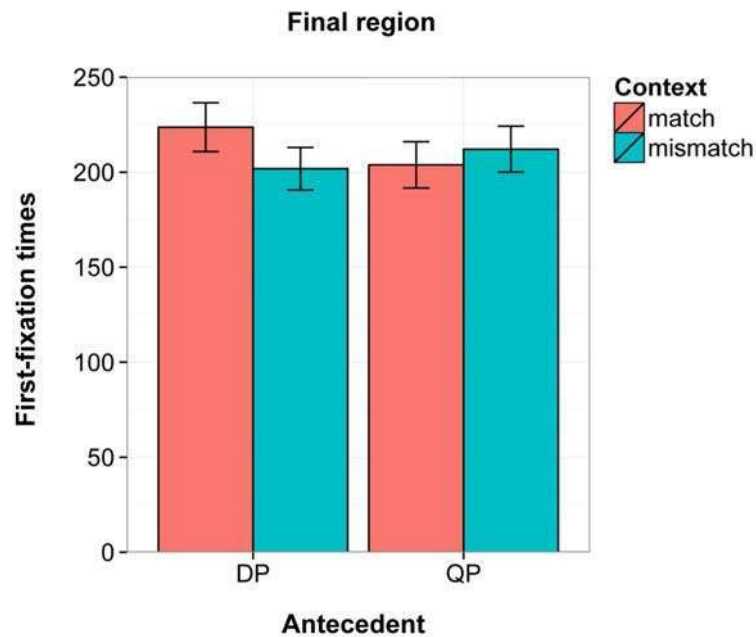


Figure 10. Mean first-fixation durations in the final region, per condition. Error bars are 95% confidence intervals

## 8. Discussion

36 For the sake of comparison, the results and conclusions from the Patterson (2012) paper are summarized here, before the results from the new analysis (given above) are discussed. Firstly there was an effect of the discourse manipulation in first-fixations in the critical region, with significantly shorter durations in the context match conditions (not found in the new analysis). In the regression-path times, there was a marginally significant effect of context in the opposite direction, that is, longer regression-path times for the context match conditions (seen also in the new analysis). In the spillover region, there was a marginal effect of context in the regression-path times (item-analysis only) with longer regression-path times in the context match conditions. This pattern was also seen in the rereading times and total viewing times. This set of results suggested that there was a very early effect of context in the expected direction, that is, matching contexts being less disruptive than mismatching contexts. This effect is fleeting, however, because it has reversed at the regression-path times in the same region, still a relatively early measure. All other results, both in the critical and the spillover regions, show longer reading times for the context match conditions. The results appeared to show, firstly, that discourse factors could influence very early stages of pronoun processing. It was suggested that the reverse effect, when the mismatching contexts had shorter reading times, could be due to repeated-name penalty effects. This was unexpected, because the

repeated-name penalty has been mainly looked for at the repeated name itself, rather than downstream of the repetition, also because repeated names are not expected to be as apparent in more complex discourses, and because it was unclear whether the repeated-name penalty could be extended to referents other than proper names. It was speculated that the likely cause of the repeated-name penalty was an upset in discourse coherence. This upset may have effects downstream of the repetition itself, and in particular could involve pronouns which have to search for or reactivate previous referents from the discourse. The discrepancy between the very early context match advantage and its later disadvantage was explained as a difference in timing for different types of discourse information. An early advantage was seen for the antecedent that had received a boost in activation from the prior context, but later during processing the overall discourse coherence was taken into account, with disrupted coherence (via the repeated NPs in the context match conditions) causing disrupted reading times. The Patterson (2012) paper therefore argued for discourse effects to be considered not as a whole, but as various different factors that can take effect at different times during pronoun processing. As far as pronoun resolution preferences with two potential subject antecedents are concerned, the statistical analysis brought out no obvious differences between the DP and the QP antecedent, despite the existence of numerically different trends for the two antecedents.

37 The advantage of the new analysis presented in this paper is that it brings out differences that were previously overlooked. Although the difference in first-fixation durations in the critical region is no longer significant, the effect in the regression-path times is seen in both analyses, and is fully significant in the new analysis. The new analysis also brings out the trends seen in the spillover region, where there was an advantage for the context mismatch condition for the DP antecedent but not for the QP. The data from the prefinal and final regions was not presented in the Patterson (2012) paper, but it is presented here and the data shows an interesting contrast. In the prefinal region, there is an advantage for the context match condition when the antecedent is a QP, but no such advantage for the DP. The effects in the final region reflect effects seen in earlier regions. In contrast to the previous analysis, in the critical region there is only a disadvantage for the context match condition, rather than an early advantage followed by a later disadvantage. Although this was not considered in Patterson (2012), one possible explanation for the reversal is that the effect at first-fixation durations was the result of participants regressing out of the critical region quickly to reread previous material in the context match conditions, whereas they fixated for longer in the mismatch conditions but then progressed to the next region. The inflated regression-path times for the context match conditions support this, so that the two results together suggest more difficulty overall with the match conditions. In any case, the first-fixation durations do not show any significant effects in the new analysis, and we have to rely on the regression-path times alone to demonstrate the disruptive effect of the match conditions. What is also seen in the new analysis is an interaction between context and antecedent, with effects of the context manipulation on the DP in one region and effects on the

QP in the next region. This pattern from the new analysis ties in with the results of the Cunnings et al. study. In Cunnings et al., where the prior context was not manipulated, there was a preference for the most recent antecedent. In the current study, the DP is always the recent antecedent so it can be expected that the DP conditions will enjoy an advantage over the QP conditions. Such an advantage can be seen indirectly in the earlier effects of context on the DP compared to the QP. This is discussed in more detail below.

38 The results of the new analysis are now summarized and discussed. In the precritical region, there was an interaction between *antecedent* and *context* in the early measures, with mismatching contexts having longer reading times when the DP was the antecedent than when the QP was the antecedent. Normally, the precritical region is checked to ensure that any effects seen at the critical region and beyond can really be attributed to the critical region (containing the pronoun), and did not start beforehand. The precritical region contained the verb for which the QP was the subject (in all four conditions). The only difference between conditions at this point was that the context sentences were different; the DP context match and the QP context mismatch contained “the queen” in the context sentence and the DP context mismatch and the QP context match began with “the soldiers”. Conditions beginning with “the queen” were easier to process at the precritical region than those beginning with “the soldiers”. If the cost of establishing a representation of the QP is affected by the content of the previous sentence, this could help to explain the pattern observed at the precritical region, since the verb must be linked specifically to the QP. A post-hoc analysis of early reading times at the QP was carried out to explore this possibility<sup>8</sup>. Tables 3 and 4 below show the means and model outcomes for three measures in this region and also in the DP region.

39 The post-hoc analysis shows that first-pass times and total viewing times of the QP were affected by the context sentence, with shorter reading times when the context sentence began with “the soldiers”. This suggests that the QP was easier to process when a set of soldiers had already been introduced into the discourse by the NP “the soldiers”. The finding that linking the verb in the precritical region to the QP was more difficult in these conditions is then perhaps counter-intuitive. But in memory literature it has been argued that more processing effort during encoding can lead to reduced effort in later reaccess, because the extra effort involved in encoding increases salience (see discussion in Van Gompel & Majid, 2004). The extra effort in encoding *every soldier* when only *the queen* has previously been introduced could have led to reduced processing times when *every soldier* was reactivated at the verb<sup>9</sup>.

8. Data treatment and modelling were the same as for the main analysis. The data was recoded to create a two-level factor called *contNP*, which indicated whether the context sentence began with “the soldiers” or “the queen”. The baseline was “soldiers”.

9. Another possibility is that, while in the precritical region, readers were able to engage in parafoveal processing of the next region which contained the pronoun thereby leading to the interaction. However this is unlikely because the critical region consisted of *that he* or *that she*, and parafoveal processing of upcoming material is limited to approximately 6 characters.

	First-fixation durations	First-pass times	Total viewing times
QP REGION			
DP, context match Context = queen	191 (82)	496 (220)	692 (381)
DP, context mismatch Context = soldiers	193 (69)	248 (203)	658 (426)
QP, context match Context = soldiers	200 (83)	437 (183)	635 (358)
QP, context mismatch Context = queen	186 (77)	483 (232)	696 (390)
DP REGION			
DP, context match Context = queen	202 (67)	408 (207)	662 (401)
DP, context mismatch Context = soldiers	210 (69)	478 (272)	792 (502)
QP, context match Context = soldiers	209 (62)	492 (342)	821 (581)
QP, context mismatch Context = queen	214 (73)	417 (230)	691 (482)

Table 3. Means for all conditions in milliseconds (with standard deviations shown in brackets) for first-fixation durations, first-pass times, and total viewing times in the QP and DP regions

	First-fixation durations		First-pass times		Total viewing times	
	Estimate (SE)	t-value	Estimate (SE)	t-value	Estimate (SE)	t-value
QP REGION						
<i>ContNP</i>	-		0.12 (0.03)	4.78**	0.09 (0.04)	2.37*
DP REGION						
<i>ContNP</i>	-		-0.13 (0.03)	-3.92**	-0.16 (0.05)	-2.87*

Table 4. Outcome of the modelling process for QP and DP regions, early measures. The estimate for the adjustment for the *contNP* factor in the model is given in the “estimate” column, and the standard error of the estimate is shown in brackets. The *t*-value is given with the significance level indicated: (\*) =  $p < .10$ , \* =  $p < .05$ , \*\* =  $p < .001$ . The baseline for *contNP* was “soldiers”. Coefficient estimates and *t*-values are not given (indicated by “-”) when the outcome of the modelling process was that the empty model was the best model

40 The pattern in the critical region for the DP conditions is different from the precritical region, suggesting that the effects here are based on encountering the pronoun. There was an effect of context in the regression-path times; readers took longer to go past the critical region when contexts were matching compared to when they were mismatching, which implies an elevated processing cost for the matching contexts. This is the opposite of what was predicted; the context match conditions were expected to boost the prominence of the antecedent, making it easier to process. The pattern for the QP conditions has not changed from the precritical region. If we take the precritical region into account here as well, what could be driving the main effect of context in the critical region is a change in pattern for the DP and no change in pattern for the QP. This is consistent with the fact that the context appears to affect the QP at a later point than the DP, as we see in the spillover and prefinal regions. In the spillover region, there is a significant or marginally significant interaction between antecedent and context in regression-path times, rereading times and total viewing times. Looking just at the DP antecedent, the matching condition takes longer to process than the mismatching condition (marginally so in regression-path times), which is also the pattern observed in the critical region. In the prefinal region, it is the QP, not the DP, that is affected by the context manipulation. In first-pass and total viewing times there is a marginal interaction between antecedent and context. There is a significant effect of context when the antecedent is a QP but not when the antecedent is a DP. The QP in this region is affected differently by the context manipulation than the DP was in the region before. Here, there is an advantage for the matching context rather than a processing cost. The DP in this region is unaffected by the manipulation of context.

41 The effects of the context manipulation are seen reasonably early in the pronoun region for the DP and continue into the spillover region. The matching context causes disruption for the DP antecedent. Effects of context on the QP antecedent emerge in the prefinal region, where matching contexts are more helpful than the mismatching contexts. The effect of the context manipulation, then, is not uniform, contrary to the predictions. In CT terms, it was expected that Cb status would boost the ranking of an entity in the Cf of U<sub>2</sub>, leading to easier processing for the pronoun in U<sub>3</sub>. In non-CT terms, establishing a topic in the first sentence should lead to easier processing when the pronoun matched the topic, because no intervening entity provided a competing sentence topic. This was not the case. It was also not the case that the recency preference observed in the Cummings et al. study overruled any effects of the context sentence. The recency preference however did emerge as a later effect of context on the QP than on the DP. This is likely to reflect some extra processing involved in recovering the QP antecedent because, not being the most recent antecedent, it was dispreferred.

42 The context did not overrule the recency preference, nor did the recency preference overrule the context. Rather, the context had a different effect on the two antecedents. How can this differential pattern be explained? One potential explanation, also discussed in the Patterson (2012) paper, is an extended repeated

name penalty. When “the queen” is introduced in the context sentence, repeated mention of it in the critical sentence, as opposed to use of a pronoun, could have incurred a repeated name penalty (as suggested by Gordon et al., 1993). The penalty is also found when the pronoun refers to the DP, causing processing difficulty in the DP/context match condition. But this explanation is contradicted by the findings of the post-hoc analysis (Tables 3 and 4 above). First-pass reading times in the DP region were shorter when the context sentence contained “the queen” compared to when it contained “the soldiers”, making it very unlikely that repeated mention had a detrimental effect on pronoun resolution.

43 It is perhaps more helpful to consider the effect that the context sentence has on the coherence of text. Coherence in CT is in part defined by the transitions between utterances, as exemplified by the ranking of CONTINUE, RETAIN and SHIFT. The principle underlying this idea of coherence is that “frequent shifting leads to a lack of local coherence” (Grosz et al., 1995: 215). However, the transitions are implemented in CT in such a way that no transition takes place between the first and second utterance. This is because shifting is concerned with maintenance of the Cb, which can only be established at the second utterance. As such, it is impossible to explain the coherence between the context sentence and the critical sentence in CT terms. But the notion of frequent shifting disrupting coherence can usefully be applied to explain the results of the current experiment. When the context sentence began with “the soldiers”, it established a set of soldiers in the discourse representation, making it easier to process the subsequent QP *every soldier*, as evidenced in the post-hoc analysis. The attention of the reader is still on the set of soldiers until *the queen* is encountered in the critical sentence. Compare this to the context sentence beginning with “the queen”. In the critical sentence, encountering the QP means establishing some kind of representation for a set of soldiers, thus switching the reader’s attention to this set. Then the DP *the queen* is encountered again (in subject position) which switches the attention of the reader again. When the context sentences begins with “the queen”, there are more switches of attention between the soldiers and the queen than there are in the context sentences beginning with “the soldiers”. The more frequent shifting between the entities leads to a lack of local coherence, and makes pronoun processing more costly as a result. Since pronouns are in some ways markers of discourse coherence, it makes sense that a more coherent discourse makes the establishment of pronoun reference easier. This fits well with the results of the current experiment, where the context sentence has a different effect on the DP antecedent conditions than it does on the QP antecedent conditions.

44 The experiment set out to discover whether an entity could receive facilitation from its previous discourse status, that is, mention in the context sentence. The previous mention clearly has an effect on linking the pronoun to the appropriate antecedent, but not in the straightforward way that was expected. The previous context was not always facilitative, and this could be due to its inducing an additional switch in the hearer’s attention, which upset the coherence of the



discourse. This has an effect on pronoun resolution, a process which relies on finding a prominent entity in the prior discourse and in some way acts as a marker of discourse coherence. An attempt was made to predict the effects of prior discourse using CT. This was not without problems, however, because of the complex nature of QP representations and the complexity of the clausal structure. Since it has been shown that QP antecedents are subject to at least some of the same salience factors as DP antecedents, further efforts should be made to understand how CT can apply in these cases. Further research is also needed to determine how to accommodate intervening relative clauses into CT utterances, and this paper suggests one way in which this could be achieved. Furthermore, CT contains some extremely useful vocabulary for understanding coherence relations. Whether or not the framework can be extended to include transitions between the first and second utterance, or indeed switches of attention within one utterance, remains a matter for future debate. Insofar as attention switches appear to play an important role in the coherence of short but complex texts such as those in the current experiment, they should be accounted for in models of coherence and pronoun resolution.

### Acknowledgements

45 The research reported here was supported by an Economics and Social Research Council postgraduate studentship to the author (awarded by the Department of Language and Linguistics, University of Essex) and by an Alexander von Humboldt Professorship to Professor Harald Clahsen (Potsdam Research Institute for Multilingualism); both are gratefully acknowledged.

### References

- ARNOLD, J.E. 2001. The Effect of Thematic Roles on Pronoun Use and Frequency of Reference Continuation. *Discourse Processes* 31 (2): 137-162.
- ARNOLD, J.E., BROWN-SCHMIDT, S. & TRUESWELL, J.C. 2007. Children's Use of Gender and Order-of-Mention during Pronoun Comprehension. *Language and Cognitive Processes* 22 (4): 527-565.
- BAAYEN, H. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.
- BAAYEN, R.H., DAVIDSON, D.J. & BATES, D.M. 2008. Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items. *Journal of Memory and Language* 59 (4): 390-412.
- BARR, D.J., LEVY, R., SCHEEPERS, C. & TILY, H.J. 2013. Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal. *Journal of Memory and Language* 68 (3): 255-278.
- BATES, D.M., MAËCHLER, M. & BOLKER, B. 2012. "lme4" package for R, version 0.999999-0, published 23.06.2012.
- BURKHARDT, P. 2005. *The Syntax-Discourse Interface*. Amsterdam – Philadelphia: J. Benjamins.

- CARMINATI, M., FRAZIER, L. & RAYNER, K. 2002. Bound Variables and C-command. *Journal of Semantics* 19 (1): 1-34.
- CLARK, H.H. & SENGUL, C.J. 1979. In Search of Referents for Nouns and Pronouns. *Memory and Cognition* 7 (1): 35-41.
- COWLES, H.W., WALENSKI, M. & KLUENDER, R. 2007. Linguistic and Cognitive Prominence in Anaphor Resolution: Topic, Contrastive Focus and Pronouns. *Topoi* 26 (1): 3-18.
- CRAWLEY, R.A., STEVENSON, R.J. & KLEINMAN, D. 1990. The Use of Heuristic Strategies in the Interpretation of Pronouns. *Journal of Psycholinguistic Research* 19 (4): 245-264.
- CUNNINGS, I., PATTERSON, C. & FELSER, C. submitted. Variable-Binding and Coreference in Sentence Comprehension: Evidence from Eye Movements.
- FUKUMURA, K., VAN GOMPEL, R.P.G. 2010. Choosing Anaphoric Expressions: Do People Take into Account Likelihood of Reference? *Journal of Memory and Language* 62 (1): 52-66.
- GARNHAM, A. 2001. *Mental Models and the Interpretation of Anaphora*. Hove: Psychology Press.
- GARNHAM, A. & OAKHILL, J. 1992. Discourse Processing and Text Representation from a "Mental Models" Perspective. *Language and Cognitive Processes* 7 (3-4): 193-204.
- GELORMINI-LEZAMA, C. & ALMOR, A. 2011. Repeated Names, Overt Pronouns, and Null Pronouns in Spanish. *Language and Cognitive Processes* 26 (3): 437-454.
- GERNSBACHER, M.A. & HARGREAVES, D.J. 1988. Accessing Sentence Participants: The Advantage of First Mention. *Journal of Memory and Language* 27 (6): 699-717.
- GORDON, P.C., GROSZ, B.J. & GILLIOM, L.A. 1993. Pronouns, Names, and the Centering of Attention in Discourse. *Cognitive Science* 17 (3): 311-347.
- GORDON, P.C. & HENDRICK, R. 1998. The Representation and Processing of Coreference in Discourse. *Cognitive Science* 22 (4): 389-424.
- GORDON, P.C. & SCEARCE, K.A. 1995. Pronominalization and Discourse Coherence, Discourse Structure and Pronoun Interpretation. *Memory and Cognition* 23 (3): 313-323.
- GROSZ, B.J., JOSHI, A.K. & WEINSTEIN, S. 1995. Centering: A Framework for Modelling the Local Coherence of Discourse. *Computational Linguistics* 21 (2): 203-225.
- HINTERWIMMER, S. & ENDRIS, C. 2009. Indefinites as Direct and Indirect Aboutness Topics. In M. ZIMMERMAN & C. FERY (eds.), *Information structure*. Oxford: Oxford University Press: 89-115.
- JÄRVIKIVI, J., VAN GOMPEL, R.P.G., HYÖNÄ, J. & BERTRAM, R. 2005. Ambiguous Pronoun Resolution: Contrasting the First-Mention and Subject-Preference Accounts. *Psychological Science* 16 (4): 260-264.
- JOHNSON-LAIRD, P.N. 1983. *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness*. Cambridge: Cambridge University Press.
- KAISER, E. 2011. Focusing on Pronouns: Consequences of Subjecthood, Pronominalisation, and Contrastive Focus. *Language and Cognitive Processes* 26 (10): 1625-1666.
- KAISER, E. & TRUESWELL, J.C. 2008. Interpreting Pronouns and Demonstratives in Finnish: Evidence for a Form-Specific Approach to Reference Resolution. *Language and Cognitive Processes* 23 (5): 709-748.

- KAMEYAMA, M. 1998. Intracentential Centering: A Case Study. In M.A. WALKER, A.K. JOSHI & E.F. PRINCE (eds.), *Centering Theory in Discourse*. Oxford: Clarendon Press: 89-112.
- KENNISON, S. & TROFE, J. 2003. Comprehending Pronouns: A Role for Word-Specific Gender Stereotype Information. *Journal of Psycholinguistic Research* 32 (3): 355-378.
- PATTERSON, C. 2012. The Effect of Local Discourse Coherence on Pronoun Resolution: An Eye-Tracking Study. *Essex Graduate Student Papers in Language and Linguistics* 13: 96-119. Available online: [http://www.essex.ac.uk/linguistics/publications/egspll/volume\\_13/PDF/Patterson.pdf](http://www.essex.ac.uk/linguistics/publications/egspll/volume_13/PDF/Patterson.pdf).
- POESIO, M., STEVENSON, R., DI EUGENIO, B. & HITZEMAN, J. 2004. Centering: A Parametric Theory and Its Instantiations. *Computational Linguistics* 30 (3): 309-363.
- QUENÉ, H. & VAN DEN BERGH, H. 2004. On Multi-Level Modeling of Data from Repeated Measures Designs: A Tutorial. *Speech Communication* 43 (1-2): 103-121.
- ROSE, R. 2007. Pronoun Resolution and the Influence of Syntactic and Semantic Information on Discourse Prominence. In A. BRANCO (ed.), *Anaphora: Analysis, Algorithms and Applications – 6th Discourse Anaphora and Anaphor Resolution Colloquium, DAARC 2007, Lagos, Portugal, March 29-30, 2007, Selected Papers*. Berlin – London: Springer: 28-43.
- SCHWABE, K. & WINKLER, S. 2007. On Information Structure, Meaning and Form: Generalizations across Languages. In K. SCHWABE & S. WINKLER (eds.), *On Information Structure, Meaning and Form: Generalizations across Languages*. Amsterdam: J. Benjamins: 1-29.
- STREB, J., HENNIGHAUSEN, E. & RÖSLER, F. 2004. Different Anaphoric Expressions are Investigated by Event-Related Brain Potentials. *Journal of Psycholinguistic Research* 33 (3): 175-201.
- STURT, P. 2003. The Time-Course of the Application of Binding Constraints in Reference Resolution. *Journal of Memory and Language* 48 (3): 542-562.
- VAN GOMPEL, R.P.G. & MAJID, A. 2004. Antecedent Frequency Effects during the Processing of Pronouns. *Cognition* 90 (3): 255-264.
- VAN RIJ, J. 2012. *Pronoun Processing: Computational, Behavioural and Psychophysiological Studies in Children and Adults*. PhD thesis. University of Groningen. Available online: [http://www.sfs.uni-tuebingen.de/~jvanrij/assets/Dissertation\\_VanRij\\_2012.pdf](http://www.sfs.uni-tuebingen.de/~jvanrij/assets/Dissertation_VanRij_2012.pdf).
- WUBS, E., HENDRIKS, P., HOEKS, J. & KOSTER, C. 2009. Tell Me a Story! Children's Capacity for Topic Shift. In J. CRAWFORD, K. OTAKI & M. TAKAHASHI (eds.), *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA 2008)*. Somerville: Cascadia Proceedings Project: 313-324. Available online: <http://www.lingref.com/cpp/galana/3/paper2331.pdf>.