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

29 | 2022
Varia

Discourse Prominence and Antecedent Mis-Retrieval during Native and Non-Native Pronoun Resolution

Cecilia Puebla and Claudia Felser

Electronic version
URL: https://journals.openedition.org/discours/11720
DOI: 10.4000/discours.11720
ISSN: 1963-1723

Publisher:
Laboratoire LATTICE, Presses universitaires de Caen

Electronic reference
Cecilia Puebla and Claudia Felser, "Discourse Prominence and Antecedent Mis-Retrieval during Native and Non-Native Pronoun Resolution", *Discours* [Online], 29 | 2022, Online since 03 May 2022, connection on 07 June 2022. URL: http://journals.openedition.org/discours/11720 ; DOI: https://doi.org/10.4000/discours.11720

discours
Revue de linguistique, psycholinguistique et informatique
http://journals.openedition.org/discours/

Discourse Prominence and Antecedent Mis-Retrieval during Native and Non-Native Pronoun Resolution

Cecilia Puebla
Potsdam Research Institute for Multilingualism,
University of Potsdam, Germany

Claudia Felser
Potsdam Research Institute for Multilingualism,
University of Potsdam, Germany

Cecilia Puebla, Claudia Felser, « Discourse Prominence and Antecedent Mis-Retrieval during Native and Non-Native Pronoun Resolution », Discours [En ligne], 29 | 2022, mis en ligne le 03 mai 2022.

URL : http://journals.openedition.org/discours/11720

Titre du numéro : *Varia*
Coordination : Saveria Colonna & Sarah Schimke

Date de réception de l'article : 05/05/2021
Date d'acceptation de l'article : 03/12/2021

Discourse Prominence and Antecedent Mis-Retrieval during Native and Non-Native Pronoun Resolution

Cecilia Puebla

Potsdam Research Institute for Multilingualism,
University of Potsdam, Germany

Claudia Felser

Potsdam Research Institute for Multilingualism,
University of Potsdam, Germany

Previous studies on non-native (L2) anaphor resolution suggest that L2 comprehenders are guided more strongly by discourse-level cues compared to native (L1) comprehenders. Here we examine whether and how a grammatically inappropriate antecedent's discourse status affects the likelihood of it being considered during L1 and L2 pronoun resolution. We used an interference paradigm to examine how the extrasentential discourse impacts the resolution of German object pronouns. In an eye-tracking-during-reading experiment we examined whether an elaborated local antecedent ruled out by binding Condition B would be mis-retrieved during pronoun resolution, and whether initially introducing this antecedent as the discourse topic would affect the chances of it being mis-retrieved. While both participant groups rejected the inappropriate antecedent in an offline questionnaire irrespective of its discourse prominence, their real-time processing patterns differed. L1 speakers initially mis-retrieved the inappropriate antecedent regardless of its contextual prominence. L1 Russian/L2 German speakers, in contrast, were affected by the antecedent's discourse status, considering it only when it was discourse-new but not when it had previously been introduced as the discourse topic. Our findings show that L2 comprehenders are highly sensitive to discourse dynamics such as topic shifts, supporting the claim that discourse-level cues are more strongly weighted during L2 compared to L1 processing.

Keywords: pronoun resolution, non-native sentence processing, discourse prominence, interference, German, eye-movement monitoring

The present research was supported by the German Research Foundation (DFG) through Grant No. FE 1138/1-1 awarded to Claudia Felser, and by an Alexander-von-Humboldt Professorship to Harald Clahsen (Potsdam Research Institute for Multilingualism). The authors would like to thank Clare Patterson for her guidance and support over different stages of this work. We thank our colleagues and members of the Potsdam Research Institute for Multilingualism for valuable discussion and constructive feedback in group lab meetings, and the audiences at the ACLC-Workshop "Doing Experiments with Theoretical Linguistics" and at the International Conference for Prominence in Language, for their insightful comments. We are grateful to two anonymous reviewers for their helpful suggestions for improvement on an earlier version of this paper. Special thanks are given to all the volunteers who participated in this study.

1. Introduction

1 Successful native (L1) and non-native (L2) language comprehension crucially relies on the ability to establish links between anaphoric expressions and their antecedents. This requires extracting from the input and integrating grammatical, semantic, and discourse-level information quickly and efficiently during processing. Current theories of sentence processing commonly assume that coming across a pronoun during comprehension triggers a cue-based search for a suitable antecedent, and the reactivation or retrieval from memory of a set of reference representations previously encoded in a comprehender's discourse model (e.g., Foraker, McElree, 2007; Lewis, Vasishth, 2005; Lewis et al., 2006). From the initial set of candidates, one is ultimately selected and integrated with the pronoun for dependency building in a subsequent resolution stage (Garrod, Terras, 2000; Sanford, Garrod, 1989). The accessibility of potential antecedents is modulated by several types of information, including structure-sensitive restrictions (i.e., binding constraints), morphological constraints such as gender or number match, discourse-level and contextual, "prominence-lending" cues that contribute to promoting one candidate over another (e.g., topichood), and the degree of referents' elaboration, that is, their syntactic complexity or semantic "heaviness" (Arnold, 2010; Troyer et al., 2016; Von Heusinger, Schumacher, 2019). These types of information interact during the time-course of processing in guiding comprehenders' referential decisions. However, inappropriate weighting of information sources or delayed application of a relevant constraint can result in temporary interference or "mis-retrieval" of feature-matching but non-target competitor candidates. Interference effects can provide evidence about which types of information sources are exploited and how they interact during dependency building. Different patterns of interference between L1 and L2 comprehenders are thus potentially informative about how, when, and to what extent these two populations make use of the different information sources that guide reference resolution.

2 The precise mechanisms modulating antecedent accessibility and the factors leading to interference during L1 vs. L2 processing are not yet fully understood. Highly proficient L2 learners can perform in a native-like manner in offline tasks, but they sometimes behave differently from native speakers during the real-time processing of pronominal expressions. Although feature-matching non-target competitor antecedents are also mis-retrieved during L1 processing under certain conditions (e.g., Chow et al., 2014; Puebla et al., 2021), L2 anaphor resolution appears especially vulnerable to interference from discourse-prominent antecedents, as several processing studies have found L2 learners to favour first-mentioned, topical antecedents for both reflexive and non-reflexive pronouns (Felser, Cunnings, 2012; Felser et al., 2009; Patterson et al, 2014; Puebla et al., 2021). Roberts et al. (2008) reported eye-tracking evidence indicating that the presence of more than one feature-matching antecedent candidate in the preceding context distracted only L2 but not L1 speakers. L2 learners also appear more likely than L1 speakers to link ambiguous pronouns to coreference antecedents than to c-commanding variable binders during real-time processing (Trompelt, Felser, 2014).

3 In an eye-tracking study on reflexive resolution, Felser and Cunnings (2012) found that introducing an antecedent in the prior context sentence increased the likelihood of it being initially retrieved during L2 but not L1 processing, irrespectively of the antecedent's structural accessibility, which suggests that L2 speakers first tried to establish the dependency at the discourse-representational level rather than by applying binding constraints. Theoretical approaches to L2 processing such as the shallow structure hypothesis (Clahsen, Felser, 2006, 2018) and the memory-interference account (Cunnings, 2017a and b) attribute these and other L1/L2 processing differences to L2 learners' overreliance on non-structural (e.g., semantic or discourse-level) information during comprehension.

4 The current study investigates how discourse-level information provided by the extrasentential context may affect the resolution of German object pronouns in grammatically constrained ("Condition B") configurations in L1 and L2 comprehension. In an eye-tracking during reading experiment, we examined whether introducing a grammatically inappropriate antecedent as the initial discourse topic affects the likelihood of it being considered for referential dependency formation in the subsequent discourse. The inappropriate antecedent was elaborated by means of relative clause (RC) modification. We also examined participants' ultimate antecedent choices in a complementary offline task.

1.1. Factors modulating antecedent accessibility

5 Sentence-internal referential dependencies are restricted by binding constraints, which determine the structural accessibility of potential antecedents. Binding constraints are syntactically mediated in that they are sensitive to configurational properties and rely on notions such as c-command and locality. C-command is a relational concept that is characterised in terms of hierarchical dominance relationships between syntactic constituents: a constituent c-commands its sister constituents and any constituents that these dominate (Reinhart, 1983). Although the precise implementation of binding constraints in cue-based models of anaphor resolution remains controversial (see e.g., Alcocer, Phillips, 2012; Kush, 2013, for discussion), they play an important role in antecedent search and retrieval, and much research has investigated how these restrictions are used in the resolution of anaphoric dependencies by native speakers (see Nicol, Swinney, 2003; Sturt, 2013, for reviews) and, to a lesser extent, by L2 learners (see Felser, 2016, 2019, for reviews).

6 The interpretation of non-reflexive pronouns (e.g., *her*, *him*) is constrained by Condition B of the binding theory (Chomsky, 1981). Condition B prohibits a pronoun from establishing a dependency with a c-commanding constituent if the two share the same local syntactic domain, as illustrated in [1]. Here the c-commanding local noun phrase (NP) *Christian* is a grammatically inappropriate (or "inaccessible") antecedent for the object pronoun *him* because both elements are coarguments of the same predicate and the NP c-commands the pronoun.

[1] Sebastian$_i$ noticed that Christian$_j$ was talking to him$_{i,*j}$

7 Condition B identifies ungrammatical antecedent candidates but does not direct the comprehender towards a suitable antecedent. Pronominal dependencies can also be established at the discourse-representational level without the mediation of c-command, via coreference assignment (Reinhart, 1983; Reuland, 2001, 2011). For example, in [1], the pronoun *him* can be linked to the c-commanding matrix subject *Sebastian*, but it can also corefer with a (non-commanding) sentence-external antecedent. Thus, the successful interpretation of personal pronouns relies on additional, non-structural factors that modulate the relative accessibility of potential referents in a coherent discourse.

8 Given that the distance between a pronoun and its antecedent may span several words or even sentences, pronouns tend to trigger the retrieval of psychologically salient entities that are more accessible in memory (e.g., Arnold, 2010; Foraker, McElree, 2007). The relative accessibility of referent candidates can be modulated at the discourse-level through mechanisms such as topichood that render referent representations highly prominent in the local discourse. Topical referents are more "active" in the discourse model and are more likely to be subsequently referred to by pronouns compared to non-topical antecedents (e.g., Ariel, 1990; Colonna et al., 2012; Gundel et al., 1993). Exactly what determines a topic's relative prominence is still under debate, however (compare, e.g., Cowles et al., 2007; Givón, 1983; Grosz et al., 1995; Gundel et al., 1993). Nevertheless, it is commonly agreed that topics are what the discourse is about (Arnold, 2010; Colonna et al., 2012; Cowles et al., 2007; Stede, 2012). They are typically introduced as first-mentioned characters in the discourse and may be more frequently mentioned or referred to throughout the discourse than non-topical entities (Kieras, 1980; Perfetti, Goldman, 1975).

9 Pronouns tend to trigger the retrieval of the current discourse topic (e.g., Arnold, 2010; Cowles et al., 2007). Although existing models of cue-based retrieval remain underspecified as to the precise implementation of discourse-level constraints such as topichood as retrieval cues (e.g., Jacob et al., 2017), the interference effects from non-target but discourse-prominent topical antecedents reported in earlier processing studies on anaphor resolution show that topichood, however implemented, can be used as a cue to guide retrieval and affects the extent to which inaccessible antecedents can be considered both during L1 (e.g., Badecker, Straub, 2002; Sturt, 2003) and L2 processing (e.g., Felser, Cunnings, 2012; Felser et al., 2009).

10 Another mechanism that modulates the accessibility of potential antecedents in memory is elaboration, which has been claimed to increase cognitive salience and facilitate retrieval (Hofmeister, 2011; Hofmeister, Vasishth, 2014; Troyer et al., 2016). Elaborated referents that exhibit a higher degree of syntactic complexity or semantic heaviness, for example through adjectivisation (e.g., *the victorious four-star general*), may have stronger memory representations and can therefore be more easily accessed than less elaborated, non-modified antecedent candidates (e.g., *the general*). Under a cue-based approach to sentence comprehension, elaboration is not a retrieval cue (i.e., it is not a cue carried by the anaphor but a property of a potential antecedent), but earlier findings have shown that it can impact retrieval,

at least during L1 processing (e.g., Chow et al., 2014; Puebla et al., 2021). It is not yet clear, however, how elaborated but binding-inaccessible antecedents may affect L2 pronoun resolution, or how antecedent elaboration may interact with "prominence-lending" cues such as topichood during L2 compared to L1 processing.

11 The present study examines how a grammatically inappropriate antecedent's relative discourse prominence modulates the likelihood of this antecedent being considered for dependency building during L1 and L2 pronoun resolution.

1.2. Elaboration and topichood in L2 vs. L1 pronoun resolution

12 Elaboration and topichood have previously been found to cause mis-retrieval of Condition B incompatible antecedents, albeit differently for L1 and L2 speakers. While there is evidence that native speakers can be distracted by elaborated non-target antecedents, L2 learners appear more sensitive to the discourse status of potential referents. Using an interference paradigm in two eye-tracking during reading experiments, Patterson et al. (2014) examined the L1 and L2 processing of English object pronouns both in canonical Condition B configurations such as [2a–c], where the embedded subject (e.g., *Mark*) is a structurally inaccessible antecedent for the pronoun *him*, and in environments containing ambiguous "short distance" pronouns such as [3a–c], where local coreference with the embedded subject (e.g., *Gavin*) is also allowed. The authors manipulated the gender congruence between the pronoun and each antecedent as a way of detecting whether an antecedent was considered for dependency formation. Assuming that evaluating an antecedent makes readers sensitive to its gender features, a reading-time difference is expected between the gender-matching and the mismatching conditions at the pronoun or shortly after (e.g., Sturt, 2003).

[2a] John remembered that Mark had taught him a new song on the guitar.

[2b] John remembered that Jane had taught him a new song on the guitar.

[2c] Jane remembered that John had taught him a new song on the guitar.

[3a] Barry saw Gavin place a gun near him on the ground with great care.

[3b] Barry saw Megan place a gun near him on the ground with great care.

[3c] Megan saw Barry place a gun near him on the ground with great care.

13 The authors found that both English native speakers and L1 German-speaking, advanced L2 learners of English behaved similarly in configurations such as [2a–c] in that they only considered the matrix subject antecedent, as reflected by longer reading times at the pronoun region and shortly after for sentences such as [2c], where the structurally accessible antecedent mismatched in gender with the pronoun (e.g., *Jane*), compared to [2a, b], where there was a gender-matching accessible antecedent (e.g., *John*). There was no difference in reading times between conditions [2a] and [2b], where the gender of the inaccessible antecedent (e.g., *Mark* vs. *Jane*) was

manipulated. This indicates that both participant groups retrieved only the accessible antecedent upon reading the pronoun, in accordance with Condition B. However, L1/L2 differences were observed for items such as [3a–c] in that only the native group considered both local and non-local antecedents. The L2 group, by contrast, again considered only the matrix subject (e.g., *Barry*) despite showing awareness of the referential ambiguity of "short distance" pronouns in a complementary offline task. Considering this finding, what might look like the efficient application of Condition B in Patterson et al.'s study may also plausibly reflect a general tendency in L2 pronoun resolution to focus on discourse-prominent antecedents.

14 In a recent study, Puebla et al. (2021) investigated the L1 and L2 processing of German object pronouns using a similar interference paradigm to examine whether, and when during processing, Condition B may be violated. The authors recorded the eye-movements of German L1 speakers and L1 Russian-speaking, proficient L2 learners of German while they read sentences such as [4], where the object pronoun (*ihn* 'him') could be linked to the non-local matrix subject (*Florian*) but not to the embedded subject (e.g., *der Kollege* 'the colleague'). The latter was rendered more elaborated compared to the local antecedents in Patterson et al.'s (2014) stimulus materials through modification (e.g., *aus Frankreich* 'from France'). Following Patterson et al. (2014), a gender-mismatch paradigm was used, with the addition of a fourth (double-mismatch) condition, where both potential antecedents mismatched the pronoun's gender.

[4] {Florian/Marlena} glaubte, dass {der Kollege/die Kollegin}
{F./M.} believed that the colleague$_{\{masc/fem\}}$
aus Frankreich ihn schon bald vorstellen würde.
from France him very soon introduce would
'{Florian/Marlena} believed that the colleague from France would introduce him very soon.'

15 Puebla et al.'s results revealed gender effects of the binding-inaccessible antecedent (e.g., *der Kollege* 'the colleague') during both L1 and L2 processing. The pattern of interference was different across groups, however. The L1 group violated Condition B initially even in the presence of a gender-matching accessible antecedent (e.g., *Florian*), as reflected by main effects of the inaccessible antecedent's gender at a point in time before gender effects of the binding-accessible antecedent emerged. The non-native readers, by contrast, showed later interference effects, and only when no gender-matching, structurally accessible antecedent was available, suggesting a "last resort" attempt to find a potential referent. The authors attributed the presence of interference effects in their study to differences in their design and materials compared to those of Patterson et al. (2014). First, the elaboration of inaccessible antecedents by means of an added prepositional phrase (PP) in Puebla et al.'s study, compared to the plain proper names (e.g., *Mark*, *Jane*) used by Patterson et al., may have increased the chances of inaccessible antecedents being considered.

Second, the absence of a double-mismatch condition in Patterson et al. made it impossible to detect a selective pattern of interference such as the "last resort" strategy observed by Puebla et al. for the L2 group. Note that a similar finding was previously reported by Chow et al. (2014, experiment 1) in a self-paced reading study examining the application of Condition B by English native speakers, where the embedded subject and binding-inaccessible antecedent was elaborated through PP or RC modification as in [5].

[5] Ethan discovered that the analyst who attended the office party had mocked him mercilessly for singing karaoke after drinking...

16 Taken together, the results from Puebla et al. (2021) and Chow et al. (2014) suggest that the degree of syntactic/semantic complexity of an antecedent may modulate its saliency in memory, which may result in elaborated antecedents more likely to be considered for retrieval than less elaborated ones. The fact that Puebla et al.'s L2 speakers did not consider the inaccessible antecedent at any point prior to their consideration of the accessible one suggests that elaboration may not be enough to cause the initial mis-retrieval of inaccessible antecedents during L2 processing, at least not in the presence of a highly prominent accessible antecedent.

17 Binding-inaccessible antecedents may be considered by L2 comprehenders in the absence of a discourse-prominent accessible one, however. Kim et al. (2015) carried out a visual-world eye-tracking study comparing the processing of reflexive and non-reflexive pronouns by adult L1 English speakers and Korean L2 speakers of English. They found that unlike L1 speakers, L2 learners violated Condition B by allowing coreference between object pronouns and a local, binding-inaccessible antecedent in contexts such as *Look at Goofy. Have Mickey touch him*, where the sentence-external accessible referent *Goofy* was an object and, as such, pragmatically/contextually less prominent than the agentive subject *Mickey*.

2. The present study

18 The present study comprises an eye-tracking experiment complemented by an untimed questionnaire task that investigated the role of antecedent discourse prominence in the L1 and L2 interpretation of German object pronouns in Condition B configurations. More specifically, we examined whether introducing a binding-inaccessible referent as the first-mentioned character and discourse topic would increase the likelihood of it being mis-retrieved when encountering a pronoun in a subsequent sentence. The accessibility of the inappropriate antecedent was further increased through RC modification, which rendered it highly elaborated. In our eye-tracking during reading experiment we examined the following research questions:

- Are elaborated and discourse-prominent, but grammatically inappropriate antecedents considered for referential dependency building during L1 and/or L2 processing?

– If yes, at what point during processing do interference effects arise?

19 Additionally, the following two questions were examined in a complementary questionnaire task:

– Are both L1 and L2 comprehenders aware of the interpretive restrictions imposed by Condition B?

– Does antecedent prominence affect L1 and/or L2 German speakers' ultimate referential choices?

20 Although participants completed the offline task after the main eye-tracking experiment, we begin with the presentation of the offline questionnaire.

3. Questionnaire task

21 We used an untimed antecedent choice task to confirm participants' awareness of Condition B and to examine whether rendering an inaccessible antecedent more discourse prominent increases its likelihood of being incorrectly selected.

3.1. Participants

22 The L1 group consisted of 50 native speakers of German (12 male, mean age: 26.46 years, SD: 7.25, range: 19–56). The L2 group was composed of 50 L1 Russian-speaking learners of German (7 male, mean age: 27 years, SD: 5.53, range: 19–45). The L2 group's onset age of acquisition (AoA) of German was 7 years or above (mean AoA: 17 years, SD: 7.13, range: 7–37). At the time of data collection, the L2 group had been exposed to German for 10 years on average (SD: 6.41, range: 1–26 years), and the group's mean German proficiency score according to the paper-and-pencil version of the Goethe placement test (courtesy of the Goethe Institute, 2011) was 24.3 points out of 30, equivalent to a C1 level of the Common European Framework of Reference for Languages (range: 16–30, B2–C2, SD: 3.51). Note that the interpretation of Russian non-reflexive pronouns is also restricted by Condition B in that a pronoun cannot be locally bound by its antecedent (e.g., Bailyn, 2012).[1] A reduced ability to apply Condition B in German on the part of our L2 participants would thus be unlikely to reflect negative L1 influence.

23 Most of the participants were university students who were recruited via e-mail invitations, through advertisements posted online and by flyers distributed in student locations. None of the participants reported a language disorder at the time of data collection. As a small reward for their participation in the study each person received either eight euros or course credit.

1. There is, however, much theoretical debate concerning the exact definition of the notion of local/binding domain and how to account for cross-linguistic differences (see Asarina, 2005; Avrutin, 1994, for discussion).

3.2. Materials

24 Our stimulus materials consisted of 16 sets of two-sentence texts, with a context sentence being followed by a second, critical sentence containing a pronoun. All critical sentences contained the third-person masculine object pronoun *ihn* ('him') and two c-commanding, gender-matching potential antecedent NPs as shown in [6a, b]. The interpretation of the pronoun was constrained by binding Condition B, which allows for coreference with NP1, the matrix subject (e.g., *Otto* in [6a, b]) but blocks coreference with NP2, the embedded subject (e.g., *der Direktor* 'the manager'). For each stimulus item, two experimental conditions were created by manipulating the inaccessible antecedent's discourse prominence in the context sentence. In the prominent condition [6a], NP2 was introduced as the first-mentioned character and subject of the context sentence, conferring on it the status of discourse topic and thus increasing its discourse-level prominence compared to the non-prominent condition [6b].

[6a] NP2 PROMINENT

Ein	kompetenter	Direktor	leitete	die	Firma.
a	competent	$\text{manager}_{\text{masc}}$	ran	the	company

'A competent manager ran the company.'

[6b] NP2 NON-PROMINENT

Eine	gewaltige	Konkurrenz	bedrohte	die	Firma.
a	fierce	competition	threatened	the	company

'A fierce competition threatened the company.'

CRITICAL:

Otto	glaubte,	dass	der	Direktor,	der
O.	thought	that	the	$\text{manager}_{\text{masc}}$	who_{masc}

zuverlässige	Kontakte	in	Schweden	suchte,
reliable	contacts	in	Sweden	searched

ihn	schon	bald	anrufen	würde.
him	very	soon	call	would

'Otto thought that the manager, who was searching for reliable contacts in Sweden, would call him very soon.'

25 In all critical sentences, NP1 (e.g., *Otto*) was a proper name and NP2 (e.g., *der Direktor* 'the manager') consisted of a definite description modified by a RC of exactly same length across items (e.g., *der zuverlässige Kontakte in Schweden suchte* 'who was searching for reliable contacts in Sweden').[2] This modification rendered the inaccessible antecedent considerably more elaborated than NP1. Note that the presence of a gender-matching and binding-accessible antecedent (NP1) makes the pronoun directly resolvable in both conditions.

2. All male proper names used in this study were frequent and stereotypically male German names.

26 The experimental items were mixed with 24 two-sentence fillers, yielding a total of 40 items per questionnaire. The fillers also contained a pronoun and two potential antecedents, but in half of them the pronoun was ambiguous in that it could refer to either antecedent. The remaining 12 fillers contained "unambiguous" plural or feminine pronouns in that only one of the potential referents was an appropriate antecedent for the pronoun. The stimulus items were pseudorandomised so that no two items of the same type appeared adjacent to each other, and were distributed over two presentation lists.

3.3. Procedure

27 Each participant was tested individually in a quiet laboratory room. Demographic information was collected through an online questionnaire that each participant completed remotely prior to the experimental session. At the beginning of the session, the testing protocol was explained in German to each participant, who was then asked to sign a consent form. The offline questionnaire task was administered in paper format after the main eye-tracking experiment. We used an untimed binary antecedent choice task designed to confirm participants' knowledge of Condition B and to examine whether a binding-inaccessible antecedent's discourse prominence influenced a pronoun's ultimate interpretation.

28 Each of the 40 two-sentence stimulus items contained an underlined pronoun. Following each item, two potential antecedents were shown as alternative response options. Participants were instructed to read each item carefully and to select one of the options even if they felt that both answers were possible. The presentation order of the two antecedent options was reversed in half of the items. The items were distributed over nine pages, with up to five items displayed per page. All participants completed the antecedent choice task within 10 minutes.

3.4. Results

	Experimental items (*n*=16)				Filler items (*n*=12)	
	Prominent		Non-prominent			
	Mean	SD	Mean	SD	Mean	SD
L1 group (*n* = 50)	98.5 (87.5–100)	0.12	99.75 (87.5–100)	0.05	98.33 (83.33–100)	0.13
L2 group (*n* = 50)	98.75 (87.5–100)	0.11	99.25 (87.5–100)	0.09	94.83 (75–100)	0.22

Note: Accuracy rates are given in percentages (ranges are in parenthesis); SD = standard deviation.

Table 1 – Results from the questionnaire task

29 The results from the questionnaire task are summarised in Table 1. Both groups' responses to the 12 "unambiguous" fillers were examined in order to confirm that participants had paid attention to the task (see Table 1, "Filler items"). As for the 16 experimental items, both participant groups showed a similar pattern of results, with accuracy rates close to ceiling in both conditions (see Table 1, "Experimental items"). These results indicate that both L1 and L2 speakers correctly ruled out Condition B inaccessible antecedents and that an antecedent's discourse prominence did not affect the participants' ultimate interpretation of the pronoun.

4. Eye-movement monitoring experiment

30 The purpose of our eye-tracking experiment was to test whether and if so, when during processing an inaccessible antecedent's discourse status affects the likelihood of it being considered as a referent for an object pronoun in Condition B constrained sentences.

4.1. Participants

31 Participants were the same as described above, but the data from two participants per group were removed due to track loss. The participants whose data were analysed included 48 German L1 speakers (12 males) and 48 L2 learners (7 males). The L1 group had a mean age of 26.63 years (SD: 7.32, range: 19–56). The L2 speakers' mean age was 27 years (SD: 5.71, range: 19–38). All participants had normal or corrected-to-normal vision.

4.2. Materials

32 The materials for the eye-tracking experiment included the 16 sets of stimulus items used in the questionnaire plus eight further sets of similar two-sentence texts, yielding a total of 24 experimental item sets. Using a gender-mismatch paradigm (e.g., Sturt, 2003), we manipulated both the inaccessible antecedent's discourse prominence and its gender congruence with the pronoun, resulting in four experimental conditions as shown in [7a–d].

33 The materials were arranged in a 2 x 2 Latin-square design with the factors Prominence (prominent, non-prominent) and Gender (match, mismatch) and distributed over four presentation lists. The experimental stimuli were mixed and randomised with 72 filler items, 24 of which belonged to a different experiment, resulting in a total of 96 items per list (plus four practice trials).

34 The set of fillers included 48 two-sentence texts containing a masculine or feminine object pronoun and two potential subject antecedents. Twelve of these closely resembled our experimental items but were simplified (e.g., contained fewer ornamental elements such as adverbs or adjectives) and divergent in other aspects (e.g., linear order and other characteristics of the potential antecedents). The remaining 24 fillers represented a variety of syntactic structures. Half of them contained assorted

pronouns (e.g., reflexives, possessives, relatives) while the other half contained no pronouns and were generally shorter, also including a few single-sentence items.

35 *Yes/no* comprehension questions followed half of all items to help ensure that participants read the stimulus texts properly for comprehension. Eight of the questions asked about the referent for a pronoun. This was to stimulate participants to pay attention to pronouns without leading them to focus consciously on pronouns, which might have given rise to strategic reading patterns. To encourage participants to pay attention to the context sentence, 16 questions asked about information contained within the first sentence. Our experimental stimuli are accessible at the Open Science Foundation website at https://osf.io/5rnsw/?view_only=ac56bbe3deb845008fa9b0cb392c8e66, along with detailed statistical model outputs.

[7a] NP2 PROMINENT & MATCH

Ein	kompetenter	Direktor	leitete	die	Firma.
a	competent	$\text{manager}_{\text{masc}}$	ran	the	company

Otto	glaubte,	dass	der	Direktor,	der
O.	thought	that	the	$\text{manager}_{\text{masc}}$	who_{masc}

zuverlässige	Kontakte	in	Schweden	suchte,
reliable	contacts	in	Sweden	searched

ihn	schon	bald	anrufen	würde.
him	very	soon	call	would

[7b] NP2 PROMINENT & MISMATCH

Eine	kompetente	Direktorin	leitete	die	Firma.
a	competent	$\text{manager}_{\text{fem}}$	ran	the	company

Otto	glaubte,	dass	die	Direktorin,	die
O.	thought	that	the	$\text{manager}_{\text{fem}}$	who_{fem}

zuverlässige	Kontakte	in	Schweden	suchte,
reliable	contacts	in	Sweden	searched

ihn	schon	bald	anrufen	würde.
him	very	soon	call	would

[7c] NP2 NON-PROMINENT & MATCH

Eine	gewaltige	Konkurrenz	bedrohte	die	Firma.
a	fierce	competition	threatened	the	company

Otto	glaubte,	dass	der	Direktor,	der
O.	thought	that	the	$\text{manager}_{\text{masc}}$	who_{masc}

zuverlässige	Kontakte	in	Schweden	suchte,
reliable	contacts	in	Sweden	searched

ihn	schon	bald	anrufen	würde.
him	very	soon	call	would

[7d] NP2 NON-PROMINENT & MISMATCH

Eine	gewaltige	Konkurrenz	bedrohte	die	Firma.
a	fierce	competition	threatened	the	company

Otto	glaubte,	dass	die	Direktorin,	die
O.	thought	that	the	$\text{manager}_{\text{fem}}$	who_{fem}

zuverlässige	Kontakte	in	Schweden	suchte,
reliable	contacts	in	Sweden	searched

ihn	schon	bald	anrufen	würde.
him	very	soon	call	would

'{A competent manager ran the company. / A fierce competition threatened the company.} Otto thought that the manager, who was searching for reliable contacts in Sweden, would call him very soon.'

4.3. Predictions

36 Mis-retrieval of the inaccessible antecedent (NP2) during processing should be reflected in reading-time differences between the gender-matching and mismatching conditions upon encountering the pronoun or shortly after. If interference from NP2 arises regardless of whether it is introduced as the discourse topic, we expect main effects of Gender, that is, a reading-time difference between the matching [7a, c] and mismatching conditions [7b, d]. On the other hand, if the inaccessible antecedent's discourse status affects the chances of it being considered, we should observe Prominence × Gender interactions, with gender effects restricted to the prominent conditions (i.e., a difference between [7a] and [7b] but not between [7c] and [7d]). Given previous findings reported in the L2 processing literature (e.g., Felser, Cunnings, 2012), an interaction pattern is expected in our L2 data but not necessarily in our L1 data.

37 Interference effects can be either facilitatory or inhibitory depending on whether the presence of a gender-matching NP2 leads to a processing speedup (facilitation) or slowdown (inhibition) relative to a gender-mismatching NP2. In target-match configurations like those used in our study, where there is always a gender-matching accessible antecedent available for the pronoun, certain instantiations of cue-based retrieval models (e.g., ACT-R, Lewis, Vasishth, 2005) would predict inhibitory interference, at least for L1 speakers, although empirical evidence is mixed with respect to the type of interference reported across anaphora resolution studies (see Jäger et al., 2017, for review and discussion).

4.4. Procedure

38 We used a desktop eye-tracker (EyeLink 1000, SR Research) to record participants' eye movements while reading the stimulus texts on a computer screen. Items were displayed one by one in black Courier New font (23 pt, bold) on a white background. The target items occupied three lines of text, with the context sentence in the first line and the critical sentence distributed over the second and third

lines. The distribution of each target item on the screen was carefully checked to ensure that the region containing the pronoun always fell in the middle of the third line. Given that pronouns are short words that tend to be skipped, we avoided placing the pronoun in line-initial position, where miscalculated fixations travelling from the previous line might increase the possibility of the pronoun being skipped. Automatic calibration and validation was performed for each participant before detailed instructions appeared on the screen. Participants were instructed to read the texts in silence at their usual pace and for comprehension. To familiarise participants with the task, four practice items were presented at the beginning of the experiment. Each trial began with a black fixation dot that controlled for drift. A stimulus text appeared when the reader looked at the dot while simultaneously pressing the corresponding button on a gamepad. The gamepad was also used to answer the end-of-trial comprehension questions.

39 During the eye-tracking session, participants sat on an adjustable office chair 80 cm from the screen. Below the screen, at a distance of 50 cm from the participant's eyes, the eye-tracker camera was located. To minimise head movements, participants were asked to remain sitting as still as possible with their forehead and chin supported by a head stabilizer. Good tracking was also achieved by recalibrating as many times as necessary during the experiment. Except for one L2 speaker whose right eye was not recognised by the eye-tracker camera, all participants had their right eye recorded. Items were divided in two blocks of 48 items each. In between blocks, participants could take a short break if desired. Depending on the length of the break, the eye-tracking task could take up to 50 minutes.

40 After the online task, the antecedent-choice questionnaire was administered, and the L2 participants additionally completed a paper-and-pencil test consisting of a gender decision task and a vocabulary task. The gender task included a selection of 20 masculine and feminine role nouns which participants had seen during the experiment as potential antecedents for the pronouns. Participants were asked to decide whether these nouns were masculine or feminine. In the vocabulary task, participants were presented with a list of 138 critical words and verb phrases from the experimental items and were asked to mark any words or phrases they were unfamiliar with. Completion of these two tasks did not take longer than five minutes.

4.5. Data analysis

41 Regions of text selected for analysis included the PRONOUN REGION, which contained the critical pronoun and the word following it, and the SPILLOVER REGION, which contained the next two words after the pronoun region. A slowdown in reading times in either of these regions was taken as an indicator of processing costs related to the resolution of the referential dependency. For each interest region we report four continuous and two binomial reading-time measures: (i) FIRST-PASS READING TIMES, the total fixation time contained in a region until the eyes first move out to earlier or later regions; (ii) REGRESSION-PATH DURATION, the summed

duration of all fixations in a region until it is first exited in a forward direction; (iii) REREADING TIME, which was calculated by subtracting the first-pass times from (iv) TOTAL READING TIMES, the sum of all fixations in a region; (v) REGRESSIONS OUT, showing whether the reader looked back from the current region into earlier parts of the text before their eyes entered a later region; and (vi) REGRESSIONS IN, which indicate whether the current region received regressive eye movements from later regions of text. Although the mapping of individual reading-time measures onto specific processing stages or distinct cognitive events remains controversial (e.g., Pickering et al., 2004), we can broadly distinguish between FIRST-PASS or "early" measures (i.e., first-pass times, regression-path duration, and regressions out) and SECOND-PASS or "late" measures (i.e., rereading time and regressions in). Total reading times is considered a cumulative measure indexing global processing load.

42 Prior to the statistical analysis the dataset was checked and experimental trials were cleaned one by one. Fixations with vertical drift were corrected manually (L1 group: 1.14% of total fixations; L2 group: 2.32% of total fixations). Individual fixations falling between regions and not clearly belonging to a run of fixations were manually removed (L1 group: 0.54% of experimental data; L2 group: 0.14% of experimental data). Fixations shorter than 80 ms within one degree of visual angle of a neighbouring fixation were merged. All other short (< 80 ms) and extremely long (> 1000 ms) fixations were automatically removed and excluded from analysis. Regions skipped in first-pass were removed from the dataset. Skipping rates for the pronoun and spillover regions were 7.32 and 0.09% for the L1 group, and 1.8 and 0% for the L2 group, respectively. Additionally, it was necessary to discard eight experimental trials (four per group) due to excessive drift or track loss (L1 group: 0.1% of experimental data; L2 group: 0.1% of experimental data). Data from these trials were treated as missing data and excluded from statistical analysis, but still included in comprehension accuracy calculations. Trials for which L2 participants reported unknown vocabulary were excluded from both accuracy calculations and statistical analysis (2.88% of total L2 data).

43 Data analysis was performed in R (R Core Team, 2016) with the package lme4 (Bates et al., 2015) using mixed modelling. The continuous reading-time measures were log-transformed to satisfy the assumptions of normality (Vasishth, Nicenboim, 2016). The resulting distribution of each transformed measure was checked for normality using the Box-Cox procedure (Box, Cox, 1964). The statistical analyses on the continuous measures were run on log-transformed data with linear-mixed effects models using the R function *lmer*. For the binomial variables, logistic regressions were calculated using the function *glmer*.

44 A between-groups analysis was initially performed on the complete dataset to explore whether the two groups' reading patterns differed statistically. All models contained the sum-coded fixed two-level factors Prominence (prominent, non-prominent), Gender (match, mismatch), Group (L1, L2), and a Prominence × Gender × Group interaction. Trial index (centred) was included as a covariate to

account for potential habituation effects over the course of the experiment. As long as the models converged, they contained the full random structure with random intercepts and slopes for both subjects and items. In case of non-convergence, the random-effects structure was gradually simplified until convergence was achieved by removing elements one by one following the recommendations in Barr et al. (2013).

4.6. Results

45 The end-of-trial comprehension questions were answered with high accuracy by both participant groups. In the native group, response accuracy reached 95% (SD: 0.03, range: 87–100%). The L2 group answered 91% of the questions correctly (SD: 0.05, range: 80–98%). These results confirm that participants had paid attention to the task and read the texts for comprehension. Below we first report the outcomes from the between-groups analysis, followed by the results from the separate per-group analyses.

4.6.1. Between-groups analysis

46 Our omnibus analysis revealed main effects of the factor Group, consistently significant ($p < .05$) across both continuous and binomial reading-time measures in both interest regions. Additionally, marginal main effects of Prominence and Gender emerged in several measures at the pronoun and spillover regions. Besides the reported main effects, the results show a number of two- and three-way interactions with Group. Gender × Group interactions were significant for total reading times ($t = -2.09$, $p = .039$) and regressions out ($z = -2.25$, $p = .025$) at the pronoun region, and for regression-path duration ($t = 2.04$, $p = .042$) and rereading time ($t = -2.13$, $p = .035$) at the spillover region. A marginal Gender × Group interaction emerged for regressions out at the spillover region ($z = -1.84$, $p = .066$). Prominence × Group interactions were significant at the spillover region for regression-path duration ($t = -2.48$, $p = .015$). Finally, there was a marginal three-way Prominence × Gender × Group interaction for first-pass times at the pronoun region ($t = -1.90$, $p = .057$), and another one for rereading time at the spillover region ($t = 1.91$, $p = .057$).

47 As the observed interactions, along with the main effects of Group, indicate a different L1/L2 reading pattern, we went on to analyse the data from the two groups separately. In the following, we report the results from the individual per-group analyses, beginning with the native group.

4.6.2. L1 results

48 Tables 2 and 3 provide an overview of the L1 group's reading times (continuous measures' means and binomial data's proportions) and statistical outcomes for both interest regions, respectively.

49 Pronoun region. Statistically significant main effects of Gender emerged in regressions out of the region containing the pronoun, where we observed a mis-

match effect: the native comprehenders made more regressions in the mismatching conditions [7b, d] than in the matching conditions [7a, c].

50 Spillover region. At the spillover region, the analysis again revealed significant main effects of Gender in regressions out. Additional significant gender effects emerged in regression-path duration. These gender effects go in the same direction as was observed at the pronoun region (mismatch effects), with the gender-mismatch conditions [7b, d] yielding higher processing costs relative to the matching conditions [7a, c].

		Pronoun region	Spillover region
First-pass reading times	(a) Prominent & match	330 (142)	440 (174)
	(b) Prominent & mismatch	336 (160)	435 (197)
	(c) Non-prominent & match	327 (148)	443 (202)
	(d) Non-prominent & mismatch	329 (157)	450 (211)
Regression-path duration	(a) Prominent & match	336 (149)	624 (544)
	(b) Prominent & mismatch	362 (195)	688 (613)
	(c) Non-prominent & match	340 (167)	765 (1017)
	(d) Non-prominent & mismatch	353 (188)	779 (847)
Rereading time	(a) Prominent & match	344 (229)	369 (224)
	(b) Prominent & mismatch	343 (196)	502 (454)
	(c) Non-prominent & match	328 (215)	494 (415)
	(d) Non-prominent & mismatch	342 (246)	435 (282)
Total reading times	(a) Prominent & match	412 (227)	553 (246)
	(b) Prominent & mismatch	441 (247)	584 (365)
	(c) Non-prominent & match	433 (245)	610 (365)
	(d) Non-prominent & mismatch	442 (260)	602 (322)
Regressions out	(a) Prominent & match	0.01 (0.11)	0.20 (0.40)
	(b) Prominent & mismatch	0.04 (0.20)	0.29 (0.45)
	(c) Non-prominent & match	0.02 (0.14)	0.25 (0.43)
	(d) Non-prominent & mismatch	0.04 (0.20)	0.28 (0.45)
Regressions in	(a) Prominent & match	0.17 (0.37)	0.08 (0.27)
	(b) Prominent & mismatch	0.21 (0.41)	0.09 (0.28)
	(c) Non-prominent & match	0.22 (0.41)	0.07 (0.25)
	(d) Non-prominent & mismatch	0.22 (0.42)	0.07 (0.25)

Table 2 – Means in milliseconds and proportions (standard deviations in parentheses) for six eye-movement measures at the pronoun and spillover regions; L1 group

	Pronoun region				Spillover region			
	Est.	SE	t (z)	p	Est.	SE	t (z)	p
	First-pass reading times							
ME Prominence	-0.003	0.011	-0.29	.77	0.005	0.012	0.39	.70
ME Gender	-0.004	0.012	-0.35	.73	0.007	0.013	0.56	.58
Prominence × Gender	-0.002	0.010	-0.17	.87	-0.010	0.015	-0.64	.53
	Regression-path duration							
ME Prominence	0.001	0.011	0.06	.95	0.033	0.020	1.63	.11
ME Gender	-0.019	0.013	-1.52	.14	-0.037	0.017	-2.14	.04*
Prominence × Gender	0.003	0.011	0.23	.82	0.007	0.019	0.36	.72
	Rereading time							
ME Prominence	-0.020	0.030	-0.68	.50	0.018	0.034	0.53	.60
ME Gender	-0.016	0.034	-0.47	.64	-0.050	0.039	-1.28	.21
Prominence × Gender	-0.020	0.029	-0.70	.48	0.059	0.036	1.64	.11
	Total reading times							
ME Prominence	0.017	0.013	1.26	.21	0.025	0.016	1.60	.12
ME Gender	-0.022	0.013	-1.66	.11	-0.008	0.013	-0.61	.55
Prominence × Gender	0.007	0.014	0.48	.64	0.001	0.012	0.09	.93
	Regressions out							
ME Prominence	0.156	0.219	0.71	.48	0.073	0.076	0.96	.34
ME Gender	-0.576	0.219	-2.63	.01*	-0.211	0.076	-2.77	.01*
Prominence × Gender	0.142	0.218	0.65	.52	0.111	0.076	1.46	.15
	Regressions in							
ME Prominence	0.088	0.081	1.10	.27	-0.129	0.115	-1.12	.26
ME Gender	-0.114	0.081	-1.41	.16	-0.055	0.116	-0.48	.63
Prominence × Gender	0.081	0.081	1.01	.31	0.012	0.115	0.10	.92

Note: R-code formula used with lmer (for continuous measures) and glmer (for binomial measures): measure ~ prominence * gender + c.(trial index) + (1 + prominence * gender | subject) + (1 + prominence * gender | item); ME = main effects; Est. = estimate; SE = standard error; * p < .05.

Table 3 – Summary of statistical analyses for six eye-movement measures at the pronoun and spillover regions; L1 group

4.6.3. L2 results

51 An overview of the L2 group's mean reading times (continuous measures) and proportions (binomial data) for our two interest regions is provided in Table 4, and the results of the statistical analysis are shown in Table 5.

52 Pronoun region. The analysis of the pronoun region revealed significant Prominence × Gender interactions in first-pass times and total reading times, and a marginal one in regression-path duration. For total reading times, the follow-up pairwise comparisons confirmed a significant gender-match effect within the non-prominent conditions, with longer reading times for the matching condition [7c] compared to the mismatching condition ([7d], t = -2.15, p = .046). The same pattern, yet statistically marginal, was observed for the interaction in first-pass times (t = -1.71, p = .098). No significant differences were found within the prominent conditions, however.

53 Spillover region. At the spillover region only marginal main effects of Prominence in regression-path duration were observed, with the prominent conditions [7a, b] yielding longer reading times than the non-prominent conditions [7c, d].

		Pronoun region	Spillover region
First-pass reading times	(a) Prominent & match	448 (233)	637 (261)
	(b) Prominent & mismatch	464 (209)	644 (322)
	(c) Non-prominent & match	469 (242)	627 (301)
	(d) Non-prominent & mismatch	434 (201)	628 (291)
Regression-path duration	(a) Prominent & match	513 (500)	903 (916)
	(b) Prominent & mismatch	495 (256)	893 (797)
	(c) Non-prominent & match	521 (353)	825 (638)
	(d) Non-prominent & mismatch	468 (246)	827 (701)
Rereading time	(a) Prominent & match	615 (584)	789 (585)
	(b) Prominent & mismatch	549 (422)	776 (679)
	(c) Non-prominent & match	758 (831)	794 (716)
	(d) Non-prominent & mismatch	591 (400)	746 (572)
Total reading times	(a) Prominent & match	711 (552)	1030 (640)
	(b) Prominent & mismatch	699 (446)	998 (719)
	(c) Non-prominent & match	806 (778)	1031 (742)
	(d) Non-prominent & mismatch	674 (460)	986 (607)
Regressions out	(a) Prominent & match	0.04 (0.20)	0.18 (0.38)
	(b) Prominent & mismatch	0.04 (0.20)	0.19 (0.39)
	(c) Non-prominent & match	0.04 (0.21)	0.16 (0.37)
	(d) Non-prominent & mismatch	0.04 (0.20)	0.15 (0.36)
Regressions in	(a) Prominent & match	0.28 (0.45)	0.25 (0.43)
	(b) Prominent & mismatch	0.31 (0.46)	0.23 (0.42)
	(c) Non-prominent & match	0.30 (0.46)	0.20 (0.40)
	(d) Non-prominent & mismatch	0.27 (0.44)	0.25 (0.43)

Table 4 – Means in milliseconds and proportions (standard deviations in parentheses) for six eye-movement measures at the pronoun and spillover regions; L2 group

	Pronoun region				Spillover region			
	Est.	SE	t (z)	p	Est.	SE	t (z)	p
	First-pass reading times							
ME Prominence	-0.007	0.012	-0.55	.59	-0.011	0.015	-0.77	.45
ME Gender	0.004	0.011	0.31	.76	0.003	0.016	0.17	.86
Prominence × Gender	0.027	0.012	2.31	.03*	-0.005	0.015	-0.35	.73
	Regression-path duration							
ME Prominence	-0.007	0.012	-0.56	.58	-0.030	0.017	-1.79	.08†
ME Gender	0.010	0.014	0.72	.48	0.002	0.017	0.10	.92
Prominence × Gender	0.023	0.012	1.86	.06†	-0.004	0.019	-0.20	.84
	Rereading time							
ME Prominence	0.036	0.032	1.11	.28	-0.036	0.033	-1.09	.29
ME Gender	0.009	0.035	0.24	.81	0.035	0.039	0.92	.37
Prominence × Gender	0.013	0.030	0.43	.67	-0.011	0.031	-0.36	.72
	Total reading times							
ME Prominence	0.005	0.015	0.31	.76	0.000	0.014	0.03	.98
ME Gender	0.018	0.017	1.07	.29	0.017	0.015	1.15	.26
Prominence × Gender	0.034	0.013	2.51	.01*	-0.008	0.013	-0.60	.55
	Regressions out							
ME Prominence	0.000	0.154	0.00	1.00	-0.100	0.083	-1.21	.22
ME Gender	0.037	0.154	0.24	.81	0.002	0.082	0.03	.98
Prominence × Gender	0.001	0.154	0.01	.99	0.032	0.083	0.38	.70
	Regressions in							
ME Prominence	-0.040	0.073	-0.54	.59	-0.048	0.077	-0.63	.53
ME Gender	-0.007	0.073	-0.09	.93	-0.043	0.077	-0.56	.57
Prominence × Gender	0.109	0.073	1.49	.14	-0.111	0.077	-1.45	.15

Note: R-code formula used for continuous (*lmer*) and binomial measures (*glmer*):
measure ~ prominence * gender + c.(trial index) + (1 + prominence * gender | subject)
+ (1 + prominence * gender | item); ME = main effects; Est. = estimate; SE = standard error;
† $p < .10$; * $p < .05$.

Table 5 – Summary of statistical analyses for six eye-movement measures at the pronoun and spillover regions; L2 group

5. Summary of results

54 On the one hand, both participant groups demonstrated awareness of Condition B in our offline task, overwhelmingly selecting accessible antecedents regardless of the inaccessible antecedent's discourse prominence. The analysis of the eye-tracking data, on the other hand, revealed gender effects of Condition B inaccessible antecedents that emerged in first-pass measures for both participant groups. However, the pattern of interference was different across the two groups. For the L1 group, we found main effects of Gender unmodulated by our Prominence manipulation in both interest regions. For the L2 group, by contrast, Gender effects were dependent on the factor Prominence, as reflected by the Prominence × Gender interaction that emerged in various measures at the pronoun region. Here, gender effects were restricted to the non-prominent conditions, where processing was facilitated when the inaccessible antecedent mismatched in gender with the pronoun compared to when it matched.

6. Discussion

55 The current study investigated the effect of an inaccessible antecedent's contextual prominence on the L1 and L2 interpretation and real-time processing of German object pronouns. The data from our offline task showed that both participant groups ruled out a binding-inaccessible antecedent with high accuracy regardless of its discourse status. In contrast, the results from the eye-tracking task revealed interference effects in early reading-time measures for both groups. While our L1 readers were distracted by the inaccessible antecedent regardless of its prominence at the discourse level, the L2 learners showed a selective interference pattern, considering the inaccessible antecedent only if it was not initially introduced as the discourse topic. These findings are discussed in more detail below.

6.1. Inaccessible antecedent mis-retrieval

56 Although both our L1 and L2 participants adhered to Condition B in the offline task, their real-time processing of the critical pronoun was affected by the presence of an elaborated but binding-inaccessible antecedent. This was the case even though we presented readers with target-match items, where a sentence-internal, gender-matching and structurally accessible antecedent for the pronoun was always available. The accessible antecedent moreover was a proper name (e.g., *Otto*), a type of referential expression particularly suitable as an antecedent for a pronoun (e.g., Sanford et al., 1988). Condition B incompatible antecedents were nevertheless considered for dependency building during processing in both L1 and L2 comprehension, although the pattern of the interference was different across groups.

57 The early gender-mismatch effects seen on our L1 group show that the native readers attempted local coreference immediately after encountering the pronoun, and regardless of the inaccessible antecedents' discourse status. Moreover, L1 reading

times were faster when the inaccessible antecedent matched in gender with the pronoun compared to when it did not match, which is indicative of facilitatory interference.

58 A pattern of facilitatory interference in target-match configurations is unexpected under certain implementations of cue-based retrieval models (e.g., ACT-R, Lewis, Vasishth, 2005). However, our L1 pattern of results replicates Puebla et al.'s (2021) finding of early facilitatory interference effects triggered by elaborated but Condition B inaccessible antecedents in native German-speaking readers. These results only partly align with those reported by Chow et al. (2014, experiment 1) in a self-paced reading study with native English speakers. Chow et al. found that binding-inaccessible, PP or RC-modified antecedents were considered only in the absence of a gender-matching accessible antecedent, a finding that they interpreted as signalling a last resort strategy or repair process when initial retrieval fails to identify a suitable antecedent. Note, however, that self-paced reading may not necessarily allow for the detection of subtle processing effects, especially if these are short-lived and limited to early processing stages (Patterson, Felser, 2019).

59 For our L2 group we also observed gender effects in first-pass measures at the pronoun region, but unlike in the L1 group these were modulated by the factor Prominence. A Prominence × Gender interaction was also found for total reading times at the pronoun region. Given that this measure includes both first and second-pass fixations, and that we did not observe any effect in second-pass measures (e.g., rereading times), this effect is likely to reflect earlier processes (e.g., Liversedge et al., 1998). The observed interactions reflected the fact that gender effects were restricted to the non-prominent conditions, where the presence of a gender-matching inaccessible antecedent increased processing cost (i.e., inhibitory interference). This pattern shows that our L2 learners selectively considered the inaccessible antecedent only in cases where it had not previously been introduced as the discourse topic, and indicates that unlike our native group, the L2 readers were sensitive to our manipulation of the extrasentential context, a finding we will discuss in more detail below.

60 Our L2 results differ somewhat from the L2 results reported by Puebla et al. (2021). First, while Puebla et al.'s L2 participants tried to link the pronoun to an inaccessible, PP-modified antecedent only during later stages of processing as a last resort strategy, like Chow et al.'s English native speakers, our L2 readers were affected by the inaccessible antecedent's gender early during processing. It is conceivable that the inaccessible antecedent's relatively stronger elaboration via RC modification in the current study resulted in a stronger mental representation of the referent, strong enough to affect the initial stages of pronoun resolution in both participant groups. A second difference concerns the type of the L2 interference observed across different studies. While the presence of a gender-matching inaccessible antecedent facilitated processing in Puebla et al's study when there was no gender-matching accessible antecedent available within the sentence, in the current study we found evidence for inhibitory interference in our L2 group. This contrast may be due

to the different configurations that were examined across studies. Note that in Puebla et al's study, facilitation was observed only in target-mismatch conditions. In contrast, in the current study we observed inhibition, but our design only included target-match items.

6.2. Sensitivity to discourse-level information in L1 and L2 processing

61 There was no evidence that our L1 participants' processing of the pronoun was affected by the inaccessible antecedent's discourse status. This was different for our L2 group, who showed effects of the inaccessible antecedent's gender that were influenced by our Prominence manipulation. However, contrary to our predictions, introducing the inaccessible antecedent as the discourse topic in the context sentence did not lead to interference effects in our L2 group. Instead, our L2 readers only considered the inaccessible antecedent in the non-prominent conditions, which might be dubbed an "anti-prominence" effect. This L2 reading pattern seems surprising when compared to previous findings showing that discourse-prominent or topical antecedents are generally favoured by non-native comprehenders in the resolution of argument reflexives (Felser, Cunnings, 2012; Felser et al., 2009) and non-reflexive object pronouns (Patterson et al., 2014, experiment 3). The fact that our L2 readers only considered the inaccessible antecedent when it had not previously been introduced indicates that this antecedent's accessibility was actually lower in the "prominent" conditions compared to the "non-prominent" ones.

62 Note that what exactly determines the topic of a discourse remains controversial. According to some researchers, initial mention and frequency of occurrence are strong cues to topichood (e.g., Kieras, 1980; Perfetti, Goldman, 1975). However, topichood operates both within the sentence and in a wider discourse, and recent proposals have put forward more nuanced definitions of topic that include the dynamic and relational aspects of discourse prominence (Von Heusinger, Schumacher, 2019). For example, Cowles et al. (2007: 6) define a discourse topic as a "referent that has been a sentential topic for more than one sentence without any intervening sentence topics". In our prominent conditions, the presence of a sentence-internal accessible antecedent – a proper name in matrix subject position (e.g., *Otto*) – is likely to have triggered a topic shift, resulting in the first-mentioned character losing its discourse topic status and the prominence associated with it. This topic shift may have moved our L2 readers' focus of attention to the binding-accessible antecedent (the matrix subject), boosting its prominence relative to the inaccessible antecedent. The marginal main effect of Prominence we saw in the L2 data at the spillover region, where regression-path times were higher for the prominent relative to the non-prominent conditions, may potentially reflect the cost associated with updating the two antecedent candidates' topic features in memory (Gabriele et al., 2017).

63 Note also that the inaccessible antecedent is mentioned twice in our prominent conditions, which means that upon its second mention in the critical sentence, the inaccessible antecedent's referent is already given or "discourse-old". In our non-prominent conditions, by contrast, the inaccessible antecedent is "discourse-new"

as it was not previously introduced in the context. New information occupies a cognitively more prominent status in the discourse model than old information (e.g., Cowles et al., 2007). Additionally, the fact that the inaccessible antecedent was linearly closer to the pronoun and more elaborated than the accessible antecedent may have contributed to a further increase of its activation level to the point of mis-retrieval.

64 Recall that the gender effects we observed in our L2 data in the non-prominent conditions were gender-match effects, where processing was more costly when the inaccessible antecedent matched in gender with the pronoun compared to when it did not (i.e., inhibitory interference, compare also Drummer, Felser, 2018). Inhibitory interference is predicted by implementations of cue-based retrieval models such as ACT-R (Lewis, Vasishth, 2005) for target-match configurations like those used in our study (e.g., Jäger et al., 2017). However, in the cue-based retrieval literature there is currently no agreement as to what exactly this type of inhibitory interference might index (e.g., Parker et al., 2017). Some researchers claim that it reflects competition between target and competitor items in memory when these match a set of cues at the time of retrieval (e.g., Badecker, Straub, 2002). Others argue that such inhibitory effects do not necessarily provide evidence in favour of retrieval interference, as they could arise as a mere consequence of feature-overwriting between two similar items during encoding (e.g., Dillon et al., 2013). Irrespective of whether the gender-match effect found in our L2 group reflects retrieval or encoding interference, the presence of gender effects shows that the inaccessible antecedent was evaluated for dependency building. Several discourse-level factors may have contributed to a decrease of the inaccessible antecedent's accessibility in our "prominent" conditions relative to the "non-prominent" ones. Our results show that L2 learners were sensitive to these discourse-level factors and to the changes in the discourse-prominence status of antecedent candidates, whilst our native speakers were not.

65 Our findings add to the growing body of research showing L1/L2 differences in the real-time processing of referential dependencies (e.g., Cunnings et al., 2017; Felser, Cunnings, 2012; Felser et al., 2009; Kim et al., 2015; Patterson et al., 2014; Puebla et al., 2021; Roberts et al., 2008; Trompelt, Felser, 2014). Confirming and extending previous findings, our results indicate that L2 comprehenders follow a strongly discourse-based pronoun resolution strategy and are thus more sensitive to subtle changes in a potential referent's discourse prominence compared to L1 speakers. Our results thus support theoretical approaches to L2 processing that argue for an overreliance on discourse-level cues in L2 compared to L1 processing (Clahsen, Felser, 2006, 2018; Cunnings, 2017a and b). However, our finding that interference effects were more selective in our L2 than in our L1 group challenges Cunnings' (2017a) claim that L2 comprehenders should generally be more susceptible to retrieval interference compared to L1 comprehenders.

66 Further research is required to systematically investigate the role of antecedent elaboration and how it affects memory retrieval specifically during L2 processing. Careful examination of the information sources used during structure-building

and encoding, and of the interplay between the multiple factors modulating antecedent accessibility during L2 compared to L1 processing, will provide a better understanding of the memory processes involved during language comprehension.

7. Conclusions

67 The current study revealed a similar pattern of results in our L1 and L2 participants' offline pronoun interpretations but differences in the two groups' reading-time patterns. While the discourse status of binding-inaccessible antecedents did not affect our participants' ultimate referential choices, the observed group differences in the patterns of interference from inaccessible antecedents indicate that the factors modulating antecedent accessibility during processing are differently weighted across populations. Similar discrepancies between real-time pronoun resolution and participants' ultimate interpretations have been reported in previous studies on bilingual anaphor resolution, which emphasizes the need for gathering data from online tasks in order to gain a deeper understanding of the mental processes underlying L2 reference resolution.

68 Our eye-movement results showed that binding Condition B can be violated during early stages of both L1 and L2 processing if inaccessible antecedents are sufficiently elaborated, even when an ideal antecedent candidate can be found within the same sentence. Our results also confirm and extend previous findings on anaphor resolution showing that L2 comprehenders are more sensitive to discourse-level information than native speakers during processing. This in turn suggests that discourse-level information may be more strongly weighted in non-native compared to native language processing. Memory search and retrieval models should be able to account for diverging cross-population processing profiles by allowing for the information sources and constraints that guide pronoun resolution to be differently weighted during comprehension.

References

Alcocer, P. & Phillips, C. 2012. Using Relational Syntactic Constraints in Content-Addressable Memory Architectures for Sentence Parsing. College Park: University of Maryland. Available online: http://www.colinphillips.net/wp-content/uploads/2014/08/alcocer_phillips2012_v2.pdf.

Ariel, M. 1990. *Accessing Noun-Phrase Antecedents*. London – New York: Routledge.

Arnold, J.E. 2010. How Speakers Refer: The Role of Accessibility. *Language and Linguistics Compass* 4 (4): 187–203.

Asarina, A. 2005. Russian Binding Theory: Two Improved Movement Approaches. Cambridge, Massachusetts: MIT Press. Available online: https://web.mit.edu/alya/www/binding.pdf.

Avrutin, S. 1994. *Psycholinguistic Investigations in the Theory of Reference*. PhD thesis. Department of Brain and Cognitive Sciences, MIT.

Badecker, W. & Straub, K. 2002. The Processing Role of Structural Constraints on the Interpretation of Pronouns and Anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28 (4): 748–769.

Bailyn, J.F. 2012. *The Syntax of Russian*. Cambridge: Cambridge University Press.

Barr, D., Levy, R., Scheepers, C. & Tily, H. 2013. Random Effects Structure for Confirmatory Hypothesis Testing: Keep it Maximal. *Journal of Memory and Language* 68 (3): 255–278.

Bates, D., Mächler, M., Bolker, B. & Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67 (1): 1–48.

Box, G. & Cox, D. 1964. An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)* 26 (2): 211–243.

Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht – Cinnaminson: Foris Publications.

Chow, W.Y., Lewis, S. & Phillips, C. 2014. Immediate Sensitivity to Structural Constraints in Pronoun Resolution. *Frontiers in Psychology* 5. DOI: https://doi.org/10.3389/fpsyg.2014.00630.

Clahsen, H. & Felser, C. 2006. Grammatical Processing in Language Learners. *Applied Psycholinguistics* 27 (1): 3–42.

Clahsen, H. & Felser, C. 2018. Some Notes on the Shallow Structure Hypothesis. *Studies in Second Language Acquisition* 40 (3): 693–706.

Colonna, S., Schimke, S. & Hemforth, B. 2012. Information Structure Effects on Anaphora Resolution in German and French: A Crosslinguistic Study of Pronoun Resolution. *Linguistics* 50 (5): 991–1013.

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. DOI: https://doi.org/10.1007/s11245-006-9004-6.

Cunnings, I. 2017a. Parsing and Working Memory in Bilingual Sentence Processing. *Bilingualism: Language and Cognition* 20 (4): 659–678.

Cunnings, I. 2017b. Interference in Native and Non-Native Sentence Processing. *Bilingualism: Language and Cognition* 20 (4): 712–721.

Cunnings, I., Fotiadou, G. & Tsimpli, I. 2017. Anaphora Resolution and Reanalysis during L2 Sentence Processing: Evidence from the Visual World Paradigm. *Studies in Second Language Acquisition* 39 (4): 621–652.

Dillon, B., Mishler, A., Sloggett, S. & Phillips, C. 2013. Contrasting Intrusion Profiles for Agreement and Anaphora: Experimental and Modeling Evidence. *Journal of Memory and Language* 69 (2): 85–103.

Drummer, J.D. & Felser, C. 2018. Cataphoric Pronoun Resolution in Native and Non-Native Sentence Comprehension. *Journal of Memory and Language* 101: 97–113.

Felser, C. 2016. Binding and Coreference in Non-Native Language Processing. In A. Holler & K. Suckow (eds.), *Empirical Perspectives on Anaphora Resolution*. Berlin: De Gruyter: 241–266.

Felser, C. 2019. Structure-Sensitive Constraints in Non-Native Sentence Processing. *Journal of the European Second Language Association* 3: 12–22. DOI: http://doi.org/10.22599/jesla.52.

Felser, C. & Cunnings, I. 2012. Processing Reflexives in a Second Language: The Timing of Structural and Discourse-Level Constraints. *Applied Psycholinguistics* 33 (3): 571–603.

Felser, C., Sato, M. & Bertenshaw, N. 2009. The On-Line Application of Binding Principle A in English as a Second Language. *Bilingualism: Language and Cognition* 12 (4): 485–502.

Foraker, S. & McElree, B. 2007. The Role of Prominence in Pronoun Resolution: Active Versus Passive Representations. *Journal of Memory and Language* 56 (3): 357–383.

Gabriele, A., Fiorentino, R. & Covey, L. 2017. Understanding the Symptoms and Sources of Variability in Second Language Sentence Processing. *Bilingualism: Language and Cognition* 20 (4): 685–686.

Garrod, S. & Terras, M. 2000. The Contribution of Lexical and Situational Knowledge to Resolving Discourse Roles: Bonding and Resolution. *Journal of Memory and Language* 42 (4): 526–544.

Givón, T. 1983. *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam – Philadelphia: John Benjamins Publishing Company.

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.

Gundel, J., Hedberg, N. & Zacharski, R. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69 (2): 274–307.

Hofmeister, P. 2011. Representational Complexity and Memory Retrieval in Language Comprehension. *Language and Cognitive Processes* 26 (3): 376–405.

Hofmeister, P. & Vasishth, S. 2014. Distinctiveness and Encoding Effects in Online Sentence Comprehension. *Frontiers in Psychology* 5. DOI: https://doi.org/10.3389/fpsyg.2014.01237.

Jacob, G., Lago, S. & Patterson, C. 2017. L2 Processing and Memory Retrieval: Some Empirical and Conceptual Challenges. *Bilingualism: Language and Cognition* 20 (4): 691–693.

Jäger, L.A., Engelmann, F., & Vasishth, S. 2017. Similarity-Based Interference in Sentence Comprehension: Literature Review and Bayesian Meta-Analysis. *Journal of Memory and Language* 94: 316–339.

Kieras, D.E., 1980. Initial Mention as a Signal to Thematic Content in Technical Passages. *Memory and Cognition* 8: 345–353. DOI: https://doi.org/10.3758/BF03198274.

Kim, E., Montrul, S. & Yoon, J. 2015. The On-Line Processing of Binding Principles in Second Language Acquisition: Evidence from Eye Tracking. *Applied Psycholinguistics* 36 (6): 1317–1374.

Kush, D. 2013. *Respecting Relations: Memory Access and Antecedent Retrieval in Incremental Sentence Processing*. PhD thesis. Department of Linguistics, University of Maryland. Available online: https://drum.lib.umd.edu/bitstream/handle/1903/14589/Kush_umd_0117E_14582.pdf;sequence=1.

Lewis, R. & Vasishth, S. 2005. An Activation-Based Model of Sentence Processing as Skilled Memory Retrieval. *Cognitive Science* 29 (3): 375–419.

Lewis, R., Vasishth, S. & Van Dyke, J.A. 2006. Computational Principles of Working Memory in Sentence Comprehension. *Trends in Cognitive Sciences* 10 (10): 447–454.

Liversedge, S.P., Paterson, K.B. & Pickering, M.J. 1998. Eye Movements and Measures of Reading Time. In G. Underwood (ed.), *Eye Guidance in Reading and Scene Perception*. Oxford – Amsterdam: Elsevier: 55–75.

Nicol, J. & Swinney, D. 2003. The Psycholinguistics of Anaphora. In A. Barss (ed.), *Anaphora: A Reference Guide*. Malden – Oxford: Blackwell Publishing: 72–104.

Parker, D., Shvartsman, M. & Van Dyke, J.A. 2017. The Cue-Based Retrieval Theory of Sentence Comprehension: New Findings and New Challenges. In L. Escobar, V. Torrens & T. Parodi (eds.), *Language Processing and Disorders*. Newcastle: Cambridge Scholars Publishing: 121–144.

Patterson, C. & Felser, C. 2019. Delayed Application of Binding Condition C during Cataphoric Pronoun Resolution. *Journal of Psycholinguistic Research* 48 (2): 453–475. DOI: https://doi.org/10.1007/s10936-018-9613-4.

Patterson, C., Trompelt, H. & Felser, C. 2014. The Online Application of Binding Condition B in Native and Non-Native Pronoun Resolution. *Frontiers in Psychology* 5. DOI: https://doi.org/10.3389/fpsyg.2014.00147.

Perfetti, C.A. & Goldman, S.R. 1975. Discourse Functions of Thematization and Topicalization. *Journal of Psycholinguistic Research* 4 (3): 257–271. DOI: https://doi.org/10.1007/BF01066930.

Pickering, M.J., Frisson, S., McElree, B. & Traxler, M.J. 2004. Eye Movements and Semantic Composition. In M. Carreiras & C. Clifton, Jr. (eds.), *The On-Line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond*. New York – Hove: Psychology Press: 33–50.

Puebla, C., Patterson, C. & Felser, C. 2021. The Timing of Interference Effects during Native and Non-Native Pronoun Resolution. In V. Torrens (ed.), *Syntax Processing*. Newcastle: Cambridge Scholars Publishing: 171–199.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available online: www.R-project.org.

Reinhart, T. 1983. *Anaphora and Semantic Interpretation*. London – Sydney: Croom Helm.

Reuland, E. 2001. Primitives of Binding. *Linguistic Inquiry* 32 (3): 439–492.

Reuland, E. 2011. *Anaphora and Language Design*. Cambridge, Massachusetts: MIT Press.

Roberts, L., Gullberg, M. & Indefrey, P. 2008. Online Pronoun Resolution in L2 Discourse: L1 Influence and General Learner Effects. *Studies in Second Language Acquisition* 30 (3): 333–357.

Sanford, A. & Garrod, S. 1989. What, When, and How?: Questions of Immediacy in Anaphoric Reference Resolution. *Language and Cognitive Processes* 4 (3–4): 235–262. DOI: https://doi.org/10.1080/01690968908406369.

Sanford, A., Moar, K. & Garrod, S. 1988. Proper Names as Controllers of Discourse Focus. *Language and Speech* 31: 43–56. Available online: https://journals.sagepub.com/doi/pdf/10.1177/002383098803100102.

Stede, M. 2012. *Discourse Processing*. San Rafael: Morgan & Claypool Publishers.

Sturt, P. 2003. The Time-Course of the Application of Binding Constraints in Reference Resolution. *Journal of Memory and Language* 48 (3): 542–562.

Sturt, P. 2013. Syntactic Constraints on Referential Processing. In R.P.G. Van Gompel (ed.), *Sentence Processing*. New York – Hove: Psychology Press: 136–159.

Trompelt, H. & Felser, C. 2014. Variable Binding and Coreference in Non-Native Pronoun Resolution. In W. Orman & M.J. Valleau (eds.), *Proceedings of the 38th Annual Boston University Conference on Language Development*. Somerville, Massachusetts: Cascadilla Press: 471–483.

Troyer, M., Hofmeister, P. & Kutas, M. 2016. Elaboration Over a Discourse Facilitates Retrieval in Sentence Processing. *Frontiers in Psychology* 7. DOI: https://doi.org/10.3389/fpsyg.2016.00374.

Vasishth, S. & Nicenboim, B. 2016. Statistical Methods for Linguistic Research: Foundational Ideas – Part I. *Language and Linguistics Compass* 10 (8): 349–369.

Von Heusinger, K. & Schumacher, P.B. 2019. Discourse Prominence: Definition and Application. *Journal of Pragmatics* 154: 117–127.