2024. *Proc Ling Soc Amer* 9(1). 5705. https://doi.org/10.3765/plsa.v9i1.5705.

The division of main and subordinate discourse information in real-time comprehension

Sanghee J. Kim & Ming Xiang*

Abstract. The study examines the role of discourse structure representation in real-time comprehension, with a focus on how *main discourse* and *subordinate discourse* information guides language processing. Using a web-based visual world eyetracking paradigm, we investigated pronoun resolution in two types of relative clauses: appositive relative clauses (ARCs) and restrictive relative clauses (RRCs).While ARCs were generally considered as secondary and side-commentary information, aligning with subordinate discourse, RRCs were usually construed as part of the main discourse. Making use of this distinction between ARCs and RRCs, the current study examined pronoun resolution with two possible antecedents. In one condition (RRC), the two antecedents were part of the same discourse whereas in the other (ARC), they were in distinct discourse units. We found stronger competition between the two possible antecedents in the former condition than the latter. Our findings can potentially be explained by an encoding interference, where linguistic entities encoded with similar discourse status information can lead to an interference effect.

Keywords. sentence comprehension; discourse structure; appositive relative clause; restrictive relative clause; pronoun resolution; web-based eyetracking

1. Introduction. Previous work on discourse context has examined various types of discourse properties. This includes information packaging such as given and new information (Birner & Ward 1998; Kaiser & Trueswell 2004), viewing discourse as strategies and moves to achieve conversational goals (Lewis 1979; Roberts 2004; Bruce & Farkas 2007; Farkas & Bruce 2010), coherence relation between discourse units (Hobbs 1985; Kehler 2002), or structured representation of discourse units and their relations to one another (Polanyi 1988; Asher & Lascarides 2003). And there has also been numerous work on how different types of discourse information impact language comprehension, such as givenness (Kaiser & Trueswell 2004), the focused or non-focused status of discourse entities (Arnold 1998; Kaiser 2011; Colonna et al. 2015), or topichood (Altmann & Kamide 1999; Clifton Jr. & Frazier 2018). The current study focuses on the effect of discourse structure in language comprehension, especially the division between *main discourse* versus *subordinate discourse* content.

For the current purpose, the main versus subordinate discourse structure is manipulated by two different types of relative clause structures in English: namely appositive relative clauses (ARCs) (1a) and restrictive relative clauses (RRCs) (1b).

(1) a. ARC content is part of the subordinate discourse
The waitress, **who sat near the girls**, was unhappy.

* We are grateful to Jennifer Arnold, Brian Dillon, Monica Do, Chris Kennedy, and Matt Wagers for their feedback and insight at different stages of this work. We thank the members at the Language, Evolution, Acquisition, and Processing (LEAP) workshop, and the Language Processing Lab at the University of Chicago, and the audience at the Syntax, Semantics, and Language Processing Lab (s/Lab) at the University of California, Santa Cruz for their suggestions and comments on the preliminary results of this project. We also appreciate the reviewers and the audience at the LSA Annual Meeting 2024 for their questions and feedback. We thank Thomas Sostarics and Jérémy Zehr for technical help. This work has been supported by the generous NSF DDRI-LING Grant (BCS-2214437). All shortcomings are our own. Authors: Sanghee J. Kim, University of Chicago (sangheekim@uchicago.edu) & Ming Xiang, University of Chicago (mxiang@uchicago.edu).

b. RRC content is part of the main discourse
The waitress **who sat near the girls** was unhappy.

In both examples, the information "the waitress was unhappy" constitutes the main assertion of the sentence, hence part of the main discourse information. However, the information *the waitress sat near the girls* contributes to a different part of the discourse structure in the two examples. In (1a), it is secondary and side-commentary information (Koev 2022), and hence part of the subordinate discourse information (Jasinskaja 2016). On the other hand, in (1b), because RRC is restricting the subject noun phrase *the waitress*, it is part of the main discourse information (e.g., Göbel 2019). Making use of this distinction between ARCs and RRCs, in the current study, we ask how the contrast between main vs. subordinate discourse structure is encoded in real-time language comprehension.

2. Processing main vs. subordinate discourse structure information. Existing work has observed different behavioral patterns when main vs. subordinate discourse information was processed. For instance, sentences with ARCs, as opposed to RRCs, were reported to show reduced difficulty in syntactically complex (Dillon et al. 2014, 2017; Kroll & Wagers 2019; Duff et al. 2023) or ungrammatical sentences (Ng & Husband 2017; McInnerney & Atkinson 2020; Kim & Xiang 2022). Longer syntactic dependencies are known to incur greater processing cost (e.g., Gibson 1998; Grodner & Gibson 2005). Hence relative to (2a), (2b) would result in greater processing difficulty and lower acceptability rating. The processing cost associated with dependency length, however, would be reduced with ARCs, such as in (2c) (Dillon et al. 2014, 2017; Kroll & Wagers 2019; Duff et al. 2023).[1]

(2) a. **The bear is** wearing a hat.
b. **The bear** that is standing on the ball the trainer rolled across the room **is** wearing a hat.
c. **The bear**, who is standing on the ball the trainer rolled across the room, **is** wearing a hat.

Pronoun resolution can also be impacted depending on whether the antecedent originates in an ARC or RRC. In both (3a) and (3b) below, the pronoun *their* is ambiguous in possibly referring to either of the two antecedents *the violinists* and *the singers*. In a self-paced-reading experiment, Kim & Xiang (2023) found longer RTs at the pronoun region in the RRC condition (3b) than the ARC condition (3a).

(3) a. *The violinists*, who admired *the singers*, invited **their** mentors to the party.
b. *The violinists* who admired *the singers* invited **their** mentors to the party.

In the ARC condition (3a), one antecedent (*the violinists*) was part of the main discourse content and the other antecedent (*the singers*) was part of the subordinate discourse content; whereas in the RRC condition (3b), both antecedents were part of the main discourse content. One possibility to explain the pronoun resolution effect in (3) is to assume that discourse status information, i.e., [+MAIN] or [+SURBODINATE], was encoded on the antecedent noun phrases when they were first encountered and encoded into working memory representations. When the

[1] Examples in (2b–2c) were modified from Kroll & Wagers (2019) (Experiment 6). The authors used parentheses instead of commas in their original ARC example.

antecedents share similar discourse status, as in (3b), their encodings are less distinctive from each other, in contrast to the situation in (3a) where the working memory representations of the two antecedents were more distinct. Under this assumption, the observed higher pronoun resolution cost in (3b) could be interpreted as the result of encoding interference between two antecedents that partially overlap in their encoding representations. In particular, antecedents that were more similar in their memory encoding may compete with each other, leading to greater difficulty for pronoun resolution.

The encoding interference account for the processing difference in (3) is in line with previous studies that have assumed encoding interference when there was a greater degree of feature overlap between multiple linguistic items that need to be encoded, leading to degraded or weaker memory representations (Nairne 1990, 2002; Oberauer & Kliegl 2006; Oberauer & Lange 2008), or "activation leveling" of the elements that share overlapping features (Villata et al. 2018). The same account could potentially be extended to explain the reduced length penalty effect in (2c) as well. The content inside ARC is part of the subordinate discourse, and it does not interfere with the rest of the main discourse content for the subject-verb dependency resolution.

One empirical challenge to the encoding interference account, however, is that the processing difference between the ARC and RRC conditions in (3) was only observed on the pronoun. Since this was also the region where memory retrieval of the pronoun antecedents took place, the effect observed in (3) may be due to retrieval interference instead of encoding interference. Although in some previous studies, encoding interference effects sometimes were argued to have emerged at the retrieval site (e.g., Gordon et al. 2001, 2004; Hofmeister & Vasishth 2014; Villata et al. 2018), it would be stronger evidence if the hypothesized encoding interference effect can be observed prior to memory retrieval of the relevant representations (e.g., Acheson & MacDonald 2011; Kush et al. 2015; Rich & Wagers 2020). Previous studies that have found processing differences between ARC and RRC structures were primarily sentence judgment task experiments (e.g., Dillon et al. 2014; Kroll & Wagers 2019; Dillon et al. 2018) or reading studies using eyetracking (e.g., Dillon et al. 2017) or self-paced-reading paradigm (e.g., Ng & Husband 2017; McInnerney & Atkinson 2020; Kim & Xiang 2022, 2023). It is possible that reading time measures may not be sensitive enough to detect the early effect of encoding interference. In this study, we use a visual world paradigm eyetracking method, following Sekerina et al. (2016), wherein the paradigm was used to examine an interference effect for resolving linguistic dependency.

3. Visual world paradigm for examining pronoun resolution and the competition effect. The visual world paradigm is particularly effective for examining how multiple representations compete for activations. For example, Sekerina et al. (2016) compared the two conditions in (4) with four images presented on the screen: *button* (target), *pen*, *key*, *earring*. In (4a), multiple objects were plausible targets to be "spotted" whereas in (4b), there was only one targeted object (e.g, *button*) that can be "sewed." When the verb information appeared in the audio stream, there were more looks to the target object (e.g., *button*) in the no-competition condition (4b) than in the competition condition (4a).

(4) Key contrast in Sekerina et al. (2016)

a. Competition condition
It was the button that the maid who returned from vacation **spotted** in the early morning.

b. No-competition condition
It was the button that the maid who returned from vacation **sewed** in the early morning.

Particularly relevant for the current purpose, the visual world paradigm has been applied to study anaphora resolution, in particular when there may be multiple possible antecedents competing with each other (Arnold et al. 2000; Kaiser et al. 2009; Clackson et al. 2011; Runner & Head 2014; Han et al. 2021). For example, Clackson et al. (2011) examined a case in which two antecedents could potentially compete due to gender feature overlap. An example from Clackson et al. (2011) Experiment 2 is given in (5).

(5) Key contrast in Clackson et al. (2011)

a. No gender feature overlap
Susan was waiting outside the corner shop. *She*$_{[+\text{FEM}]}$ watched as *Mr. Jones*$_{[+\text{MASC}]}$ bought a huge box of popcorn for **her** over the counter.

b. Gender feature overlap
Peter was waiting outside the corner shop. *He*$_{[+\text{MASC}]}$ watched as *Mr. Jones*$_{[+\text{MASC}]}$ bought a huge box of popcorn for **him** over the counter.

Due to binding constraints, there was only one grammatical antecedent for the bold-faced pronouns in (5a) and (5b), namely the matrix subject. The intervening noun phrase, Mr. Jones, either did not overlap (5a) or overlapped (5b) in gender with the grammatical antecedent.Four images were presented on the screen as each stimulus sentence was auditorily presented. The images included the two characters (e.g., Susan and Mr. Jones), the object (e.g., *popcorn*) mentioned in the stimulus sentence, and one distractor inanimate object not mentioned in the stimulus. More looks were found towards the grammatical antecedent (e.g., looks towards Susan upon hearing *her* in the (5a) condition). Crucially, however, the proportion of looks to the target was reduced in the overlap condition (5b), suggesting the overlap in gender features between the grammatically licit and illicit entities leads to "competition," which was manifested as a smaller eye gaze difference between the two relevant images. Similar observations were also reported in Runner & Head (2014).

The current study uses the visual world paradigm to examine the contrast demonstrated in (3). We hypothesize that greater overlap in discourse structure information between two linguistic representations can have a real-time impact on comprehension, modulating participants' eye-movement behavior. Specifically, we expect to observe smaller eye gaze differences between the two plausible antecedents when they share the same discourse structure information (3b) compared to when they do not (3a). We also examine the time course of when the eye gaze difference would be exhibited.

4. Experiment.

4.1. SUBJECTS. Sixty-five native speakers of American English were recruited via Prolific. Two participants were omitted from the main analysis as they took part in the study twice, resulting in a total of 63 participants for the analysis (mean age = 30.36; age range: 19–40). The duration of the experiment was around 35 minutes, and the participants received 8.00 USD in compensation.

4.2. DESIGN AND MATERIALS. An example set of experimental item is presented in (6). As explained earlier in (3), the primary interest in the design is whether the two grammatical antecedents for the pronoun *their* share similar discourse status. The discourse status of the antecedents was manipulated through clause type (ARC vs. RRC). In the two ARC conditions (6a–6b), one antecedent was part of the main discourse content and the other is part of the subordinate discourse content. In the two RRC conditions (6c–6d), both antecedents were part of the main discourse content. Additionally, we varied whether there was a modifier following the pronoun: short vs. long (e.g., an additional modifier *musical* in (6b & 6d). This was done to provide some extra buffer time to process the pronoun *their*. The contrast in the RC structure was signaled by adding pauses both before *who* and after the RC boundary (e.g., the offset of *singers*), following previous studies on the intonational pattern of ARC structures (Dehé & Kavalova 2007; Dehé 2014; Watson & Gibson 2004). For all the critical items, the critical pronoun is always *their*. The two antecedent NPs were created from various vocation names.

(6) Experimental conditions (Clause type x Modifier length)

a. ARC & Short-modifier
The violinists$_{[+\text{MAIN}]}$, who admired *the singers*$_{[+\text{SUBORDINATE}]}$, invited **their** mentors to the party.

b. ARC & Long-modifier
The violinists$_{[+\text{MAIN}]}$, who admired *the singers*$_{[+\text{SUBORDINATE}]}$, invited **their** musical mentors to the party.

c. RRC & Short-modifier
The violinists$_{[+\text{MAIN}]}$ who admired *the singers*$_{[+\text{MAIN}]}$ invited **their** mentors to the party.

d. RRC & Long-modifier
The violinists$_{[+\text{MAIN}]}$ who admired *the singers*$_{[+\text{MAIN}]}$ invited **their** musical mentors to the party.

4.3. PROCEDURE. A total of 32 critical and 20 filler items were pseudo-randomly presented auditorily in a visual world paradigm. The experiment was implemented on PCIbex (Zehr & Schwarz 2018), and the trials were distributed in a Latin square design. All filler trials also contained a pronoun that needed to be resolved.

We did not video-record any information during the experiment; only the subjects' eye-gaze information was recorded. Upon completing a consent form and an instruction phase, participants proceeded to the calibration stage. During this initial calibration process, the subjects were asked to look at a green dot (size of 48 px width and 48 px height (approximately 36 pt)) appearing on the screen and to click it when it appeared. A total of 9 dots were presented; the first dot started at the center of the participants' screen (the position of the dot was calculated based on the user's screen size), and the rest appeared in random order, at each edge point of the screen (top left, top center, top right, center left, center right, bottom left, bottom center, and bottom right). Each dot stayed on the screen for 750 ms.[2] The calibration precision was calculated by using the Euclidean distance between the stimulus point, e.g., mid-center dot (X_1, Y_1), and the estimated gaze at the

[2] In the newer PennController version 2.1, the calibration dots do not stay on the screen for a fixed amount of time. Once participants click on one dot, it disappears and the next dot appears on the screen.

n-th dot (X_2, Y_2). The precision score was calculated as in Equation 1. The threshold for acceptable precision score was set as 60, and the participants were asked to repeat the calibration process when they failed to reach the threshold. The study would be aborted if a participant made 3 failed attempts to pass the initial calibration phase.

$$Precision = 100 - \frac{\sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}}{1/2 * WindowHeight} * 100 \quad (1)$$

Calibration was conducted before each trial as well, but the between-trial calibration score was calculated only based on the precision of the mid-center calibration dot. The between-trial calibration score threshold was set to 50.

Once the participant passed the calibration, a fixation cross ("+") appeared at the screen center for 500 ms. Subsequently, four images in black-and-white line drawings were displayed on each quadrant in a randomized order (Figure 1). Each quadrant was a size of 50vw * 50vh (50% of the width and 50% of the height of the participant's screen), and each image had a size of 20vw * 20vh. Two of the images depicted the two possible antecedents and the remaining two images were irrelevant distractors. Visual stimuli were produced with OpenAI's DALL-E 2 (Ramesh et al. 2022), an AI system generating images when given a language prompt.

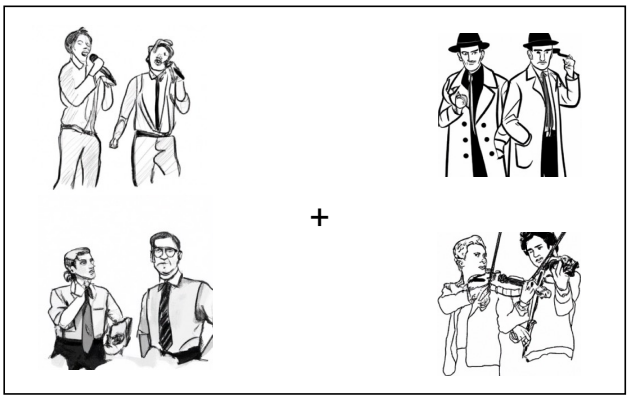

Figure 1. An example of the four images presented to the participants.
Note. Clockwise from the top-left: *the singers*, *the detectives*, *the violinists*, *the teachers*.

After 2200 ms of the image display (image preview time), the auditory stimulus started playing and participants' eye-movement recording also started simultaneously. The eye-movement recording continued for 500 ms after the end of the auditory stimulus. An antecedent selection task followed, with a question prompt asking, "Whose mentors were invited?" Participants were instructed to click on one of the four pictures.

4.4. ANALYSIS. Both antecedent selection and eye-movement patterns were analyzed. The antecedent selection results were presented in section 5.1. Before analyzing the eye-movement data, participants were considered for removal if their eye gaze went outside the quadrants above 50% of the recorded data. No participant was removed based on this criterion. To assess variations in sampling interval inherent in web-based eyetracking data, we excluded data points from trials with an average sampling interval that exceeded 100 ms. The mean sampling interval across trials was 33.57 ms (median = 32.82 ms). After applying the sampling interval threshold, 0.01% of the data were removed. Subsequently, we eliminated data points where the eye gaze did not fall

within a quadrant (an additional 9.71% data point removal). We also excluded data points where an incorrect image was selected for the trial, i.e., neither of the two NPs selected in the trial (an additional 1.06% data point removal). Finally, data points falling outside the designated monitor screen were removed, constituting an additional 0.91% data point removal.

We aggregated the eye-movement data into 100 ms time bins following Slim & Hartsuiker (2022). The proportions of eye-gaze fixations on the four pictures were computed for each time bin (e.g., 0–100 ms, 100–200 ms, ...) across all participants. Taking into account the 200 ms oculomotor delay, we offset the onset of the auditory stimuli by 200 ms. Given our primary focus on the competition between the two grammatical antecedents, we computed the eye-gaze difference between NP1 and NP2. This involved subtracting the proportion of eye gazes to NP2 from that of NP1 for each bin. Using this eye-gaze difference score as the dependent variable, we performed mixed-effects linear regression models with the `lmerTest` package (Bates et al. 2015) in R (R Core Team 2022).

We identified two broad windows of analysis. The first includes the second NP and the two words prior to that. Hence, the first window started from the onset of the relative pronoun (*who*) and ended at the offset of the second NP (e.g., *the singers*). A separate analysis was carried out for each word within this 3-word window. For the first window, our primary interest is the encoding effect of the antecedent noun phrases prior to pronoun resolution. The second window of interest started from the onset of the matrix verb (e.g., *invited*) and ended at the offset of the object NP (e.g., *mentors*). Again a separate analysis was carried out for each word within this 3-word window (*invited*, *their*, and *mentors*). For this window, the primary interest was on the pronoun, but we were also interested in whether there was any effect on the verb, prior to pronoun resolution taking place. Additionally, each word window was identified by including the eye gaze information from the onset to the offset of the word.

In the first pass of the analysis, we included Length, Clause, and their interactions as the fixed effect and a by-subject random intercept. None of the models we conducted found statistically reliable 2-way interaction in any of the word windows (all $|t|s < 1$). Since our primary interest is only Clause type, we removed the Length variable from the models and included the Clause effect (RRC vs. ARC) as a fixed effect and a by-subject random intercept. The Clause variable sum-coded: RRC = -0.5 and ARC = 0.5. In the results reported below, we followed the convention of considering results statistically significant when $|t| > 2$ (Gelman & Hill 2006).[3]

5. Results.

5.1. ANTECEDENT SELECTION. The antecedent selection preference results are presented in Figure 2. Irrespective of conditions, NP1 exhibited a higher preference, approximately 65%, while NP2 showed a preference of around 35%. Two of the distractors were rarely selected.

5.2. EYE-GAZE PROPORTIONS. Figure 3 shows the proportion of eye gaze on each image for every 100 ms time bin, averaged across all participants.

The first window The difference between eye gazes on NP1 and NP2 is illustrated in Figure 4. Regression models revealed no statistically reliable effects at any of the word regions ($|t|s < 2$ for all three individual words, WHO (*who*), VERB1 (e.g., *admired*), and NP2 (e.g., *the singers*)).

[3] See Kim & Xiang (2024) for results analyzed using a cluster-based permutation analysis (Maris & Oostenveld 2007; Ito & Knoeferle 2023). The key findings are identical that statistically significant effects are found only in the second window and not in the first window.

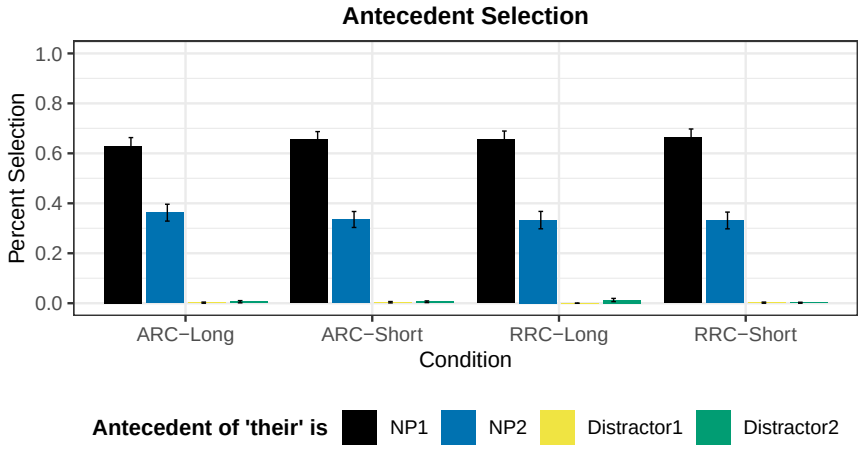

Figure 2. Antecedent preference.

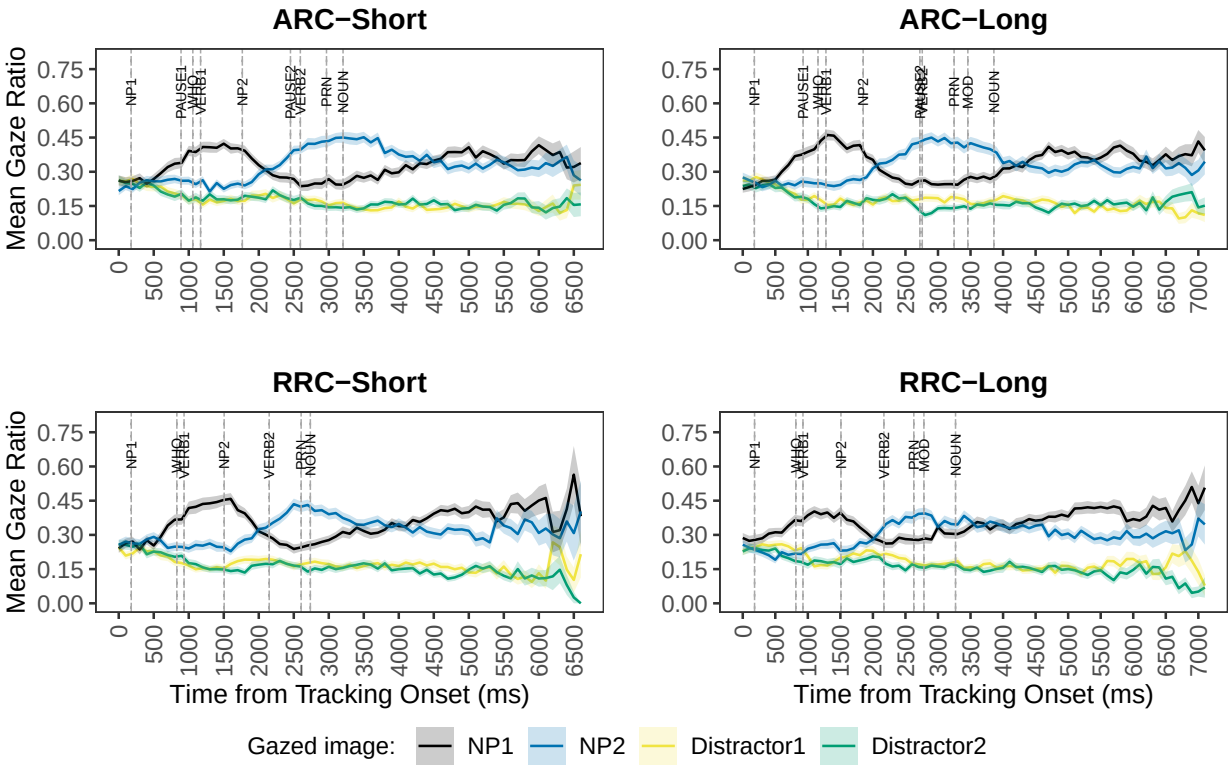

Figure 3. Proportion of eye gaze on each image, offset for 200 ms.
Note. The labels above the vertical lines denote the onset of each word averaged across items. NP1=*the violinists*; VERB1=*admired*; NP2=*the singers*; VERB2=*invited*; PRN (pronoun)=*their*; MOD (modifier)=*musical*; NOUN=*mentors*; PAUSE1/2 represents pauses inserted in the ARC condition, as in the trial "The violinists, who admired the singers, invited their (musical) mentors to the party." Shaded ribbons indicate the standard errors of the mean.

The second window Eye-gaze difference in the second window is illustrated in Figure 5. Regression models showed that there was an effect of Clause in all three windows, VERB2 (β =

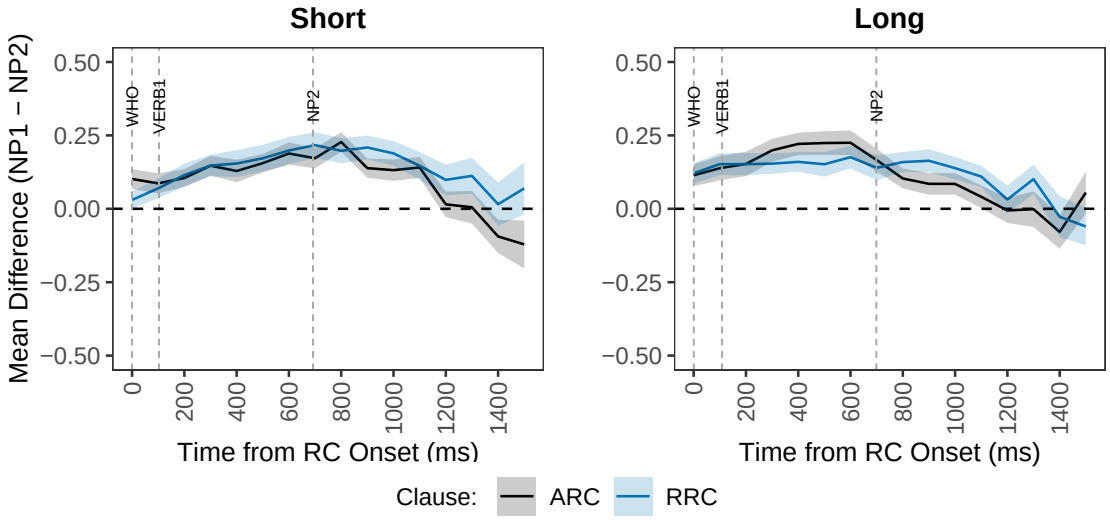

Figure 4. Eye-gaze difference between NP1 and NP2 in the first window.
Note. The X-axis is time-locked to the onset of "WHO" across items. The labels on top of the vertical lines indicate the onset of each word averaged across items. WHO=*who*; VERB1=*admired*; NP2=*the singers*, as in the trial "The violinists(,) who admired the singers(,) invited their mentors to the party." Shaded ribbons indicate the standard errors of the mean.

$-0.098, se = 0.030, t = -3.296$), pronoun (PRN) ($\beta = -0.085, se = 0.034, t = -2.486$), and noun (NOUN) ($\beta = -0.081, se = 0.032, t = -2.517$) windows, where the eye-gaze difference between NP1 and NP2 was greater in the ARC conditions compared to the RRC conditions.

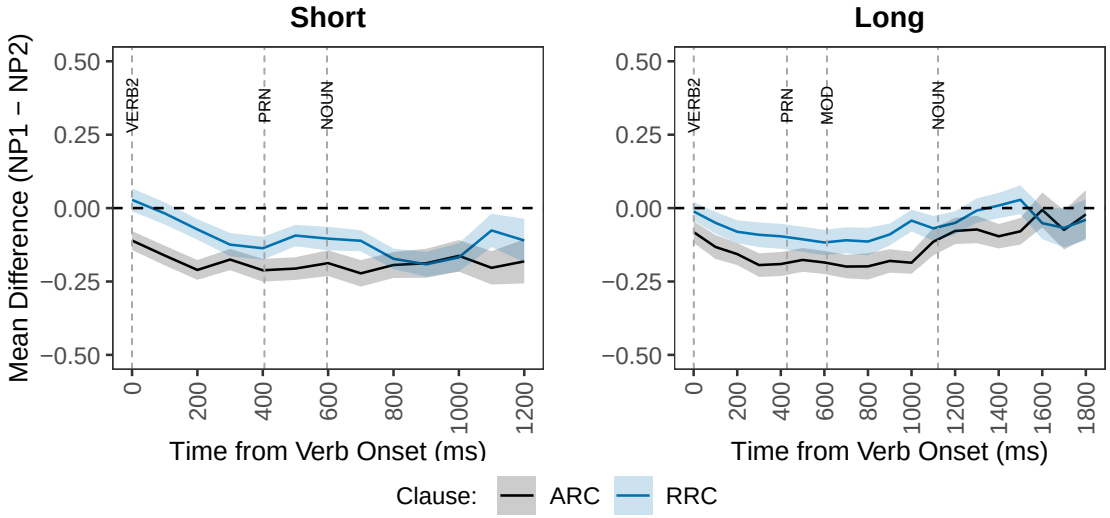

Figure 5. Eye gaze difference between NP1 and NP2 in the second window.
Note. The X-axis is time-locked to the onset of VERB2 across items. The labels on top of the vertical lines indicate the onset of each word averaged across items. VERB2= *invited*; PRN (pronoun)= *their*; MOD (modifier)=*musical*; NOUN=*mentors*, as in the trial "The violinists(,) who admired the singers(,) invited their (musical) mentors to the party." Shaded ribbons indicate the standard errors of the mean.

6. Discussion. In the antecedent selection task, there was an across-the-board preference for the first NP as the antecedent of *their*. The preference for the NP1 in the selection task is also consistent with the eye-gaze results. As shown in Figure 3, in all the conditions the eye gaze towards the NP1 rises around 1000 ms after the onset of the final word of the sentence. This general preference for NP1 is in line with previous observations that participants prefer to choose the subject of the same clause when faced with ambiguous antecedents, as noted in Arnold (2010). It is worth noting that although the ultimate antecedent preference is the same across all the conditions, there are condition-specific online eye-gaze differences. This suggests that the observed eye-gaze differences between ARC vs. RRC conditions cannot be solely attributed to subjects' antecedent preferences.

The analyses of eye-gaze proportions found evidence to suggest that participants were indeed sensitive to the discourse status of potential antecedents of a pronoun. When the two antecedents shared discourse status, i.e. when both antecedents were part of the main discourse content in the RRC conditions, competition arises and the eye-gaze difference between the two antecedents was smaller than in the ARC conditions. An important observation is that the ARC vs. RRC difference did not emerge immediately after the second antecedent (NP2) appeared in the auditory input (see Figure 4). Instead, the effect appeared first on the matrix verb (e.g., *invited*) and then continued into the pronoun region. The fact that the observed effect appeared before the pronoun suggests that it is not simply triggered by the need for pronoun resolution. There are different possibilities as to why the effect started on the verb. One possibility is that at the matrix verb, participants need to activate the corresponding subject NP. The memory retrieval of the subject NP was prone to interference from the intervening NP, resulting in competition between the two NPs. The competition between the two NPs was stronger in the RRC conditions since the two NPs shared discourse status features. Alternatively, the effect at the matrix verb might be a spillover effect from the previous time window, in which the encoding of the second noun phrase was initiated. As shown by the experimental stimuli example in (6), the NP2 antecedent immediately preceded the matrix verb. It was possible that the process to fully encode the NP2 representation, including its discourse status information, was not entirely completed by the time the matrix verb appeared. In the current design, it is difficult to tease apart whether the observed effect on the verb was truly driven by the matrix verb or it was a spill-over effect from encoding the previous noun. Future work is needed to address this issue (see Kim & Xiang (2024) for a discussion).

7. Conclusion. In the current study, an eyetracking experiment showed that discourse structure information, specifically the distinction between main vs. subordinate discourse status information, impacts pronoun resolution. Our findings can potentially be explained through encoding interference, i.e. two memory representations sharing similar discourse information features can result in processing interference. But stronger conclusions can only be made with further work.

References

Acheson, Daniel J. & Maryellen C. MacDonald. 2011. The rhymes that the reader perused confused the meaning: Phonological effects during on-line sentence comprehension. *Journal of Memory and Language* 65(2). 193–207. https://doi.org/10.1016/j.jml.2011.04.006.

Altmann, Gerry T. M. & Yuki Kamide. 1999. Incremental interpretation at verbs: Restricting the

domain of subsequent reference. *Cognition* 73(3). 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1.
Arnold, Jennifer E. 1998. *Reference form and discourse patterns*. Stanford, CA: Stanford University dissertation.
Arnold, Jennifer E. 2010. How speakers refer: The role of accessibility. *Language and Linguistics Compass* 4(4). 187–203. https://doi.org/10.1111/j.1749-818X.2010.00193.x.
Arnold, Jennifer E., Janet G. Eisenband, Sarah Brown-Schmidt & John C. Trueswell. 2000. The immediate use of gender information: Eyetracking evidence of the time-course of pronoun resolution. *Cognition* 76. B13–B26. https://doi.org/10.1016/S0010-0277(00)00073-1.
Asher, Nicholas & Alex Lascarides. 2003. *Logics of conversation*. Cambridge: Cambridge University Press.
Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.
Birner, Betty J. & Gregory Ward. 1998. Information status and noncanonical word order in English. In *Studies in language companion series*. Philadelphia: John Benjamins.
Bruce, Kim B. & Donka F. Farkas. 2007. Context structure for dialogues. Unpublished manuscript, University of California, Santa Cruz.
Clackson, Kaili, Claudia Felser & Harald Clahsen. 2011. Children's processing of reflexives and pronouns in English: Evidence from eye-movements during listening. *Journal of Memory and Language* 65(2). 128–144. https://doi.org/10.1016/j.jml.2011.04.007.
Clifton Jr., Charles & Lyn Frazier. 2018. Context effects in discourse: The question under discussion. *Discourse Processes* 55(2). 105–112. https://doi.org/10.1080/0163853X.2017.1330029.
Colonna, Saveria, Sarah Schimke & Barbara Hemforth. 2015. Different effects of focus in intra- and inter-sentential pronoun resolution in German. *Language, Cognition and Neuroscience* 30(10). 1306–1325. https://doi.org/10.1080/23273798.2015.1066510.
Dehé, Nicole. 2014. *Parentheticals in spoken English: The syntax-prosody relation*. Cambridge: Cambridge University Press.
Dehé, Nicole & Yordanka Kavalova. 2007. Parentheticals: An introduction. In Nicole Dehè & Yordanka Kavalova (eds.), *Parentheticals*, 1–24. Amsterdam: John Benjamins.
Dillon, Brian, Charles Clifton Jr. & Lyn Frazier. 2014. Pushed aside: Parentheticals, memory and processing. *Language, Cognition and Neuroscience* 29(4). 483–498. https://doi.org/10.1080/01690965.2013.866684.
Dillon, Brian, Charles Clifton Jr., Shayne Sloggett & Lyn Frazier. 2017. Appositives and their aftermath: Interference depends on at-issue vs. not-at-issue status. *Journal of Memory and Language* 96. 93–109. https://doi.org/10.1016/j.jml.2017.04.008.
Dillon, Brian, Lyn Frazier & Charles Clifton Jr. 2018. No longer an orphan: Evidence for appositive attachment from sentence comprehension. *Glossa* 3(1). https://doi.org/doi.org/10.5334/gjgl.379.
Duff, John, Pranav Anand, Adrian Brasoveanu & Amanda Rysling. 2023. Pragmatic representations and online comprehension: Lessons from direct discourse and causal adjuncts. *Glossa Psycholinguistics* 2(1). https://doi.org/10.5070/G6011198.
Farkas, Donka F. & Kim B. Bruce. 2010. On reacting to assertions and polar questions. *Journal of Semantics* 27(1). 81–118. https://doi.org/10.1093/jos/ffp010.

Gelman, Andrew & Jennifer Hill. 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68(1). 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1.
Göbel, Alexander. 2019. Final appositives at the right frontier: An experimental investigation of anaphoric potential. In M. Teresa Espinal, Elena Castroviejo, Manuel Leonetti, Louise McNally & Cristina Real-Puigdollers (eds.), *Proceedings of Sinn und Bedeutung* 23, 451–467. Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès). https://doi.org/10.18148/sub/2019.v23i1.543.
Gordon, Peter C., Randall Hendrick & Marcus Johnson. 2001. Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27(6). 1411. https://doi.org/10.1037/0278-7393.27.6.1411.
Gordon, Peter C., Randall Hendrick & Marcus Johnson. 2004. Effects of noun phrase type on sentence complexity. *Journal of Memory and Language* 51(1). 97–114. https://doi.org/10.1016/j.jml.2004.02.003.
Grodner, Daniel & Edward Gibson. 2005. Consequences of the serial nature of linguistic input for sentenial complexity. Cognitive Science 29(2). 261–290. https://doi.org/10.1207/s15516709cog0000$_7$.
Han, Chung-hye, Keir Moulton, Trevor Block, Holly Gendron & Sander Nederveen. 2021. Pronouns are as sensitive to structural constraints as reflexives in early processing: Evidence from visual world paradigm eye-tracking. *Frontiers in Psychology* 12. 611466. https://doi.org/10.3389/fpsyg.2021.611466.
Hobbs, Jerry R. 1985. *On the coherence and structure of discourse*. Stanford, CA: CSLI Technical Report 85-37.
Hofmeister, Philip & Shravan Vasishth. 2014. Distinctiveness and encoding effects in online sentence comprehension. *Frontiers in Psychology* 5. 1237. https://doi.org/10.3389/fpsyg.2014.01237.
Ito, Aine & Pia Knoeferle. 2023. Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods* 55(7). 3461–3493. https://doi.org/10.3758/s13428-022-01969-3.
Jasinskaja, Katja. 2016. Not at issue any more. Unpublished manuscript, University of Cologne.
Kaiser, Elsi. 2011. Focusing on pronouns: Consequences of subjecthood, pronominalisation, and contrastive focus. *Language and Cognitive Processes* 26(10). 1625–1666. https://doi.org/10.1080/01690965.2010.523082.
Kaiser, Elsi, Jeffrey T. Runner, Rachel S. Sussman & Michael K. Tanenhaus. 2009. Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition* 112(1). 55–80. https://doi.org/10.1016/j.cognition.2009.03.010.
Kaiser, Elsi & John C. Trueswell. 2004. The role of discourse context in the processing of a flexible word-order language. *Cognition* 94(2). 113–147. https://doi.org/10.1016/j.cognition.2004.01.002.
Kehler, Andrew. 2002. *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.
Kim, Sanghee J. & Ming Xiang. 2022. Memory retrieval selectively targets different discourse units. The 35th Annual Conference on Human Sentence Processing (HSP2022), University

of California at Santa Cruz, CA. https://doi.org/10.17605/OSF.IO/2UMRE. March 24–26 (plenary talk).

Kim, Sanghee J. & Ming Xiang. 2023. Memory encoding/retrieval is sensitive to discourse status: Through the lens of pronoun resolution. The 36th Annual Conference on Human Sentence Processing (HSP2023), University of Pittsburgh, PA. https://doi.org/10.17605/OSF.IO/KH68W. March 9–11 (plenary talk).

Kim, Sanghee J. & Ming Xiang. 2024. Encoding discourse structure information during language comprehension: Evidence from web-based visual world paradigm experiments. In Larissa K Samuelson, Stefan Frank, Mariya Toneva, Allyson Mackey & Eliot Hazeltine (eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society*, https://escholarship.org/uc/item/3m93r04n.

Koev, Todor. 2022. *Parenthetical meaning* (Oxford Studies in Semantics and Pragmatics). Oxford: Oxford University Press.

Kroll, Margaret & Matthew W. Wagers. 2019. Working memory resource allocation is not modulated by clausal discourse status. Unpublished manuscript, University of California, Santa Cruz.

Kush, Dave, Jeffrey Lidz & Colin Phillips. 2015. Relation-sensitive retrieval: Evidence from bound variable pronouns. *Journal of Memory and Language* 82. 18–40. https://doi.org/10.1016/j.jml.2015.02.003.

Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8. 339–359. https://doi.org/10.1007/BF00258436.

Maris, Eric & Robert Oostenveld. 2007. Nonparametric statistical testing of EEG-and MEG data. *Journal of Neuroscience Methods* 164(1). 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024.

McInnerney, Andrew & Emily Atkinson. 2020. Syntactically unintegrated parentheticals: Evidence from agreement attraction. The 33rd Annual CUNY Human Sentence Processing, University of Massachusetts Amherst: Amherst, MA. March 19–21 (oral presentation).

Nairne, James S. 1990. A feature model of immediate memory. *Memory & Cognition* 18. 251–269. https://doi.org/10.3758/BF03213879.

Nairne, James S. 2002. The myth of the encoding-retrieval match. *Memory* 10(5-6). 389–395. https://doi.org/10.1080/09658210244000216.

Ng, Anne & Matthew Husband. 2017. Interference effects across the at-issue/not-at-issue divide: Agreement and NPI licensing. The 30th Annual CUNY Human Sentence Processing, MIT:
Cambridge, MA. March 30–April 1 (poster presentation).

Oberauer, Klaus & Reinhold Kliegl. 2006. A formal model of capacity limits in working memory. *Journal of Memory and Language* 55(4). 601–626. https://doi.org/10.1016/j.jml.2006.08.009.

Oberauer, Klaus & Elke B. Lange. 2008. Interference in verbal working memory: Distinguishing similarity-based confusion, feature overwriting, and feature migration. *Journal of Memory and Language* 58(3). 730–745. https://doi.org/10.1016/j.jml.2007.09.006.

Polanyi, Livia. 1988. A formal model of the structure of discourse. *Journal of Pragmatics* 12(5–6). 601–638. https://doi.org/10.1016/0378-2166(88)90050-1.

R Core Team. 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

Ramesh, Aditya, Prafulla Dhariwal, Alex Nichol, Casey Chu & Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. ArXiv manuscript. https://doi.org/10.48550/arXiv.2204.06125.

Rich, Stephanie & Matthew W. Wagers. 2020. Semantic similarity and temporal contiguity in subject-verb dependency processing. The 33rd Annual CUNY Human Sentence Processing, University of Massachusetts Amherst: Amherst, MA. March 19–21 (oral presentation).

Roberts, Craige. 2004. Context in dynamic interpretation. In Laurence R. Horn & Gregory Ward (eds.), *The handbook of pragmatics*, 197–220. Oxford: Blackwell.

Runner, Jeffrey T. & Kellan D. L. Head. 2014. What can visual world eye-tracking tell us about the binding theory. In Christopher Piñán (ed.), *Empirical issues in syntax and semantics*, vol. 10, 269–286. Paris: CSSP Colloque de Syntaxe et Sémantique à Paris.

Sekerina, Irina A., Luca Campanelli & Julie A. Van Dyke. 2016. Using the visual world paradigm to study retrieval interference in spoken language comprehension. *Frontiers in Psychology* 7. 873. https://doi.org/10.3389/fpsyg.2016.00873.

Slim, Mieke Sarah & Robert J. Hartsuiker. 2022. Moving visual world experiments online? A web-based replication of Dijkgraaf, Hartsuiker, and Duyck (2017) using PCIbex and WebGazer.js. *Behavior Research Methods* 1–19. https://doi.org/10.3758/s13428-022-01989-z.

Villata, Sandra, Whitney Tabor & Julie Franck. 2018. Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in Psychology* 9. 2. https://doi.org/10.3389/fpsyg.2018.00002.

Watson, Duane & Edward Gibson. 2004. The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes* 19(6). 713–755. https://doi.org/10.1080/01690960444000070.

Zehr, Jérémy & Florian Schwarz. 2018. PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832.