Title
Encoding discourse structure information during language comprehension: Evidence from web-based visual world paradigm experiments

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Kim, Sanghee J.
Xiang, Ming

Publication Date
2024

Encoding discourse structure information during language comprehension: Evidence from web-based visual world paradigm experiments

Sanghee J. Kim (sangheekim@uchicago.edu)
Department of Linguistics, 1115 E. 58th Street
Chicago, IL 60637 USA

Ming Xiang (mxiang@uchicago.edu)
Department of Linguistics, 1115 E. 58th Street
Chicago, IL 60637 USA

Abstract

This study explores the way discourse structure-related information is used during encoding linguistic representations, using the distinction between *main* and *subordinate* information as a case study. We use the two contrasting constructions: (a) "The singers$_{\text{MAIN}}$ who admired the violinists$_{\text{MAIN}}$ invited their mentors to the party"; and (b) "The singers$_{\text{MAIN}}$, who admired the violinists$_{\text{SUBORDINATE}}$, invited their mentors to the party." While both contain discourse-main information, (b) includes discourse-subordinate information in the clause ("who admired *the violinists*"). Importantly, *the singers* and *the violinists* are both plausible antecedents for *their*, but the overlap in discourse structure information between the two NPs differs: (a) has an overlap ({MAIN, MAIN}); (b) has no overlap ({MAIN, SUBORDINATE}). We found evidence through two web-based eye-tracking experiments using a visual world paradigm that the overlap in discourse structure leads to a competition effect between the two NPs, evidenced by smaller eye-gaze differences in (a) compared to (b). We also find that this competition effect manifests early, even before the relevant information needs to be retrieved, i.e., before pronoun resolution.

Keywords: sentence processing; cue-based retrieval theory; encoding; interference effect; competition; discourse structure; pronoun resolution

Background

An ongoing question in psycholinguistics addresses how speakers encode, maintain, and retrieve linguistic representations in working memory (WM) while comprehending sentences (Just & Carpenter, 1992; MacDonald, Just, & Carpenter, 1992; Caplan & Waters, 1999; Gibson, 1998, 2000; Warren & Gibson, 2002). Specifically, a large body of work has supported content-addressable cue-based direct-access memory retrieval mechanism during the memory retrieval process (McElree, 2000; McElree, Foraker, & Dyer, 2003; Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Van Dyke, 2007). From this standpoint, a successful resolution of linguistic dependencies hinges on retrieving the "correct" element in a dependency chain, with retrieval guided by cues matching specific features of the target to be retrieved.

Several features have been identified to be used for linguistic representations and retrieval cues (Parker, Shvartsman, & Van Dyke, 2017, for a review). These include morphosyntactic features related to number (Wagers, Lau, & Phillips, 2009, a.o), gender (Badecker & Straub, 2002, a.o), case (Fedorenko, Babyonyshev, & Gibson, 2004, a.o), and person (Gordon, Hendrick, & Johnson, 2001; Gordon, Hendrick, & Levine, 2002), or animacy (Van Dyke, 2007, a.o), and semantic plausibility (Van Dyke & McElree, 2006, a.o). Expanding on previous work, the present study examines whether discourse-related information is also used during encoding.

There are numerous interpretations and ways to understand discourse and types of discourse information, but in this work, we focus on the role of discourse structure, viewing discourse as a structured representation of discourse units and their interrelations (Asher & Lascarides, 2003; Polanyi, 1988). We particularly focus on the different degrees of importance linguistic content has in a discourse, namely information being part of MAIN discourse or SUBORDINATE (SUBORD.) discourse.[1] A critical contrast between the two types of discourse information is illustrated in (1).[2] The contrast in discourse structure information is made by the use of two types of relative clauses: restrictive relative clauses (RRCs) (1a) and appositive relative clauses (ARCs) (1b). ARCs are known as subordinate and side-commentary information (Koev, 2013) whereas content inside RRCs is part of main, primary information as it provides necessary information such as by restricting the noun phrase (NP) among possible sets (Partee, 1975).

(1) a. The waitress [who sat near *the girl(s)*] surprisingly {**was**/***were**} unhappy about all the noise.
b. The waitress, [who sat near *the girl(s)*], surprisingly {**was**/***were**} unhappy about all the noise.

Recent work has found evidence that such distinction in main vs. subordinate discourse affects the *retrieval* process in a way that linguistic content inside subordinate discourse does not lead to a retrieval interference effect. A standard number agreement attraction effect is expected in a sentential structure as in (1a), where in the case of the ungrammatical sentence (*were* as the matrix verb), reading times are reduced when a plural distractor noun (*the girls*) is used compared to a singular distractor noun (*the girl*) (Wagers et al., 2009). However, studies have found that when the distractor noun is positioned within a subordinate discourse structure (inside the ARC as in (1b)), the attraction effect disappears. The absence of a number agreement attraction effect

[1] This division does not necessarily imply a syntactic relationship. For instance, a syntactically subordinate clause can be part of a main discourse structure.

[2] (1a) was tested in Parker and An (2018); (1b) was examined in Kim and Xiang (2022).

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society*.

has been reported in similar constructions, all in which the distractor was part of subordinate discourse information (Ng & Husband, 2017; McInnerney & Atkinson, 2020). Other work has also shown that once the linguistic content that intervenes between long-distance dependencies becomes part of subordinate discourse information, the structure leads to reduced processing cost (Dillon, Clifton, & Frazier, 2014; Dillon, Clifton, Sloggett, & Frazier, 2017; Kroll & Wagers, 2019; Duff, Anand, Brasoveanu, & Rysling, 2023).

The Current Work

This study builds on previous work on retrieval and focuses on how discourse structure information is *encoded*. Two main questions are addressed: (i) how the distinction in discourse structure influences the representation of linguistic items during encoding, and (ii) when this distinction is realized. To address these questions, we design experiments such that discourse structure information is not necessarily a retrieval cue but is used during encoding, as illustrated in (2). *The singers* and *the violinists* are both possible antecedents for *their*. Crucially, even when *the violinists* is part of main discourse (2a) or subordinate discourse (2b), it serves as a plausible antecedent in both cases. As such, discourse structure information is encoded but not used as a cue to resolve the pronoun.

(2) Key experimental materials in the current study

a. Discourse information overlap condition
The singers$_{\text{MAIN}}$ who admired *the violinists*$_{\text{MAIN}}$ invited **their** mentors to the party.

b. Discourse information no-overlap condition
The singers$_{\text{MAIN}}$, who admired *the violinists*$_{\text{SUBORD.}}$, invited **their** mentors to the party.

Crucially, the conditions differ in the degree of the overlap in discourse structure information between the two NPs (*the singers*, *the violinists*): overlap in (2a) {MAIN, MAIN} but no overlap in (2b) {MAIN, SUBORDINATE}. Previous studies have shown that overlap of features during encoding affects processing (Hofmeister & Vasishth, 2014; Sekerina, Campanelli, & Van Dyke, 2016; Villata, Tabor, & Franck, 2018), leading to competition for features and potential degradation in linguistic representations in WM (Nairne, 1990, 2002; Oberauer & Kliegl, 2006; Oberauer & Lange, 2008). The competition effect typically results in longer reading times (Gordon et al., 2001, a.o) or fewer eye gazes on possible targets (Sekerina et al., 2016, a.o).

On the Competition Effect

Following Sekerina et al. (2016), we use eye-tracking experiments in a visual world paradigm to explore the competition effect arising from feature overlap. Sekerina et al. (2016) employed a similar paradigm to investigate competition during retrieval, predicting that similar distractors would lead to difficulty in integrating previous linguistic items, resulting in reduced eye gaze toward the target. In their study, four pictures were displayed, with all four NPs as potential targets in the *interference* condition ("It was the {button/key/pen/earring} that the maid who returned from vacation **spotted** in the early morning") and only one possible target in the *non-interference* condition ("It was the {button/key/pen/earring} that the maid who returned from vacation **sewed** in the early morning")—only *buttons* can be *sewed*. They observed a significant interference effect, with fewer looks to the target in the interference condition due to the presence of competitors, compared to the non-interference condition.

Predictions Making use of the contrast shown in (2), we examine the eye-gaze difference between the two possible NP antecedents. Previous work has shown that when there is competition between two representations, there will be smaller eye-gaze differences between visual images associated with the competing representations (Runner & Head, 2014; Sekerina et al., 2016). This is possibly due to the fact that people's visual attention is drawn to the target image and its competitor. Hence, if the overlap in discourse structure information leads to competition between the two NP antecedents, we predict the eye-gaze differences between the two NPs will be (2a) (overlap) $<$ (2b) (no-overlap).

On the Time Window of the Effect

For interference effects driven by feature similarities between representations, previous studies have mostly found the effect during memory retrieval. For instance, Sekerina et al. (2016) (see above) found the competition effect at the time window when the encoded items needed to be retrieved (at the main verb, e.g., *spotted*, *sewed*). Some studies have found the effect prior to the exact retrieval site, but nonetheless close to it. For example, Rich and Wagers (2020) compared cases where the relevant NPs share semantic features to different extent: *high semantic overlap* ("*The knife* that *the sword* was placed near had been recently **sharpened**") and *low semantic overlap* ("*The knife* that *the shirt* was placed near had been recently **sharpened**"). The *high overlap* condition led to longer reading times, with the competition effect appearing around *had been recently*, before the relevant semantic information needed to be retrieved and integrated at the verb *sharpened*.

Time windows for analysis We examine two windows for the analysis. The **first window** includes the RC, from the onset of *who* until the offset of the second NP (*the singers*). The **second window** begins from the matrix verb (*invited*) and includes the pronoun (*their*) and the head noun (*mentors*).

Predictions It is possible that the competition effect, indexed by the gaze differences on the two NP antecedent images, could arise during the first window, after the second NP was encountered (*the violinists*). This would be indicative of an encoding interference effect due to feature overlap between the two encoded NP representations. It is also possible that the effect could manifest only when retrieval is initiated, and we would expect the effect during the second window, when the verb or the pronoun is encountered. We examine these two competing possibilities.

Experiment 1

Methods

Participants We recruited 63 native speakers of American English (age mean=30.36; range: 19–40) through Prolific. The experiment took approximately 35 minutes. The subjects were compensated 8.00 USD for their participation.

Design and Materials We used a fully crossed 2x2 design, varying Clause (RRC vs. ARC) and Modifier-length (Short vs. Long) (Table 1). In the RRC structure, both NPs were part of MAIN information. In the ARC structure, the first NP was part of the MAIN while the second NP was part of the SUBORDINATE discourse structure. The Modifier-length condition was included to offer more time to resolve the pronoun; the Long condition had a modifier (e.g., *musical*) before the head noun. Different vocation names were used for the NPs. All embedded RCs were in a subject-extracted structure. Only *their* was used as the pronoun. The study involved 32 critical and 20 filler items.

Table 1: An example item in Experiment 1. Commas were realized as pauses. The Long condition had a modifier before the noun (below in parentheses, e.g., *musical*). NP1 = *the violinists*; NP2 = *the singers*.

Clause	Modifier	Auditory stimuli	Discourse information: {NP1, NP2}
RRC	Short/ Long	*The violinists* who admired *the singers* invited **their** (musical) mentors to the party.	{MAIN, MAIN}
ARC	Short/ Long	*The violinists*, who admired *the singers*, invited **their** (musical) mentors to the party.	{MAIN, SUBORD.}

Auditory stimuli were generated using the Google Text-to-Speech `gtts` Python library. Pauses were added before and after the RC boundary in the ARC condition (mean = 0.327 seconds), aligning with previously reported intonational patterns of ARC structures (Dehé & Kavalova, 2007; Dehé, 2014; Watson & Gibson, 2004). No pauses were introduced at the RC boundary in the RRC condition. Each target auditory stimulus was around 7.055 seconds.

Visual stimuli were produced with DALL-E, an AI system generating images given an input with natural language. These stimuli were created with black-and-white line drawings on a white background. Each trial included four images: two *targets* (e.g., *violinists*, *singers*), and two *distractors* (e.g., *detectives*, *teachers*) (Figure 1). The image of the head noun (e.g., *mentors*) was not presented on the screen.

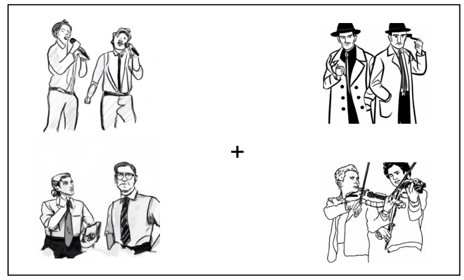


Figure 1: An example of the screen of the target trial. Clockwise from the top-left: *the singers* (target), *the detectives* (distractor), *the violinists* (target), *the teachers* (distractor).

Procedure The web-based visual world paradigm eye-tracking experiment was conducted using PCIbex (Zehr & Schwarz, 2018). Calibration involved a button-clicking task with 9 dots, and participants had three attempts to meet a precision threshold score of 60. During trials, the calibration precision threshold was 50, and participants had 2 attempts to meet this threshold. Trials began with a fixation cross, followed by a 2200 ms display of four pictures (2 targets and 2 distractors in a randomized order), after which eye movement recording started. The recording terminated 500 ms after the auditory stimulus offset. An antecedent selection task followed, asking participants to click on the picture corresponding to the referent (e.g., "Whose mentors were invited?").

Analysis We set an antecedent selection accuracy threshold of 85% to exclude participants who chose a distractor NP more than 15% of the time, yet all participants passed the threshold. We removed data points with incorrect image selection or eye gaze that fell outside quadrants. Following the approach in Slim and Hartsuiker (2022), the remaining data were aggregated into 100 ms time bins. Data points recorded 200 ms post the onset of auditory stimuli were analyzed, considering oculomotor delay.

We conducted a cluster-based permutation analysis (CPA) (Maris & Oostenveld, 2007; Ito & Knoeferle, 2023) to examine adjacent time bins that showed significant eye-gaze differences between the two target antecedents ($p < 0.05$). The analysis was employed through the `permutes` package in R (Voeten, 2023). The eye-gaze difference in the original dataset was examined using mixed-effects linear regression models in R (Bates, Mächler, Bolker, & Walker, 2015), with fixed effects including Clause (RRC vs. ARC), Modifier Length (Short vs. Long), and their interaction, along with a by-subject random effect.[3] Subsequently, the data were randomly permuted 1000 times to establish a null hypothesis distribution. Finally, each cluster within the original dataset was compared against this null hypothesis distribution.

The **first window** included the onset of *who* until the offset of *the singers*, without including the pauses. The **second win-**

[3]Sum coding for the fixed effects: RRC = -0.5 and ARC = 0.5; Short = -0.5 and Long = 0.5.

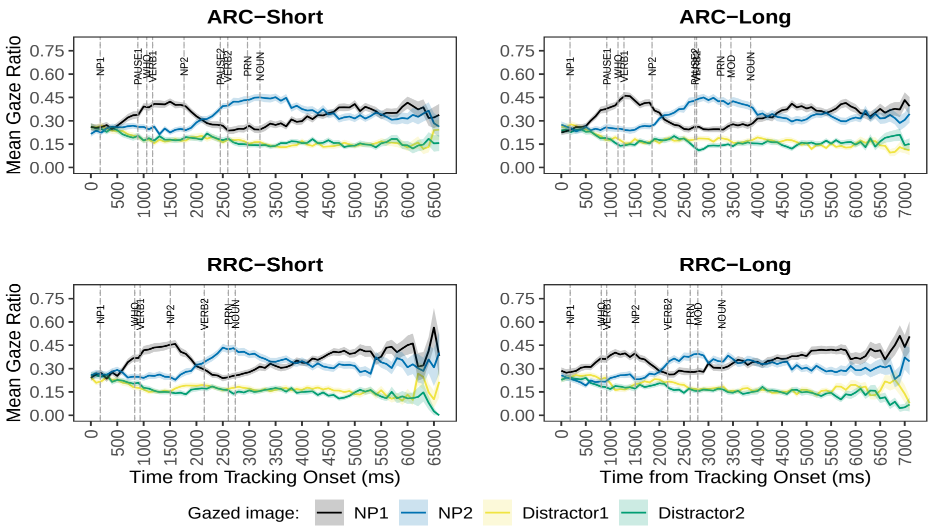


Figure 3: Eye-gaze ratio in Experiment 1. The labels above the vertical lines denote the average onset of each word across items. NP1 = *the violinists*; WHO = *who*; VERB1 = *admired*; NP2 = *the singers*; VERB2 = *invited*; PRN (pronoun) = *their*; MOD (modifier) = *musical*; NOUN = *mentors*; PAUSE1/2 = pauses inserted in the ARC condition in "The violinists, who admired the singers, invited their (musical) mentors to the party." Shaded ribbons indicate the standard errors of the mean.

dow included the matrix verb (*invited*) until the offset of the noun (*mentors*) in the Short condition. In the Long condition, the same duration was analyzed in the Long condition—it included the matrix verb, pronoun, the modifier (*musical*), and the beginning of the noun.

Results

Antecedent Selection NP1 was much preferred as the antecedent (∼65%) than NP2 (∼35%). This pattern was the same irrespective of Clause or Modifier-length (Figure 2). The distractors were rarely chosen.

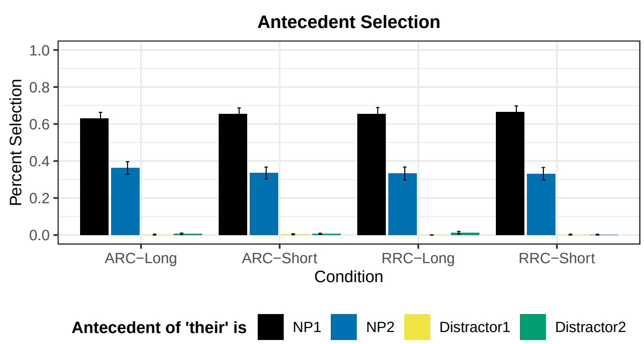
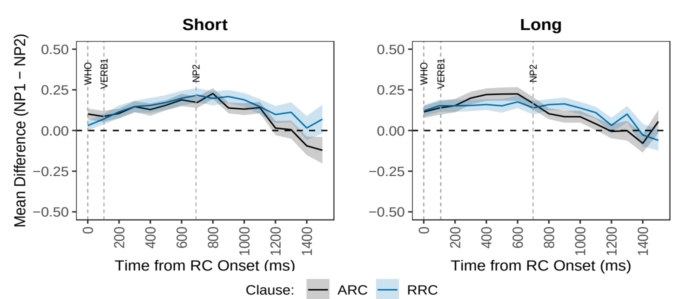


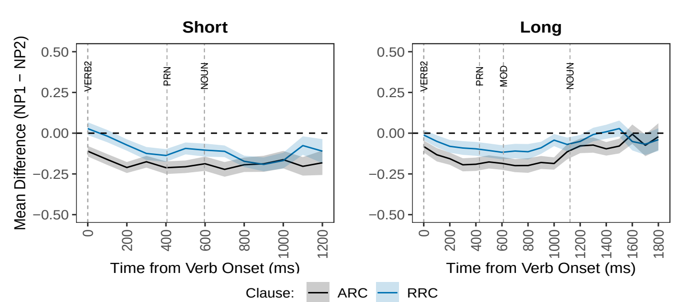
Figure 2: Antecedent preference in Experiment 1.

Eye-gaze Ratio Figure 3 shows the mean eye-gaze proportion on the images at each time bin.

Eye-gaze Differences The mean eye-gaze difference between NP1 and NP2 in the **first window** is shown in Figure 4a. The CPA suggested that no statistically significant clusters were found ($ps > 0.05$). In the **second window**, we found a significant cluster (0 ms–800 ms) for the Clause effect ($p < 0.001$). This suggests that the difference (ARC > RRC) started from the verb onset (Figure 4b). No significant clusters were found for other effects.



(a) First window. The X-axis is time-locked to the "WHO" onset.



(b) Second window. The X-axis is time-locked to the "VERB2" onset.

Figure 4: Eye-gaze ratio difference between NP1 and NP2 in the two windows of interest in Experiment 1. Shaded ribbons indicate the standard errors of the mean. The labels on the vertical lines are the same as those in Figure 3.

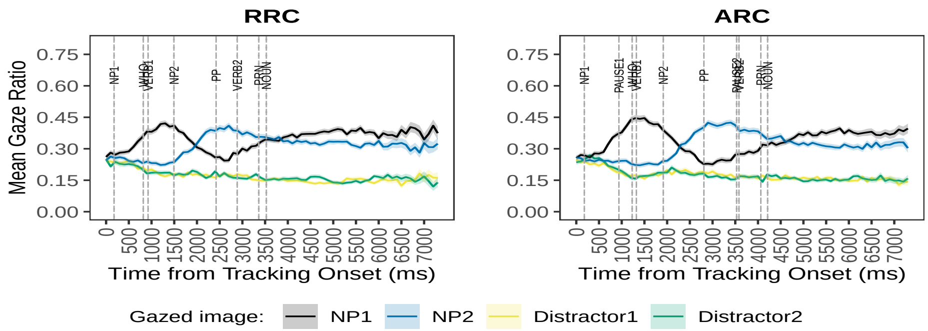


Figure 5: Eye-gaze ratio in Experiment 2. The labels above the vertical lines denote the average onset of each word across items. NP1 = *the violinists*; WHO = *who*; VERB1 = *admired*; NP2 = *the singers*; PP = *without reason*; VERB2 = *invited*; PRN (pronoun) = *their*; NOUN = *mentors*; PAUSE1/2 = pauses inserted in the ARC condition in "The violinists, who admired the singers without reason, invited their mentors to the party." Shaded ribbons indicate the standard errors of the mean.

Discussion

Participants showed a preference for the first NP as the antecedent of *their* in the pronoun-antecedent selection task across conditions. This was also reflected in the eye-gaze pattern, where the eye gaze towards the first NP rose around 1000 ms after the onset of the head noun. This preference for the first NP underscores a general inclination to choose the subject of the same clause when resolving ambiguous antecedents (Arnold, 2010, a.o.). This suggests that the observed eye-gaze results cannot be solely attributed to subjects' preferences for a specific antecedent.

The analysis of eye-gaze results revealed a competition effect due to overlap in discourse structure information, leading to smaller eye-gaze ratio differences between possible antecedents (difference: RRC < ARC). The effect began only after the onset of the matrix verb. This could reflect a reactivation of the encoded linguistic items at the verb retrieval site, even though discourse structure information is not a retrieval cue. Alternatively, the effect could be a delayed effect of encoding interference at the retrieval site, as shown similarly in Sekerina et al. (2016). However, since the offset of the second NP and the rightmost boundary of the RC overlap, it is difficult to tease apart these two possibilities.

To address this confound, in Experiment 2, we include a buffer window before the RC boundary to offer extra time to encode the distinct discourse structure information. We added a prepositional phrase (PP) after the second NP before the matrix verb. If the competition effect is realized as early as during the encoding stage, we expect to see the eye-gaze differences before the second window.

Experiment 2

Methods

We recruited 108 native speakers of American English through Prolific (age mean=30.37; range: 18–40). The same set of items as in Experiment 1 was utilized, with some modifications(Table 2). A PP (e.g., *without reason*) was added at the end of RC, and the noun modifier was removed given the absence of its effect in Experiment 1. The procedure was the same as in Experiment 1, while the between-trial calibration threshold score was set at 30.

Table 2: An example item of Experiment 2. Commas were realized as pauses. NP1 = *the violinists*; NP2 = *the singers*.

Clause	Auditory stimuli	Discourse information: {NP1, NP2}
RRC	*The violinists* who admired *the singers* without reason invited **their** mentors to the party.	{MAIN, MAIN}
ARC	*The violinists*, who admired *the singers* without reason, invited **their** mentors to the party.	{MAIN, SUBORD.}

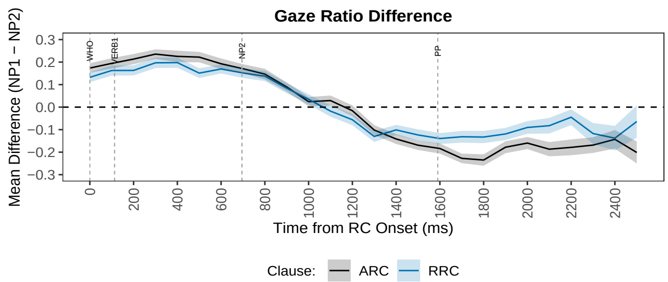
The same data filtering process as in Experiment 1 was used. Since the calibration threshold was set low in this experiment, we additionally removed data from the participants whose mean calibration score across trials was 50 or below (13.9% of data removal). The same CPA was used along with linear mixed-effects random effects models; a fixed effect of Clause (RRC vs. ARC) and a by-subject random effect were included in the models. The **first window** included the onset of *who* to the offset of the PP (e.g., *reason*). **The second window** started from the onset of the matrix verb (*invited*), pronoun (*their*), until the offset of the noun (*mentors*).

Results

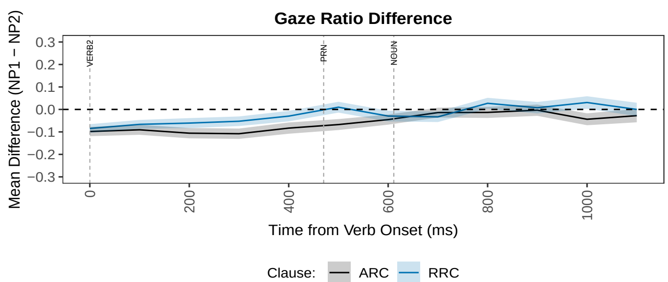
Antecedent Selection NP1 was chosen as the antecedent more frequently (∼65%) than NP2 (∼35%) irrespective of Clause type.[4] Distractors were rarely selected.

Eye-gaze Ratio Figure 5 shows the mean proportion of eye gaze, averaged across participants.

[4]Figure is not included due to space constraint.



(a) First window. The X-axis is time-locked to the "WHO" onset.



(b) Second window. The X-axis is time-locked to the "VERB2" onset.

Figure 6: Eye-gaze ratio difference between NP1 and NP2 in the two windows of interest in Experiment 2. Shaded ribbons indicate the standard errors of the mean. The labels on the vertical lines are the same as those in Figure 5.

Eye-gaze Differences In the **first window** (Figure 6a), we found a significant cluster in 1700 ms–1900 ms and 2100 ms–2300 ms after the onset of the RC ($p < 0.01$), which suggests that the effect (i.e., ARC > RRC in eye-gaze difference) started around the onset of the PP. In the **second window** (Figure 6b), no significant cluster groups were identified ($ps > 0.1$). The ARC vs. RRC contrast did not lead to eye-gaze differences during retrieval.

Discussion

Replicating the finding in Experiment 1, the first NP was predominantly selected as the possible referent of the pronoun. This preference was uniform between clauses, again suggesting that eye-gaze difference cannot be explained solely by antecedent preference.

In line with the findings in Experiment 1, the RRC condition ({MAIN, MAIN} overlap) had smaller eye-gaze differences than the ARC condition ({MAIN, SUBORDINATE}). The purpose of Experiment 2 was to examine whether the eye-gaze difference we saw at the second window in Experiment 1 was a delayed effect from the end of the first window. In Experiment 2, we found the effect during the PP window, right after the encoding of the second NP, and before the retrieval site. This indicates that the results in Experiment 1 were a reflection of a competition effect due to the overlap in discourse structure information. Additionally, it suggests that the effect can be realized as early as before retrieval is initiated.

General Discussion

We return to the two main goals of the current study. First, we examined the way discourse structure information (MAIN or SUBORDINATE) is encoded and affects processing. Our findings suggest that when the encoded representations share the same discourse structure information (as in the RRC condition), it results in a competition effect, compared to when they have distinct discourse information (as in the ARC condition). This was shown by the more pronounced eye-gaze differences between the two encoded NPs in the ARC condition compared to the RRC condition.

Secondly, we addressed the issue of how early the effect of competition due to discourse structure overlap is realized. In Experiment 1, the eye-gaze difference was found only in the second window, right after the onset of the first word, the matrix verb; this effect lasted around 800 ms. This seemed to suggest that the overlap effect is realized when retrieval of prior linguistic representations is triggered. However, when more time was given to encode the second NP before the end of the RC in Experiment 2, the eye-gaze difference started as early as the first window of interest, specifically after the second NP and before the termination of the RC. The effect lasted around 600 ms (with a gap of around 200 ms of time window in between). This suggests that the encoding interference effect can be realized even before retrieval is triggered. The effect in Experiment 1 could be interpreted as the same effect we observed in Experiment 2 but simply overlapped with the onset of the retrieval site.

The results collectively suggest that the competition effect due to overlap in linguistic information can be manifested as early as during encoding. The finding of the *early* realization of the competition effect is in line with other previous work that reported an encoding interference effect even before retrieval (Acheson & MacDonald, 2011; Kush, Johns, & Van Dyke, 2015; Rich & Wagers, 2020).

Is it about discourse?

We suggested that discourse structure information is used during the encoding process, and that its overlap can lead to a competition effect. Yet, rather than the discourse structure effect, one could propose an alternative account that the ARC condition simply "benefited" from the extra time given before the RC. There was an additional time of 300 ms with the pause inserted before the onset of *who* in the ARC condition. As earlier studies show, a longer time to encode can make the linguistic representations stronger and more salient in memory (Karimi, Diaz, & Wittenberg, 2020, a.o). Hence, the second NP in the current study, with the extra encoding time prior to its encoding, could have had a higher activation in memory—not due to the absence of overlap in discourse information but due to longer encoding time. A follow-up experiment that uses filler words and disfluency such as "umm" or "well", for example, instead of a pause before the RC, would be useful to tease apart the effect of discourse structure information vs. extra time to encode.

Acknowledgments

The project has been supported by the generous NSF DDRI Grant (BCS-2214437). It has benefited from discussions with the members of the University of Chicago Language Processing Lab, and the audience at the UC Santa Cruz S/Lab, LSA Annual Meeting 2024, and the UC Davis Ferreira Lab. We thank Jennifer Arnold, Matt Wagers, and Brian Dillon for their feedback at the initial stages of this project. We are grateful to Jérémy Zehr, Sarah Hye-yeon Lee, and Monica Do for their help in conducting the web-based eyetracking experiment, and to Thomas Sostarics for his advice in generating the auditory stimuli and conducting data analysis. We are thankful to the participants on Prolific who were patient to go through multiple rounds of the calibration process. Finally, we thank the anonymous reviewers for their insight and suggestions. All errors are our own.

References

Acheson, D. J., & MacDonald, M. C. (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.

Arnold, J. E. (2010). How speakers refer: The role of accessibility. *Language and Linguistics Compass*, *4*(4), 187–203.

Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge: Cambridge University Press.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 748.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*(1), 77–94.

Dehé, N. (2014). *Parentheticals in spoken English: The syntax-prosody relation*. Cambridge: Cambridge University Press.

Dehé, N., & Kavalova, Y. (2007). Parentheticals: An introduction. In N. Dehè & Y. Kavalova (Eds.), *Parentheticals* (Vol. 106, pp. 1–24). Amsterdam: John Benjamins.

Dillon, B., Clifton, C., Jr, & Frazier, L. (2014). Pushed aside: Parentheticals, memory and processing. *Language, Cognition and Neuroscience*, *29*(4), 483–498.

Dillon, B., Clifton, C., Jr, Sloggett, S., & Frazier, L. (2017). Appositives and their aftermath: Interference depends on at-issue vs. not-at-issue status. *Journal of Memory and Language*, *96*, 93–109.

Duff, J., Anand, P., Brasoveanu, A., & Rysling, A. (2023). Pragmatic representations and online comprehension: Lessons from direct discourse and causal adjuncts. *Glossa Psycholinguistics*, *2*(1).

Fedorenko, E., Babyonyshev, M., & Gibson, E. (2004). The nature of case interference in on-line sentence processing in Russian. In K. Moulton & M. Wolf (Eds.), *Proceedings of the 34 Annual Meeting of the North East Linguistic Society* (Vol. 34, pp. 215–226). Amherst, MA: GLSA.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. P. Marantz, Y. Miyashita, & O. Wayne (Eds.), *Image, language, brain* (pp. 95–126). Cambridge, MA: MIT Press.

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1411.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, *13*(5), 425–430.

Hofmeister, P., & Vasishth, S. (2014). Distinctiveness and encoding effects in online sentence comprehension. *Frontiers in Psychology*, *5*, 1237.

Ito, A., & Knoeferle, P. (2023). Analysing data from the psycholinguistic visual-world paradigm: Comparison of different analysis methods. *Behavior Research Methods*, *55*(7), 3461–3493.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122—149.

Karimi, H., Diaz, M., & Wittenberg, E. (2020). Sheer time spent expecting or maintaining a representation facilitates subsequent retrieval during sentence processing. In S. Hartmann & U. Hahn (Eds.), *Proceedings of the Annual Conference of the Cognitive Science Society* (Vol. 42, pp. 2728–2734).

Kim, S. J., & Xiang, M. (2022). *Memory retrieval selectively targets different discourse units.* The 35th Annual Conference on Human Sentence Processing (HSP2022), University of California at Santa Cruz, CA. (March 24–26 (plenary talk))

Koev, T. (2013). *Apposition and the structure of discourse*. Unpublished doctoral dissertation, Rutgers University, New Brunswick, NJ.

Kroll, M., & Wagers, M. (2019). Working memory resource allocation is not modulated by clausal discourse status. *Unpublished manuscript, University of California, Santa Cruz.*

Kush, D., Johns, C. L., & Van Dyke, J. A. (2015). Identifying the role of phonology in sentence-level reading. *Journal of Memory and Language*, *79*, 18–29.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, *24*(1), 56–98.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research*, *29*(2), 111–123.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*(1), 67–91.

McInnerney, A., & Atkinson, E. (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, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*, 251–269.

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, *10*(5-6), 389–395.

Ng, A., & Husband, M. (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, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*(4), 601–626.

Oberauer, K., & Lange, E. B. (2008). Interference in verbal working memory: Distinguishing similarity-based confusion, feature overwriting, and feature migration. *Journal of Memory and Language*, *58*(3), 730–745.

Parker, D., & An, A. (2018). Not all phrases are equally attractive: Experimental evidence for selective agreement attraction effects. *Frontiers in Psychology*, *9*, 1566.

Parker, D., Shvartsman, M., & Van Dyke, J. A. (2017). The cue-based retrieval theory of sentence comprehension: New findings and new challenges. In T. Parodi, L. Escobar, & T. Vicenç (Eds.), *Language processing and disorders* (pp. 121–144). Newcastle: Cambridge Scholars Publishing.

Partee, B. (1975). Montague grammar and transformational grammar. *Linguistic Inquiry*, *6*(2), 203–300.

Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics*, *12*(5–6), 601–638.

Rich, S., & Wagers, M. W. (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))

Runner, J. T., & Head, K. D. (2014). What can visual world eye-tracking tell us about the binding theory. In C. Piñán (Ed.), *Empirical issues in syntax and semantics* (Vol. 10, pp. 269–286). Paris: CSSP—Colloque de Syntaxe et Sémantique à Paris.

Sekerina, I. A., Campanelli, L., & Van Dyke, J. A. (2016). Using the visual world paradigm to study retrieval interference in spoken language comprehension. *Frontiers in Psychology*, *7*, 873.

Slim, M. S., & Hartsuiker, R. J. (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.

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(2), 407.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, *55*(2), 157–166.

Villata, S., Tabor, W., & Franck, J. (2018). Encoding and retrieval interference in sentence comprehension: Evidence from agreement. *Frontiers in Psychology*, *9*, 2.

Voeten, C. C. (2023). permutes: Permutation tests for time series data [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=permutes` (R package version 2.8)

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206–237.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition*, *85*(1), 79–112.

Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, *19*(6), 713–755.

Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX).* doi: 10.17605/OSF.IO/MD832