

Introduction

- The difficulty of processing various types of relative clauses has been extensively studied in humans and is one of the main points of evaluation for my neural network model of sentence comprehension and production (Rohde, 2002, PhD. Thesis)
- Like humans in self-paced, moving window experiments, the model is constrained to process sentences sequentially with no review.
- The model answers fill-in-the-blank comprehension questions, but it cannot refer back to the sentence to do so.
- Thus, the appropriate human measure for comparison with the model should be fill-in-the-blank comprehension following single-pass exposure to the sentence.
- How well can we comprehend single relative clause sentences, and other types, following a single reading without subsequent rehearsal?

Problems with Earlier Work

- There have been a number of online reading studies of relative clauses, most recently Gibson et al. (in press).
- Some earlier work has failed to completely control for semantic differences across conditions.
- Many online experiments do not emphasize the offline measure and do not control it carefully.
- Most often, the offline measure consists of true/false statements.
- Unless they are well-designed, true/false statements can often be answered without a real thematic understanding of the sentence.
 - ▷ The violinist that flattered the sponsor insulted the conductor.
 - ▷ T/F: The violinist flattered the cellist.

A More Serious Problem

- The true/false statements used to test comprehension are often phrased actively, with the same word order as some, but not all, conditions.
- If the word order in the query statement matches that in the original sentence, participants could answer the question based on surface-level comparisons, rather than on a deeper thematic understanding.
- Read this sentence and, when you're done, close your eyes and recall how it sounded.
- Open your eyes now. You probably had the whole thing, or much of it, buffered in a short-term phonological store.
- Participants may be replaying sentences or parts of them from this store to help answer comprehension questions, either by reparsing or making surface comparisons.
- But the phonological buffer will not be very useful in comprehending fluent spoken dialogs...
- and the model doesn't have one.

Preventing Replay

- In order to assess true first-pass comprehension, we must prevent the reader from replaying the sentence while answering questions.
- This was done with a phonological masking task.
- Half of the subjects (Exp. A) performed normal self-paced reading with multiple-choice comprehension questions.
- The other half read the sentence the same way. But then, three random syllables appeared on the screen and were highlighted in a looping sequence (333ms/syllable).

dah poo kai
dah **poo** kai
dah poo **kai**
dah poo kai

- Participants began chanting the syllables in rhythm and continued while answering the comprehension questions.
- Presumably, the chanting will interfere with phonological replay. Due to attentional demands, it may interfere somewhat with other processing, but similarly across conditions, we hope.

Example Sentences

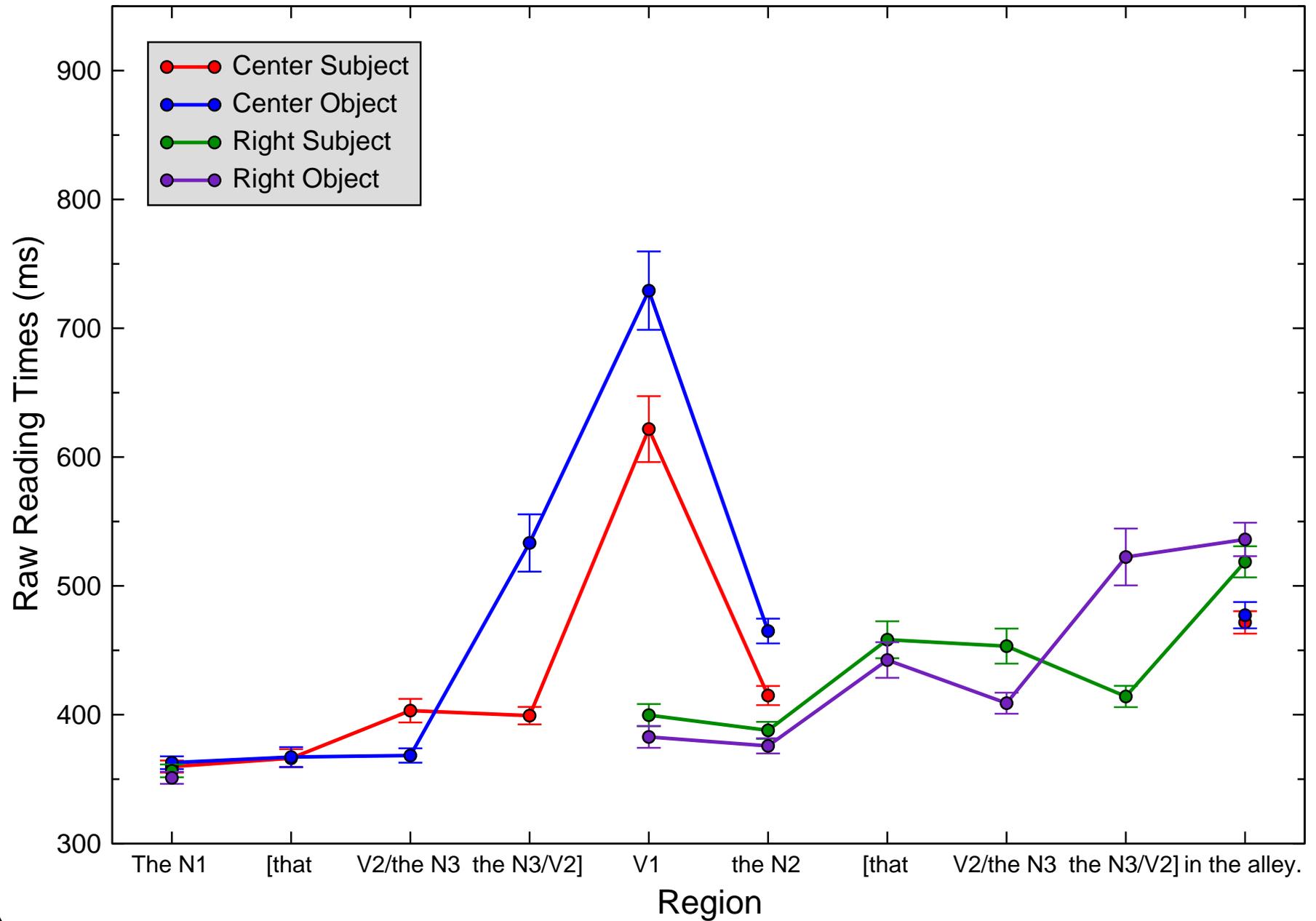
- CS: Center-embedded Subject-relative
 - ▷ The reporter [that attacked the senator] ignored the president at the meeting.
- CO: Center-embedded Object-relative
 - ▷ The reporter [that the senator attacked] ignored the president at the meeting.
- RS: Right-branching Subject-relative
 - ▷ The reporter ignored the president [that attacked the senator] at the meeting.
- RO: Right-branching Object-relative
 - ▷ The reporter ignored the president [that the senator attacked] at the meeting.
- For each condition and item, every possible arrangement (12) of the three nouns and two verbs were used, fully counter-balancing semantics.
 - ▷ The reporter [that the senator attacked] ignored the president at the meeting.
 - ▷ The senator [that the reporter attacked] ignored the president at the meeting.
 - ▷ The president [that the senator ignored] attacked the reporter at the meeting.
 - ▷ The president [that the reporter ignored] attacked the senator at the meeting.

⋮

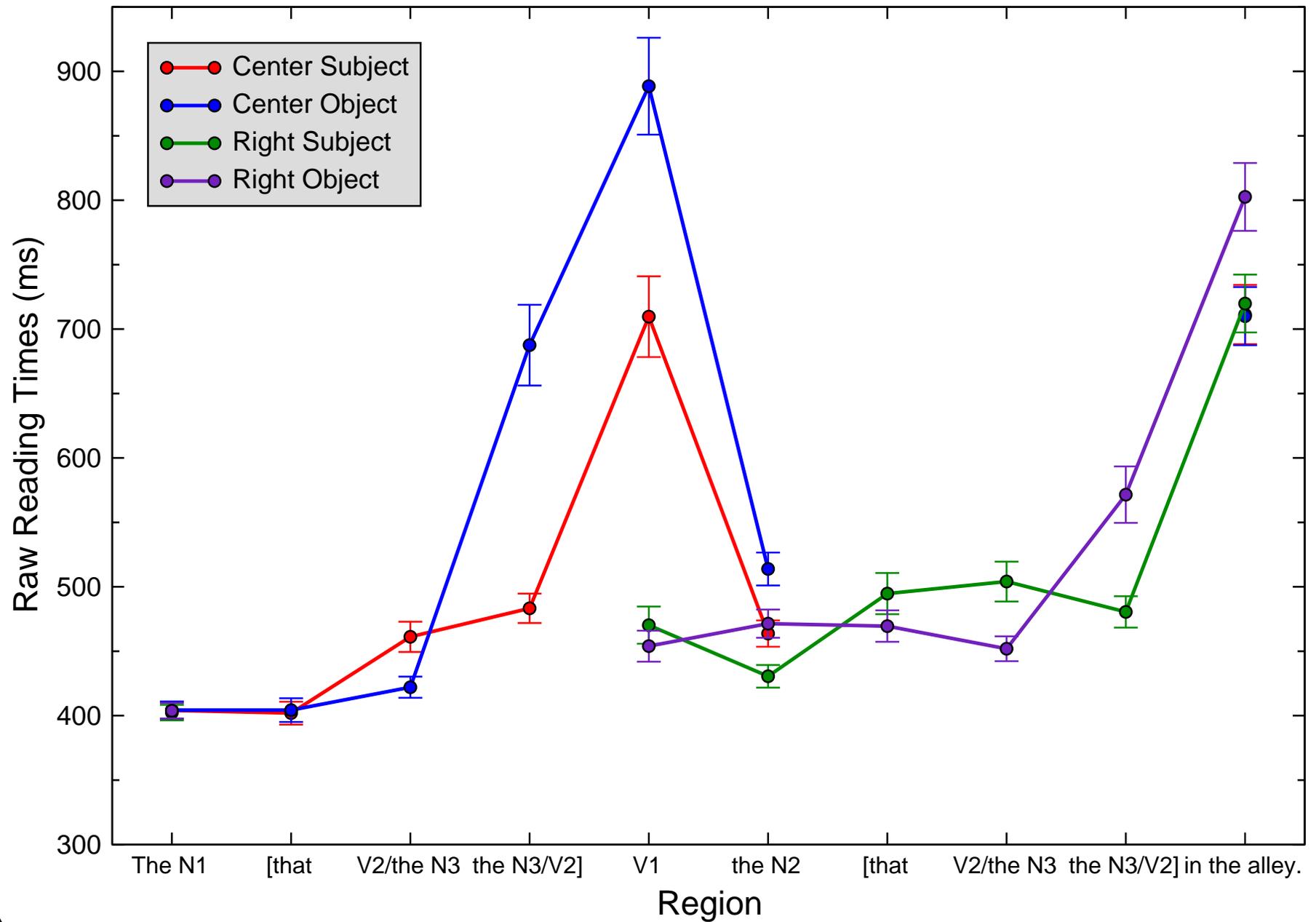
Methods

- Two versions: Exp. A had no phonological masking, Exp. B had masking. Identical otherwise.
- 40 experimental items. 60 fillers. 48 participants per experiment.
- 10 additional participants were replaced in Exp. A and 12 in Exp. B because of excessively bad comprehension or slow reading on the filler items or because they were deemed to be non-native speakers.
- Sentences were presented with masked, moving-window, self-paced reading.
- Two questions followed each experimental item; one or two followed each filler (overall average: 1.5).
- For each participant, a linear regression equation was computed to predict the reading time as a function of word length. This prediction was subtracted from the raw times to obtain residual reading times.

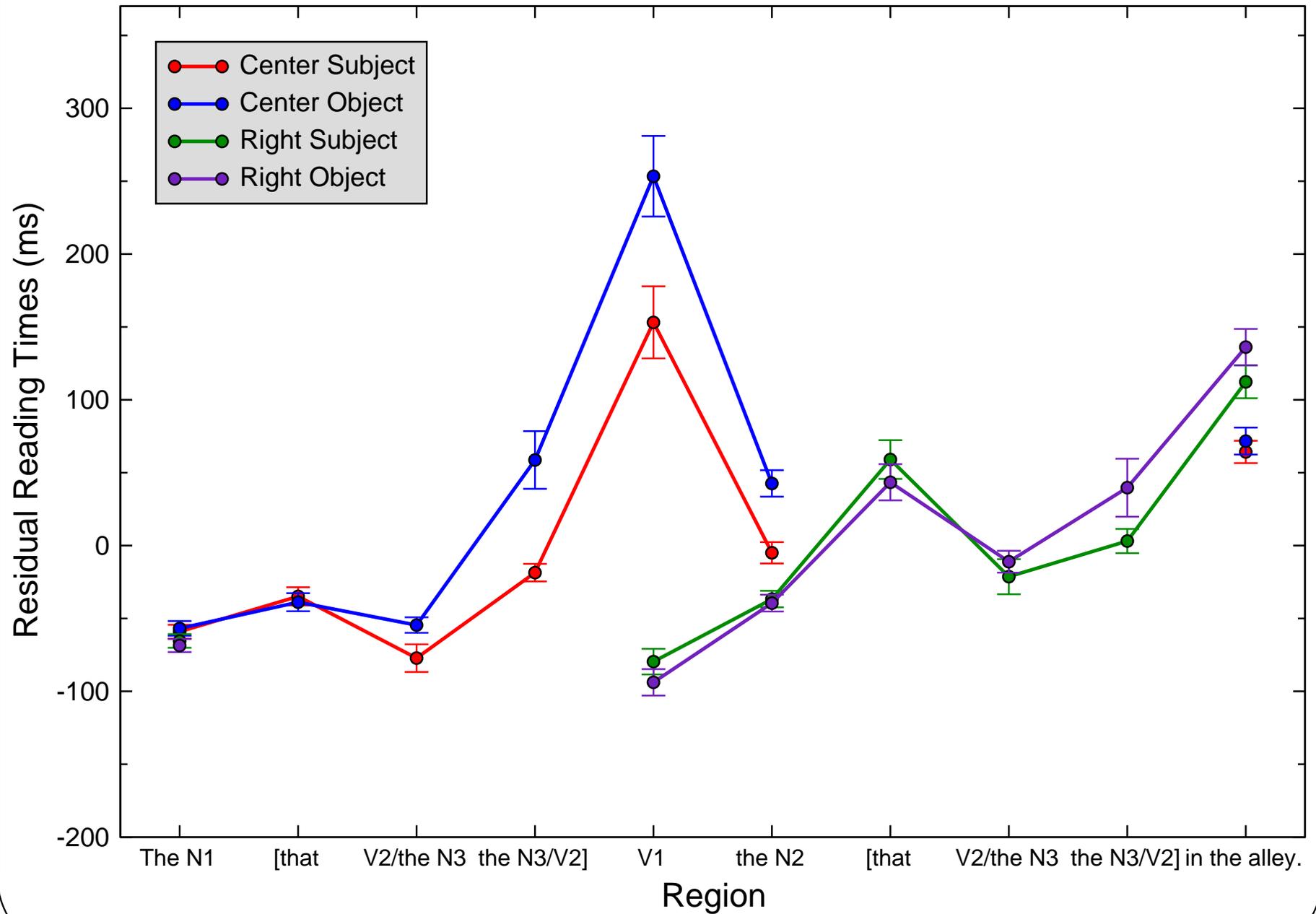
Exp. A: Raw Reading Times



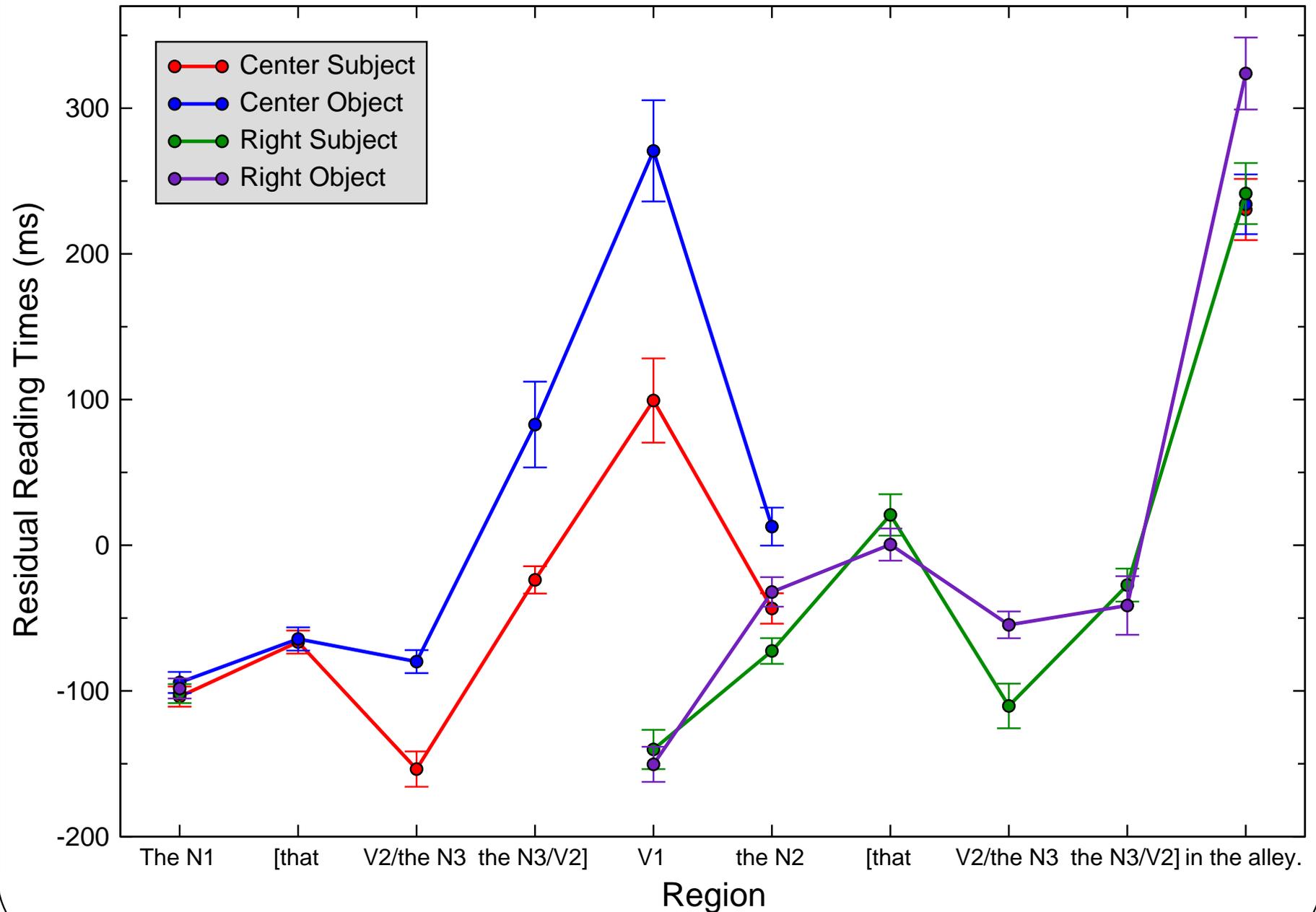
Exp. B: Raw Reading Times



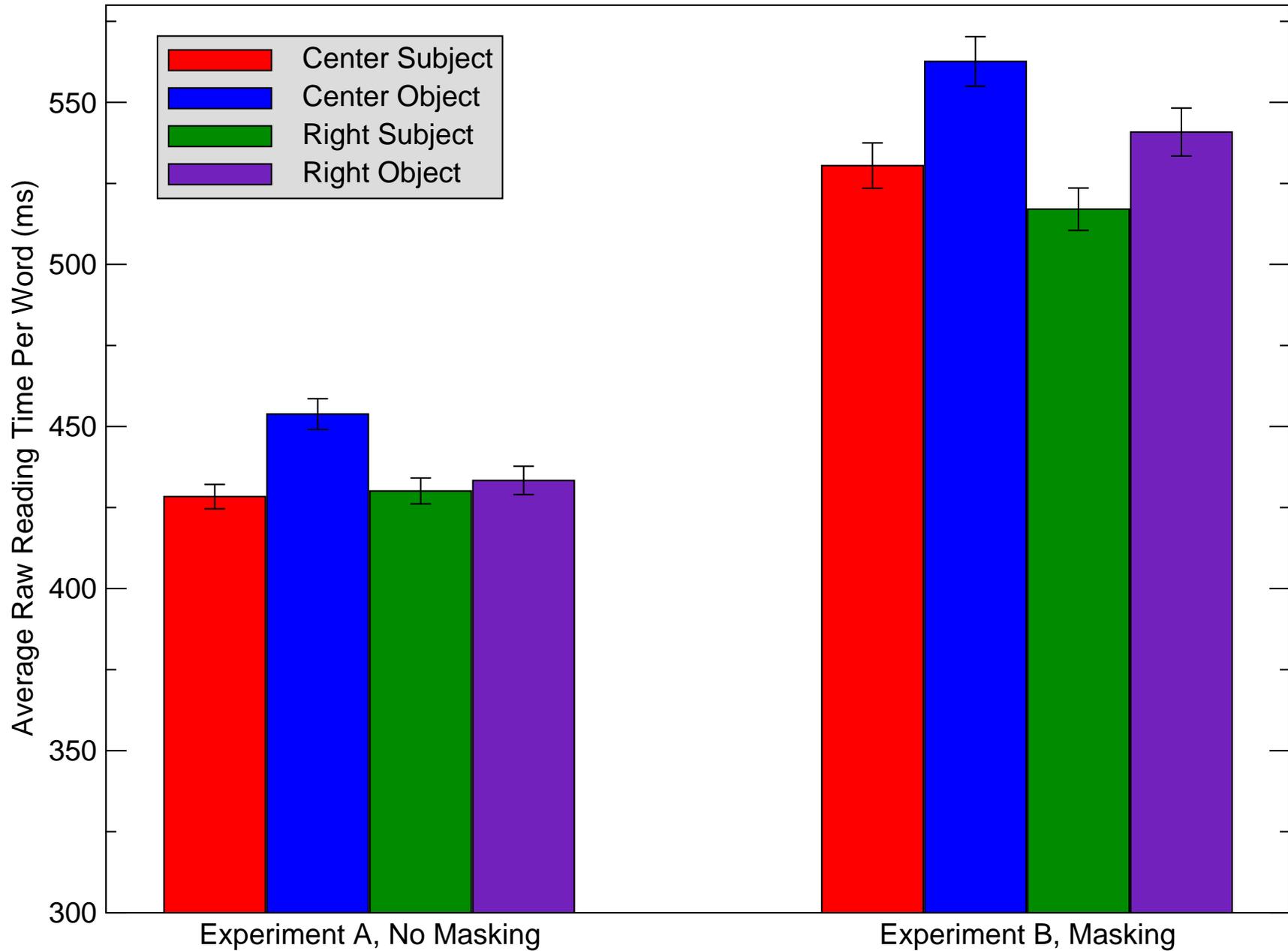
Exp. A: Residual Reading Times



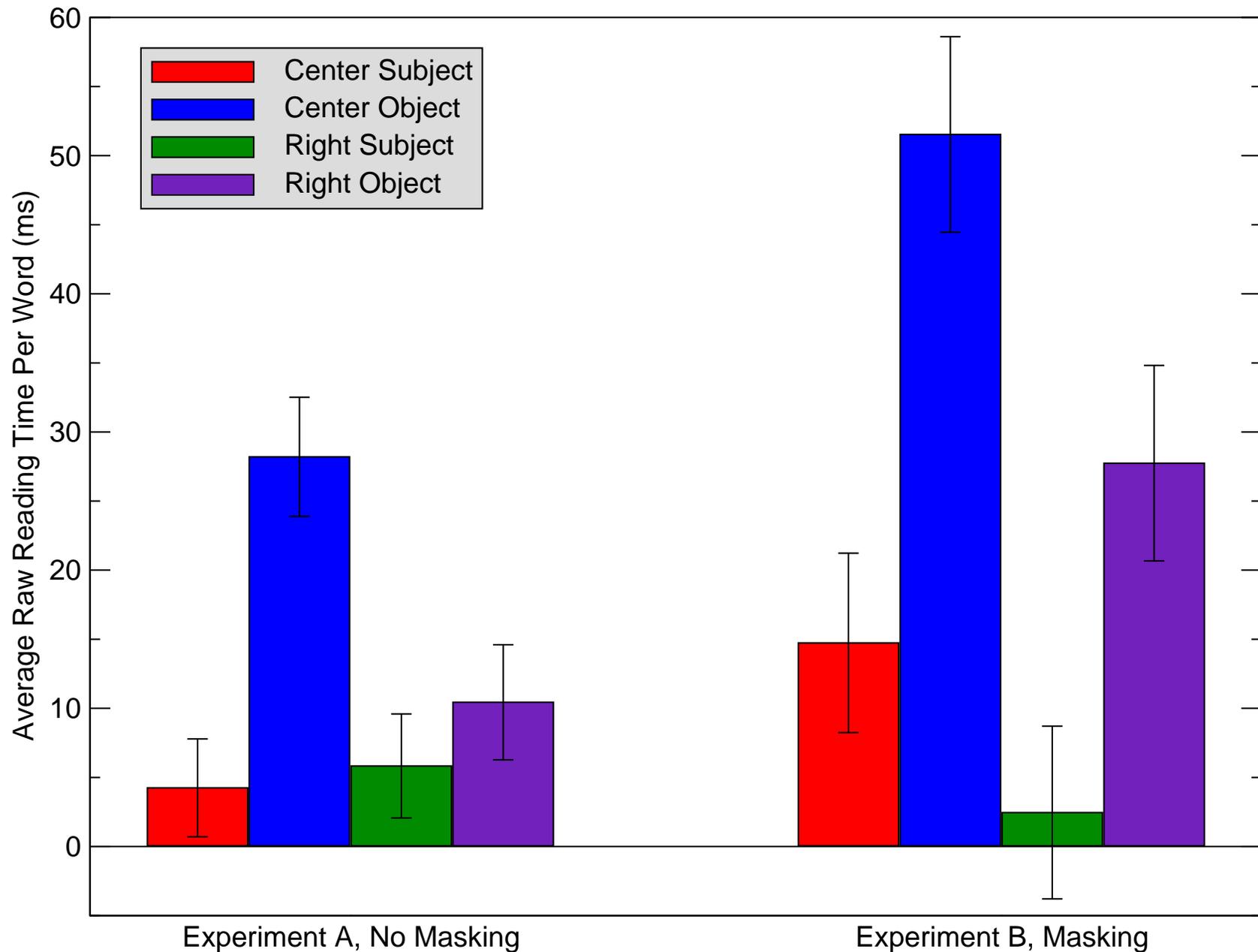
Exp. B: Residual Reading Times



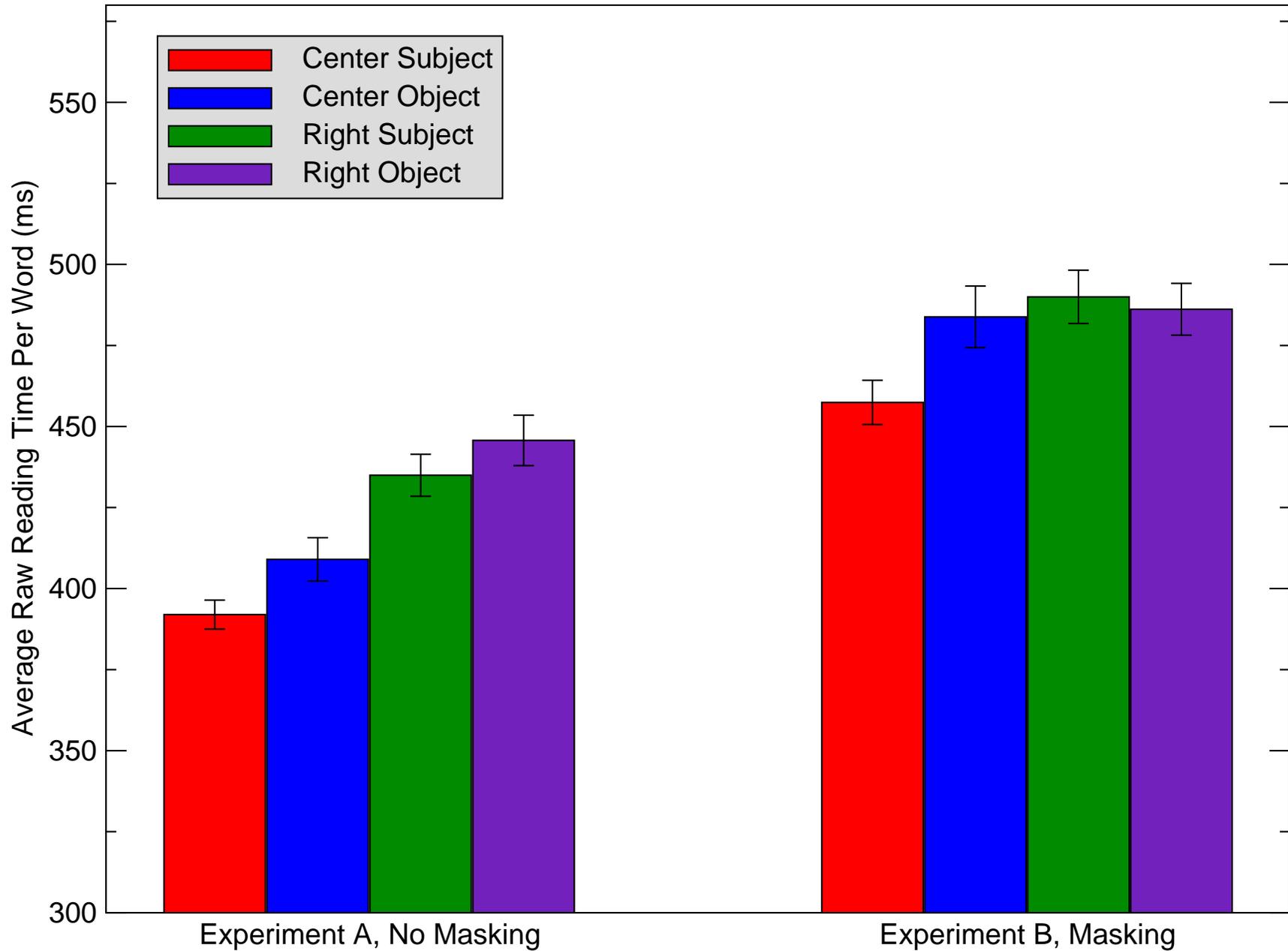
Whole Sentence: Raw Reading Times



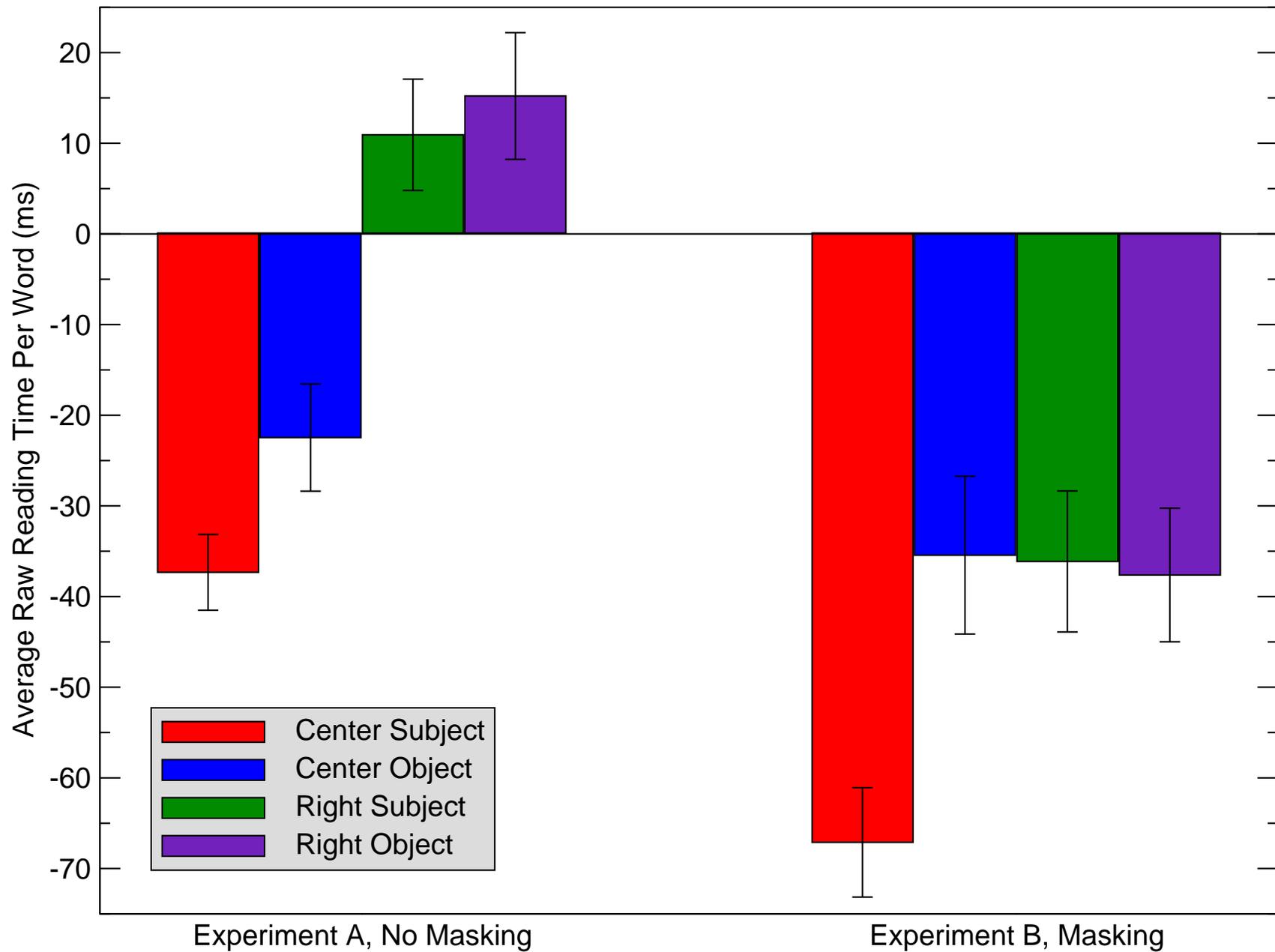
Whole Sentence: Residual Reading Times



Relative Clause: Raw Reading Times



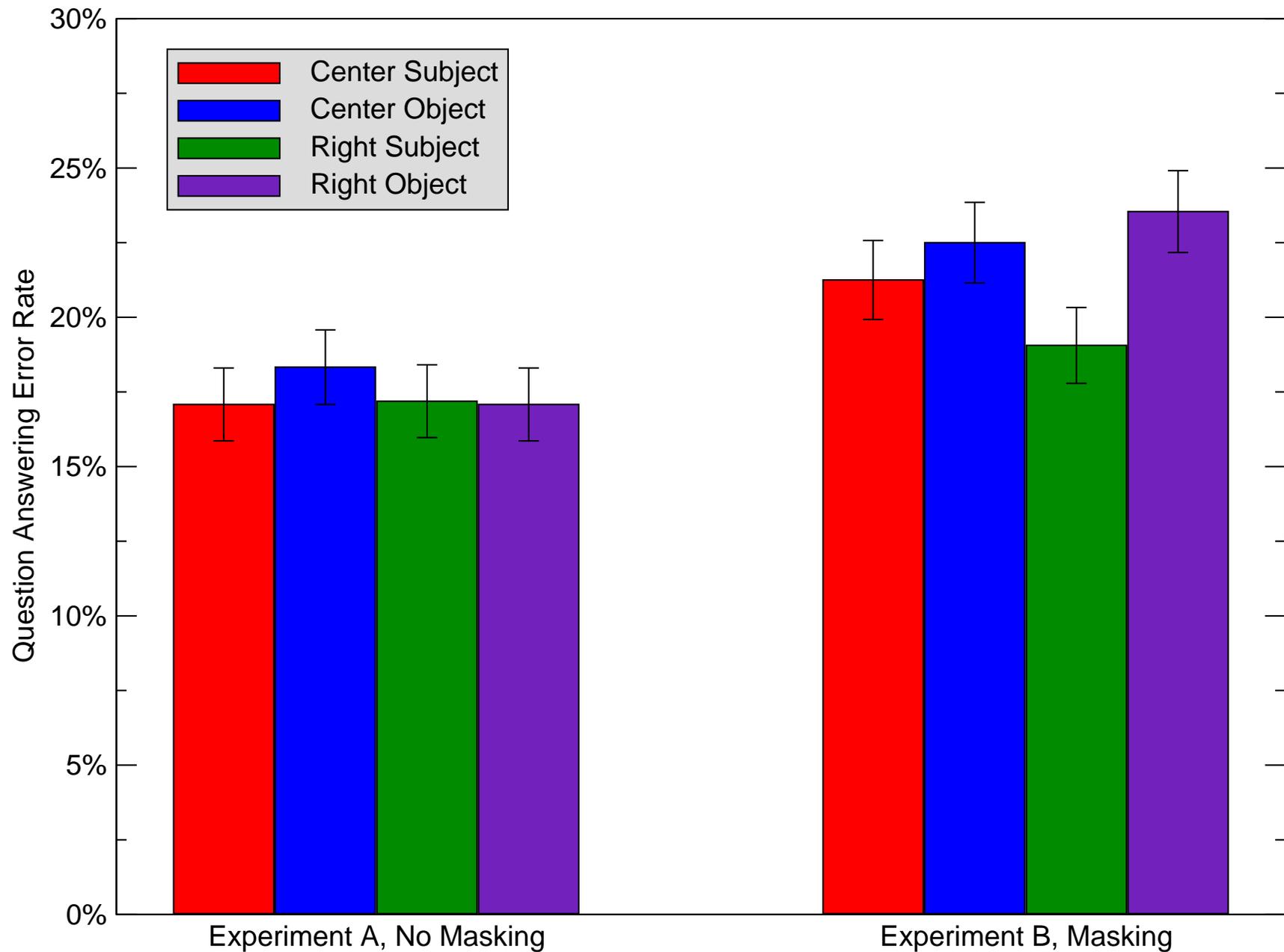
Relative Clause: Residual Reading Times



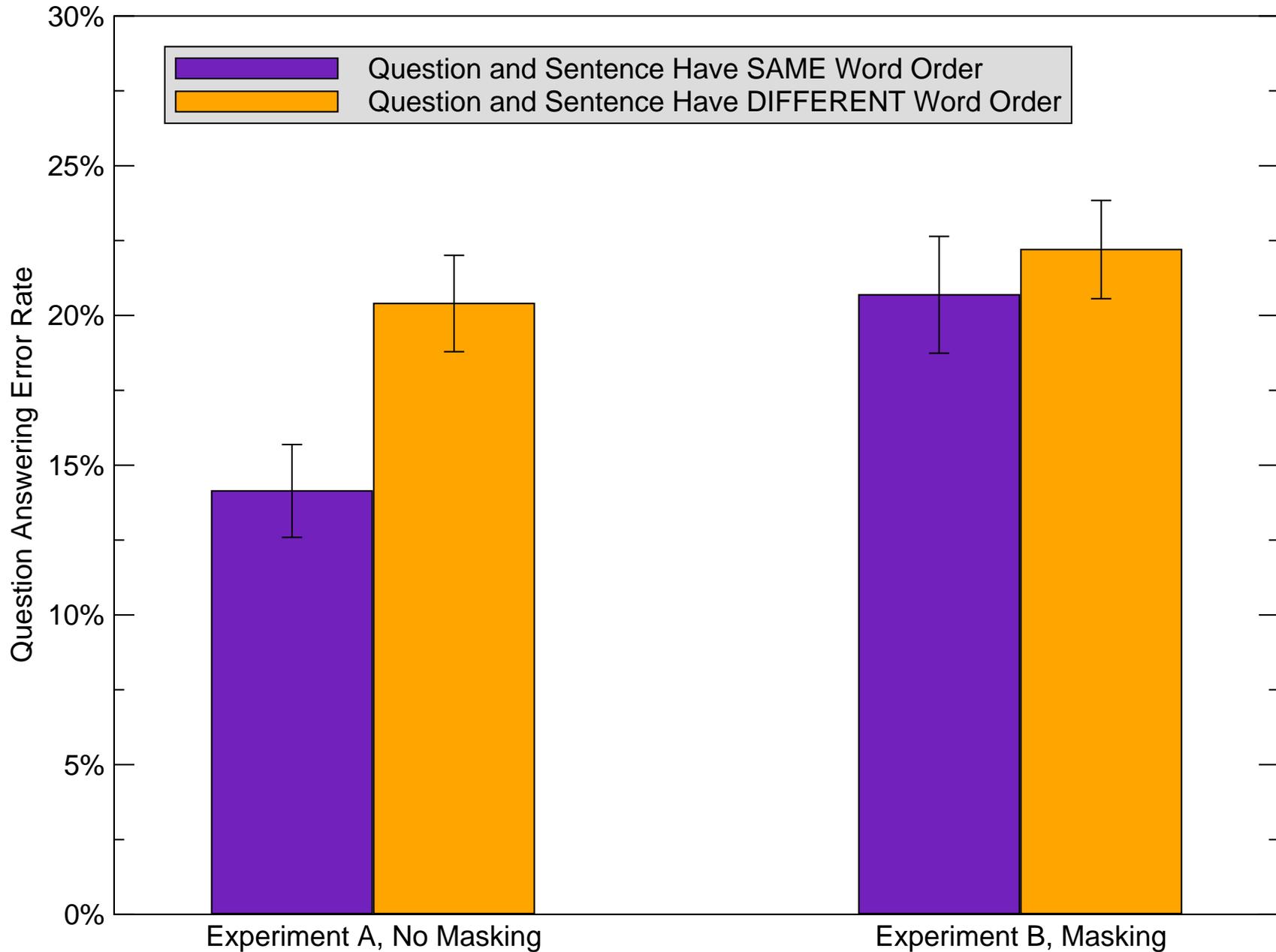
Reading Time Conclusions

- Readers are slower overall in Exp. B.
- The demands of the offline task cause a significant strategy shift—readers must actually try to comprehend the sentence on the first pass.
- Thus, reading times may fully reflect the necessary thematic processing.
- Wrap-up effects are enhanced, as are differences in the right-branching conditions, and the slow-down on the main verb after a CO.
- Whole sentence: In Exp. A, CO is slower than the others. In Exp. B, two main effects, no interaction, as in Gibson et al. (in press).
- Relative clause: In Exp. A, two main effects, no interaction. In Exp. B, CS is faster than the others.

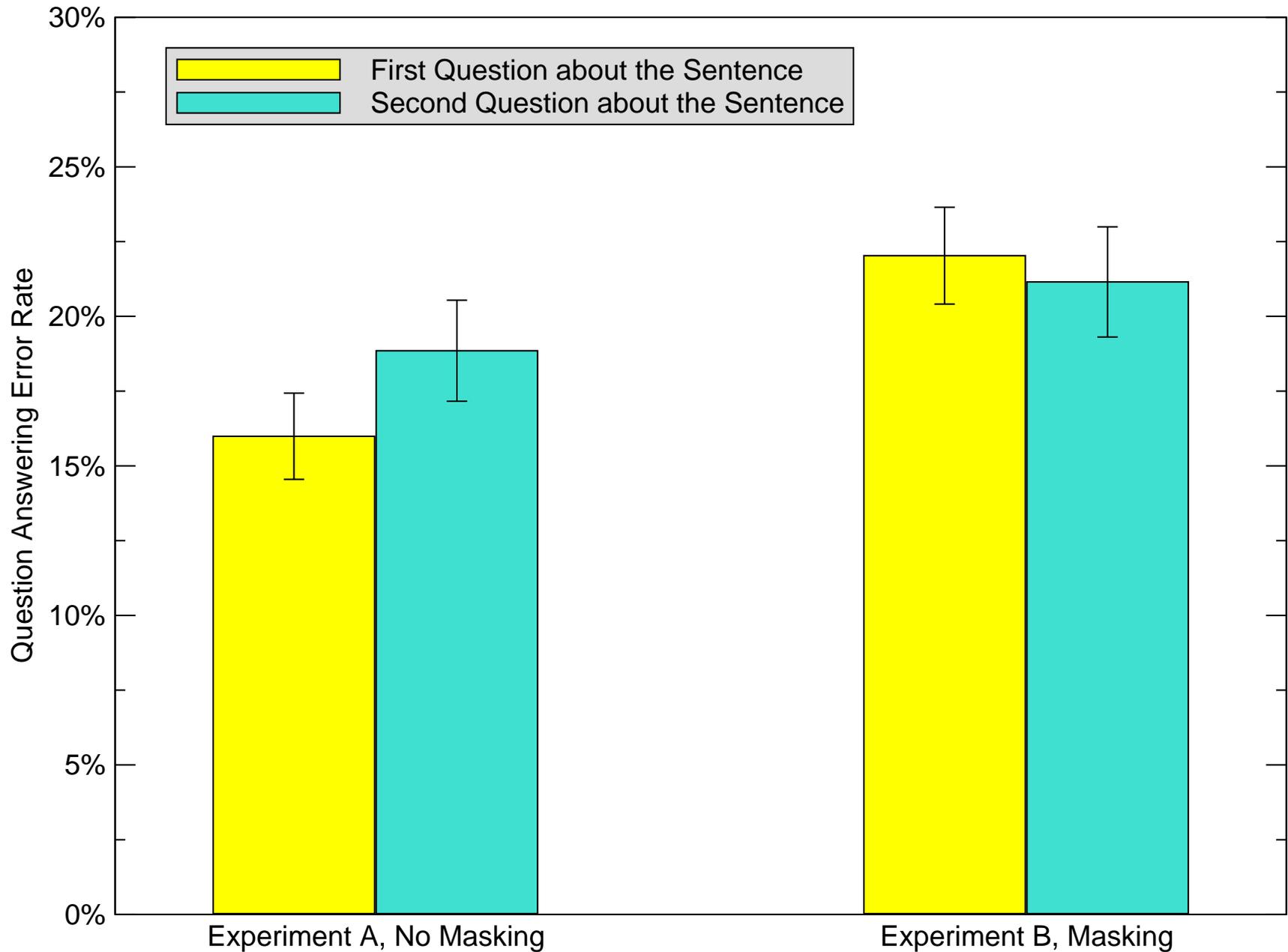
Overall Question Answering Error Rate



Effect of Word Order on Error Rate



Effect of Question Number on Error Rate



Question Answering Conclusions

- On Exp. A, there are no significant effects of location or RC type on error rate. CO is numerically higher, but $p > 0.4$ for all tests.
- Therefore, when controlling for plausibility biases and word-order biases, readers do not find CO sentences more difficult.
- Overall, error rates are higher for Exp. B.
- In this case, there is a main effect of clause type and a difference between RS and RO.
- Without masking, readers are much better at answering questions in which the word order matches the sentence. Presumably, they are using replay and surface form comparisons.
- In Exp. B, this effect goes away. **With masking, readers are not able to replay and the relative comprehension error rates better reflect true thematic comprehension.**
- Supporting this conclusion: Without masking, readers are better at the first question of the two, $p = 0.02$, when the phonological trace is fresh. With masking, the effect goes away.