UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Mouse tracking for syntactic and semantic interference

Permalink
https://escholarship.org/uc/item/9cx5n325

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Carlin-Coleman, Gabriel
Kloostra, Li
Bonfil, Ben
et al.

Publication Date
2026

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Mouse tracking for syntactic and semantic interference

Gabriel Carlin-Coleman (horizon.gabriel12@gmail.com)[1,2], Li Kloostra (l.kloostra@uu.nl)[1],
Ben Bonfil (b.bonfil@uu.nl)[1] & Jakub Dotlačil (j.dotlacil@uu.nl)[1]
[1]Institute for Language Sciences, Utrecht University
[2]Department of Linguistics, University of Toronto

Abstract

This study investigates the effectiveness of 'mouse tracking for reading' (MoTR), a novel incremental processing paradigm developed by Wilcox et al. (2024). MoTR adapts the eye-tracking procedure for online implementation. The present experiment is a replication of Mertzen et al.'s (2023) English-language eye-tracking task. Compared to Wilcox et al.'s original study, this work replicated not just syntactic but also semantic effects on reading, tested longer and more complex sentences, and directly compared the behaviour of trackpad and regular mouse users. Evidence of semantic interference was observed, with limited evidence of syntactic interference. The replication of these results adds to the evidence that MoTR can usefully supplement or even replace eye tracking.

Keywords: mouse tracking; semantic interference; syntactic interference; English

Introduction

Research on the mechanisms underlying sentence processing has frequently posed questions which can be answered only with precise behavioral data collection methods. One such technique is eye tracking while reading (ETR), which records eye movements and fixations during reading (Rayner & Liversedge, 2004). Patterns of eye movements during reading are assumed to be reflective of language processing (Conklin & Pellicer-Sánchez, 2016; Kliegl et al., 2006; Matin, 1974). However, a difficulty associated with ETR is its relatively high cost in terms of time, effort, and money. ETR experiments require specialized equipment, must be performed serially in a lab setting, and need at least one trained experimenter to be present. By contrast, other methods that tap into linguistic processing using reading, such as self-paced reading (SPR), may be administered via the internet and can therefore collect data from many participants at once without requiring special equipment, training, or extensive time commitments. The downside, of course, is that online experiments typically cannot provide the same level of spatiotemporal precision or naturalism of experience that ETR offers (Rayner, 1998).

The challenge for psycholinguists is to select a paradigm that strikes an appropriate balance between efficiency and precision for the experiment at hand. Though ETR is generally considered the 'gold standard' (Rayner & Liversedge, 2004) against which other self-paced incremental processing paradigms are measured (Wilcox et al., 2024), more economical techniques like SPR or the maze task (Forster et al., 2009) represent acceptable alternatives. In an SPR, sentential stimuli are read one part at a time, with experiment participants advancing through the sentence with button presses. In a maze task, participants see (all or part of) sentences in word pairs representing possible sentence continuations; one word will be the correct/appropriate continuation while the other is an inappropriate distractor, and participants advance by selecting one or the other. For both methods, reaction time (to a sentence chunk or to a word pair) is the only correlate of language processing. SPR, the maze task, and similar paradigms can be implemented online and take advantage of participant crowdsourcing platforms. However, compared to ETR these techniques lack naturalism: normal reading is not done in predetermined chunks nor are readers prevented from glancing back and forth between different parts of the sentence. A material effect of these drawbacks is that reading speed in SPR is half of that in ETR (Rayner, 1998), while maze task reaction times tend to be slower than both ETR and SPR (Wilcox et al., 2024); maze tasks also bring a built-in confound to the interpretation of reading behaviour since they require subjects to read a distractor word that is not actually part of the target sentence and which can influence reaction times.

A recent development that may improve on existing online incremental processing paradigms is reported by Wilcox et al. (2024). Their method, called 'Mouse tracking for reading' (MoTR), presents blurred-out stimuli which can be made temporarily readable by hovering a computer cursor over the sentence, un-blurring material within the 'spotlight' area around the cursor for as long as it is nearby. The aim of the authors was to imitate the experience of an ETR study, treating mouse movements as an analog for gaze and applying a blurring effect to mimic parafoveal vision. Wilcox et al. carried out two experiments. In the first they analyzed general reading patterns. In the second they attempted to replicate the well-known late attachment effect in English parsing (Cuetos & Mitchell, 1988; Frazier, 1979). Wilcox et al. found evidence that MoTR and ETR reading measures are reasonably well-correlated and that MoTR captured effects on reading speed more precisely than SPR. Notably, when comparing MoTR, ETR, and 'bidirectional' SPR - a version of SPR in which readers can freely jump backwards and forwards between chunks (Paape & Vasishth, 2021) - the authors found that MoTR regressive reading behaviour and likelihood of word-skipping were closer to that of ETR than bidirectional SPR was.

MoTR is not the first attempt to transpose eye tracking into an Internet-accessible form. WebGazer (Papoutsaki et al.,

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society* ©2026 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

2018; Papoutsaki et al., 2016), or online eye tracking, uses computer webcams to track participant eye movements. There is a substantial literature reporting on its successes when used with a visual world paradigm (e.g. Degen et al. (2021), Slim and Hartsuiker (2022), and van der Cruyssen et al. (2024), exp. 2). However, as Slim et al. (2024) report, WebGazer appears to have lower spatial and temporal precision than in-lab eye tracking, which may make it inappropriate for finer-grained reading studies.

One area in which precise incremental processing techniques are needed is agreement attraction, a type of 'grammaticality illusion' (Phillips et al., 2011) in which an erroneous agreement relation is established between two elements, typically a subject noun and a verb. The sentence in (1) is ungrammatical due to the mismatch between its singular subject ('key') and its plural-inflected verb ('are'). However, experiment participants have been found to systematically treat such sentences as grammatical, due to the presence of the plural non-subject distractor noun ('cabinets'). Errors are observed in both production (Bock & Miller, 1991) and comprehension (Jäger et al., 2020; Wagers et al., 2009).

(1) *The key to the cabinets are on the table.

Attraction-like effects are not restricted only to subject-verb number agreement; more broadly, distractors are linked to other kinds of interference effects on the basis of various linguistic features, such as noun animacy or syntactic role (subject/object). Distractor interference effects can show up as slowed or speeded processing times during reading comprehension experiments (Mertzen et al., 2023; Wagers et al., 2009). However, the exact timing, magnitude, and even the existence of interference effects vary between studies and between languages tested: Mertzen et al. (2023) found a semantic interference effect that was delayed in German compared to English; Chromý et al. (2023) failed to find any agreement attraction effect in Czech; and Nicenboim et al. (2023) estimate that the size of the number agreement effect is small (9ms). Thus, there is a need for spatially- and temporally-precise methods to help elucidate the nature of interference effects, particularly as more and more languages are tested.

Studies of interference effects, then, face the challenge of requiring reasonably high-precision paradigms, while also potentially needing access to diverse participant demographic groups due to the cross-linguistic variability of the phenomenon. It is no surprise that few interference studies have used both ETR and recruited participants from different linguistic backgrounds. Mertzen et al. (2023) represents one of the very few that tested speakers of two different languages (English and German) using an eye tracking paradigm. However, work on interference could benefit greatly from methods with precision approaching that of ETR and which are implementable online.

A recent study by Oğuz et al. (2025) has indeed used MoTR to study agreement processing in Russian. They succeeded in replicating reading slowdowns, found by (Fuchs et al., 2025), caused by gender mismatches between nouns and either verbs or adjectives. It remains to be seen, however, how well MoTR can replicate cases of non-syntactic interference, or cases where agreeing elements are distant from each other. This latter point is particularly important, since MoTR is less likely to elicit regressive movements than ETR (Wilcox et al., 2024).

The current study

In this paper, we describe an experiment using MoTR, administered over the internet, seeking to evaluate its ability to replicate subtle syntactic and semantic interference effects on parsing of long and complex sentences. We adapt the setup, procedures, and analysis developed by Wilcox et al. (2024), who originally looked only at parsing of unmanipulated narrative discourses and at short sentences that promoted high or low attachment readings. Additionally, we depart from Wilcox et al. in allowing our participants to choose to complete the experiment using a computer trackpad. Our goal is to replicate ETR results using MoTR, building support for MoTR as a viable method of data collection for psycholinguists.

Using MoTR, we sought to replicate the English-language experiment of Mertzen et al. (2023). The study manipulated syntactic and semantic properties of distractor nouns, attempting to elicit reading time slowdowns as evidence of memory retrieval interference effects. The findings of Mertzen et al. support the cue-based retrieval model of memory access in language, which predicts interference effects based on features of memory items (Lewis & Vasishth, 2005).

By selecting this study, we are able to test MoTR using longer sentences, a useful step towards applying MoTR to research about multi-sentence discourses, a growing area of interest for researchers interested in retrieval interference (e.g. Schmitz et al. (2025) and Winkowski et al. (2024)). Furthermore, retrieval interference effects tend to be small and their investigation benefits from the use of sensitive data collection methods, including eye tracking. Our study is thus an 'acid test' for MoTR, since success in finding evidence for an effect as subtle as retrieval interference would suggest that MoTR is suitable for a wider range of research questions.

Method

As a replication study of Mertzen et al. (2023), our experiment followed the same design, materials, and general procedure as the original (English-language) experiment, with the notable difference that we used MoTR rather than ETR. The study received ethical approval from the Faculty Ethics Assessment Committee of Utrecht University.

Participants 64 participants (30 female, 34 male) were recruited online via Prolific. All participants were native English speakers, declared no language-related disorders, and ranged in age from 19 to 78 years (median = 36). One was excluded from analysis due to performance on filler item comprehension questions. Additionally, participants would have been excluded if they could not make their browser window reach minimum size dimensions.

All participants used a computer (not e.g. a tablet). 46 reported completing the experiment using a regular mouse, while 18 reported using a trackpad (none reported using something different).

Materials The same 40 experimental sentences from the English-language experiment of Mertzen et al. (2023) were adopted as stimuli, but only 60 of the 92 fillers from that study were used. We chose to shorten our experiment in order to reduce fatigue and physical strain, which were reported by some pilot participants. Two practice sentences were shown before the main experiment (also taken from Mertzen et al.).

There were 4 conditions, the experiment had a 2×2 within-subjects Latin square design. Experimental sentences contained multiple embedded clauses, and had a 'target' subject noun, an 'introducer' noun, and a 'distractor' noun. The target was always the subject of a verb that followed the clauses containing the introducer and distractor nouns. The introducer was always the subject of a relative clause modifying the target and was the subject of a different verb that immediately followed. Targets and introducers were always animate (and human). The distractor noun varied across 4 experimental conditions in terms of whether it was animate or inanimate, and whether it was the subject of its clause or an (indirect) object. The clause with the distractor was separated from the verb (whose subject was the target noun) by an adverb. An example stimulus in all 4 conditions is shown in (2).

(2) It was assumed that the priest whose visitor complained

about the strict nun	[-subject, +animate]
about the strict church	[-subject, -animate]
that the church was strict	[+subject, -animate]
that the nun was strict	[+subject, +animate]

probably departed through the backdoor after the mass.

The regions of interest were the same across conditions and consisted of a precritical region, the adverb (*probably* in the example above), a critical region, the verb (*departed* in the example above), and a postcritical 3 word-long spillover region following the verb (*through the backdoor* in the example above). Regions of interest were always on the second line and were never at the start or end of the line. A wrap-up series of words separated the spillover region from the line end, and at least one word separated the preverbal adverb from the line start. This follows standard practice in eye tracking studies since faster and possibly shallower processing is expected at line starts and ends (Hofmeister et al., 1999; Schotter & Dillon, 2025). The same may be true for mouse tracking. Alternatively, the greater effort involved in moving the mouse diagonally to a new line may lead to slower apparent reading times at line ends. Either possibility would confound analyses of reading speed in these areas.

Design and mouse tracking The study was implemented in jsPsych and hosted on the data server of Utrecht University. The experiment required the browser window to be at or above a minimum height of 600px and width of 1000px. The window was automatically full-screened. Sentences were presented in 18 point Courier New font. Courier New is a monospace font, meaning all letters have equivalent width, which made it easier to control the placement of words and lines when designing the experiment. The maximum length of each line was 80% of the minimum width meaning line breaks appeared in the same places for all participants. All stimuli required line breaks due to their length. Stimuli were fixed in the center of the screen so that their positions were not adjusted based on window size. The line height (distance between lines) was 10 times the font size, a substantially larger line spacing than that used by Wilcox et al. We chose a larger spacing so that readers were very unlikely to reveal the next line early or hold their mouse between both lines, revealing both simultaneously. This made determining which line participants were reading completely unambiguous.

Sentences were blurred until revealed by a 'spotlight' around the cursor. The blurring effect forced readers to move their mouse when reading, giving 'fixation'-like reading behaviour, while preserving information about line breaks and word length which would be available in an eye tracking study (Wilcox et al., 2024). The spotlight temporarily revealed parts of the occluded sentences and was 102px wide and 38px high. The center of the spotlight was offset 12 pixels to the left of the cursor. Wilcox et al. made this choice in their original study in order to mirror how gaze typically picks up more characters to the right than to the left. The spotlight was offset 6 pixels above the cursor so that it would not get in the way of words. A gradual blurring effect was applied around the edge of the spotlight to mimic the out-of-focus experience of parafoveal human vision. All values described here exactly match those used by Wilcox et al. A still image from our experiment and one from Wilcox et al.'s first experiment are shown in Figure 1.

The jsPsych 7.3.3 mouse tracking extension was used to capture mouse movements.[1] The extension records X and Y window coordinates and timestamps of mouse movement events, as well as the bounding rectangles of elements. We established bounding rectangles around each word in a stimulus. Mouse movements outside of the rectangles were not recorded. The rectangles of words on the same line bordered each other horizontally. Each rectangle was 72px high; widths varied depending on the length of the word. There was empty space, not covered by any rectangle, between the two lines of text in each stimulus; this space was 108px in height. All mouse movements within the rectangles were recorded. A mouse movement is defined as any time that the mouse moves. A movement of a single pixel would lead to the initial mouse position and timestamp being recorded. The experi-

[1] The repository to run MoTR in jspsych is available at: https://github.com/UiL-OTS-labs/jspsych-mouse-reading

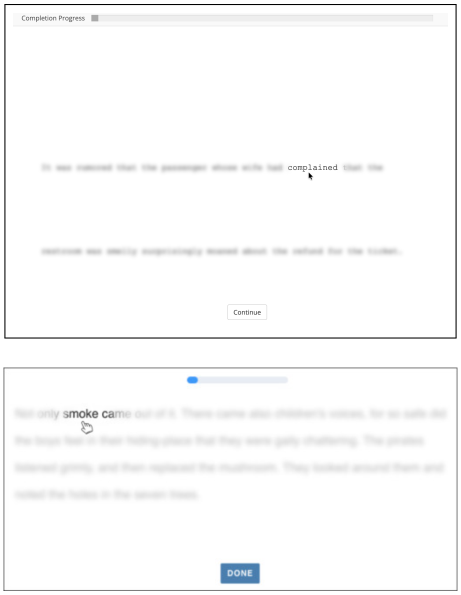

Figure 1: Screenshots of our MoTR experiment (above) and the Wilcox et al. study (below).

ment script automatically recorded whether a movement was an 'entry' or 'leave' event. An 'entry' is a mouse movement into a rectangle that follows a mouse movement outside of that rectangle. A 'leave' is a mouse movement in a rectangle that is followed by a movement outside of that rectangle.

Procedure After accessing the study via Prolific and providing informed consent, participants read instructions and completed two practice trials. The instructions explained how to use the mouse cursor to reveal the sentences and told participants they would have to answer questions about what they read. Following Mertzen et al. (2023), all test trials had comprehension questions in which participants had to choose among four options, one of which was 'Don't know'. Half of filler trials were also followed by comprehension questions with four answer options. Rest breaks were not coded into the experiment script (whereas Mertzen et al. offered a break at the halfway point, but had 32 additional items). There was no limit on how long participants could stay on a given trial, so it was possible for them to proceed at their own pace or take pauses when they wished, although Prolific automatically imposed a maximum time of 97 minutes after which compensation is not awarded. No participants who completed the experiment timed out.

Analysis Data were prepared for analysis by a Python script which calculated region-level 'associations' (analogous to eye tracking fixations) based on mouse 'leave' and 'enter' movements. Before proceeding to the analysis, entire trials containing associations greater than 10 seconds were excluded. This was intended to exclude trials where readers momentarily stopped paying attention to the experiment or paused their participation. Then, associations below 50ms and above 3000ms in duration were removed. This led to the removal of less than 0.2 per cent of the data. After applying this cutoff, individual associations were coded as 'first pass' associations if there were no associations temporally preceding it that were assigned to a 'higher' (further along in the stimulus) word/region. Associations were coded as 'second pass' if there was an earlier association assigned to a higher word/region. Wilcox et al. applied a higher minimum cutoff of 160ms. We chose not to do the same because such a high cutoff point led to the removal of significant portions of the data (37 per cent of the data would be removed with this cutoff point). Furthermore, Wilcox et al. chose this higher cutoff because they observed longer associations in mouse tracking compared to eye tracking, but we argue that this does not necessarily mean that processing of individual words is slower in mouse tracking. The longer MoTR associations may be due to the extra effort it takes to move a mouse compared to moving the eyes, and not due to slower processing of words while they are revealed. It is possible that words with association times between 50ms (our cutoff) and 160ms are read and processed in MoTR.[2] This is an area that deserves more study, if MoTR is to be generally adopted by psycholinguists.

We discuss three reading measures:

- Right-bounded duration (RBD): the sum of all associations within a given region before progressing to a later part of the sentence.
- Total association duration (TAD): the sum of all associations within a given region.
- Re-reading time (RRD): the sum of all second pass associations within a given region.

We fitted hierarchical Bayesian models with a shifted log-normal likelihood to association measures in the pre-critical, critical, and post-critical regions. Three model specifications were considered. The *basic interaction model* included animacy (animate: +0.5; inanimate: −0.5), subjecthood (subject: +0.5; object: −0.5), and their interaction as fixed effects. The *model with nested fixed effects* included subjecthood, with animacy nested within the subjecthood manipulation (i.e., separate animacy contrasts for subject and object distractors). The *full interaction model* included animacy, subjecthood, tracking device (mouse: +0.5; trackpad: −0.5), and all interactions among these predictors. All models included maximal random-effects structures, with random intercepts and slopes for within-subject predictors by participant and by item, correlated random effects, and no by-participant random slopes for tracking device or its interactions, as tracking device was a between-subject manipulation. Weakly informative priors were used throughout, following the common practice in the field. Posterior distributions were sampled using four chains with 3,000 iterations per chain, of which half were used for

[2]We also ran our analyses with the 160ms cutoff to more closely match Wilcox et al. (2024) but this did not lead to any clear change in the general patterns - i.e., effects that excluded zero still did so and vice versa, with a couple of exceptions, mentioned below.

warm-up. There were no divergent transitions and all $\hat{R}$ values were below 1.05, indicating satisfactory convergence.

Results & Discussion

Here, we summarize those posterior distributions whose 95% credible intervals (CrI) exclude or almost exclude zero, indicating positive or negative effects of the experimental manipulation. RBD and TAD measures by condition across the critical regions are shown in Figure 2. Figure 3 visualizes posterior distributions for TAD.

At the preverbal (precritical) adverb, there were main effects of animacy on total association duration (TAD) and rereading duration (RRD). The estimated effects were positive: 95% CrI [22.4,87.3]ms for TAD and 95% CrI [17.8,117.3]ms for RRD. The estimated effect of animacy on right bounded reading duration (RBD) was also largely positive, 95% CrI [-6.9,45]ms. There was an interaction between animacy and subjecthood on TAD (95% CrI [-125.6,-0.6]ms and RRD (95% CrI [-195.8,-14]ms. Models with nested animacy suggest the effect of animacy is driven by the object conditions (TAD 95% CrI [43.1,133.7]ms, RBD [-8.9,60.6]ms). Estimates of the effect of subjecthood in this region all centered on 0.

At the critical verb, there was a positive main effect of animacy on TAD (95% CrI [12.7,63.1]ms) and an almost positive effect of animacy on RBD (95% CrI [-1.8,35.6]ms). Models with animacy nested in syntactic manipulations suggest that these effects were driven mainly by object conditions[3] (TAD 95% CrI [15.3,83.4]ms, RBD 95% CrI [-8.5,45.9]ms). A marginally positive main effect of subjecthood was observed in this region, on TAD (95% CrI [-10,39.4]ms).

In the post-critical spillover region, all estimates centered on 0, for both effects of interest.

The inclusion of the tracking device in the full interaction model showed no clear main effects of tracking device or interaction with the experimental manipulation in the regions and the measures reported here.

Comparison with eye tracking We do not report regression-path duration (RPD), unlike Mertzen et al. This is because our experiment script only recorded mouse events over words and did not track time spent with the cursor hovering over parts of the screen. As a result, we cannot determine RPD with complete confidence. Unlike Mertzen et al., we calculated right-bounded duration, a related first-pass reading measure, though we make no direct comparison between RPD and RBD results.

Like Mertzen et al., we detected a precritical effect of animacy on TAD; our estimate was roughly similar in magnitude to theirs (ours [22.4,87.3]ms, theirs [20, 73]ms). Their effect of animacy on first-pass duration was somewhat matched by our almost positive effect on RBD. We did not replicate their effect of subjecthood in the precritical region (which they found in several measures). There was an interaction effect on TAD in both studies, marginal in eye tracking (ours [-125.6,-0.6]ms, theirs [-64, 8]ms).

At the critical verb, both studies observed main effects of animacy on total reading times, though the ETR estimate crosses 0 (ours [12.7,63.1]ms, theirs [-8,33]ms). Both studies reported marginal effects of subjecthood on total reading times (ours [-10,39.4]ms, theirs [-3,33]ms). Mertzen et al. observed effects of subjecthood and animacy on probability of regression, which we did not.

In the spillover region, we observed no reliable effects, whereas Mertzen et al. report a negative effect of subjecthood on total reading times.

Discussion MoTR replicated a retrieval interference effect caused by animate distractors, previously found by Mertzen et al. (2023) using eye tracking. We showed limited evidence for a replication of an interference effect caused by subject distractors. The evidence for these effects was reliable mainly for TAD, though there was marginal evidence for an effect of animacy on RBD. The MoTR experiment also replicated the precritical effect of animacy on reading.

Retrieval interference effects suggest that certain linguistic features are active as cues during language processing, instrumental to the relationship between memory and sentence parsing (Van Dyke, 2007). Feature match accounts, such as cue-based retrieval theory, argue that items in memory compete for retrieval when both possess features matching the relevant retrieval cue (Lewis & Vasishth, 2005). In comprehension, competition can result in slowed reading times; the mouse and eye tracking findings of interference by animate and subject distractors thus bear out the predictions of cue-based retrieval. However, the precritical slowdowns reported by both our study and Mertzen et al.'s are not predicted by cue-based retrieval. Mertzen et al. argue that precritical effects are the result of inhibitory interference during *encoding*, which continues to affect reading speed in the following regions. Another possible explanation is that early, parafoveal processing of the preverbal adverb is affected by features of the distractor; however, similar findings by Schoknecht et al. (2025) using SPR would appear to rule this out as an explanation. Regardless, the replication of critical and precritical effects in the present study provides encouragement for the future use of MoTR as an alternative to eye tracking for the study of memory encoding and retrieval.

The mouse tracking results provided much stronger evidence for an effect of animacy than for an effect of subjecthood. One reason for this could be MoTR's lower sensitivity compared to ETR. Wilcox et al. (2024) noted that MoTR reading times were generally slower than those of ETR and that MoTR elicited only a portion of regressive movements that would be expected based on ETR behavior. If syntactic interference in English is subtler than semantic, MoTR may not have been precise enough to pick up the effect. Indeed, Mertzen et al. (2023)'s own comparison of their English and German results suggest that German readers prioritized syn-

[3]Using a 160ms cutoff, models with nested animacy showed an effect on TAD in subject conditions (95% CrI [5.7,67.1]ms) and a very marginal effect on objects (95% CrI [-10,48.4]ms).

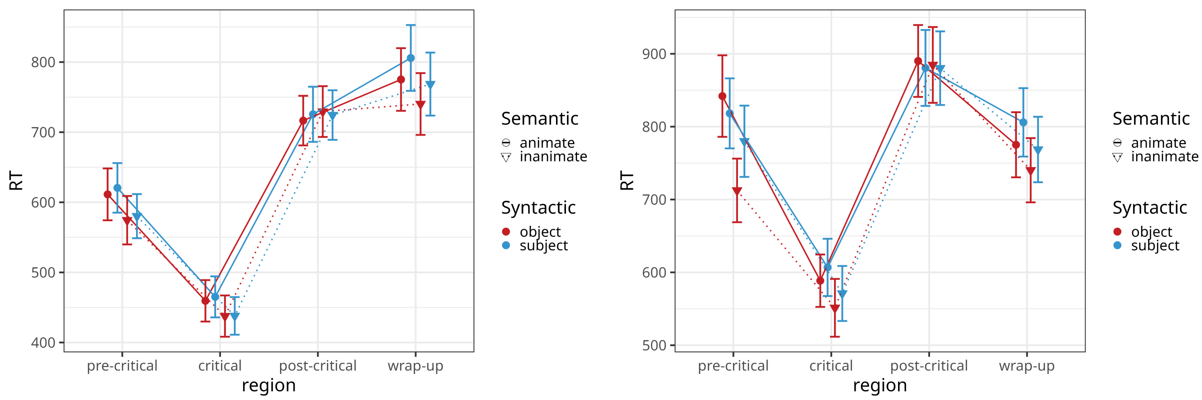

Figure 2: RBD (left) and TAD (right) per region. The points represent means and the bars are 95% confidence intervals.

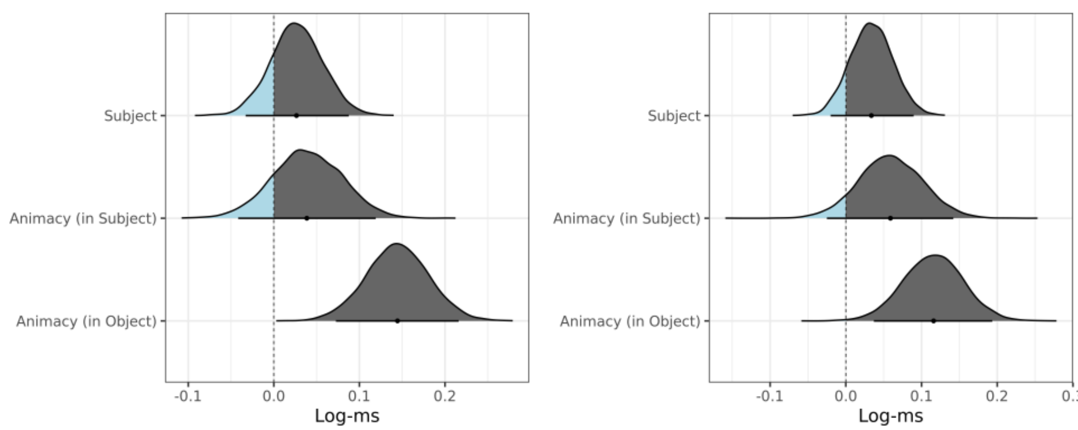

Figure 3: Posterior distributions for TAD for the pre-critical (left) and critical (right) region in the models with animacy nested in syntactic manipulations. The thick lines represent 95% credible intervals.

tactic cues more than English readers did. On the other hand, previous work by Cunnings and Sturt (2018) and Van Dyke and McElree (2011) instead showed 'gating' of semantic interference in English, whereby semantic features of distractors impacted processing only after syntactic features were evaluated. Van Dyke and McElree (2011), specifically, found semantic interference only when distractors were subjects. Our results show the opposite pattern: as our nested analysis reveals, the main effect of animacy was driven most strongly by the object conditions.

The lowered sensitivity of MoTR compared to ETR could be the reason for these somewhat anomalous results. Notably, in object conditions, distractors were separated from the critical verb by 1 word, whereas in subject conditions, they were separated by 3 words. Given that MoTR reading times are slower and regressive movements less likely than in ETR, it may be the case that object distractors were 'fresher' in memory than Subject distractors, and therefore more capable of influencing processing at the point of retrieval.

Our results let us tentatively suggest that trackpads are suitable for MoTR studies. Allowing trackpad usage did not confound our results. However, just a small proportion (28%) of participants completed the experiment using a trackpad, which may limit the ability of our analysis to detect reliable effects of tracking device. Future work on this issue could prove useful, since trackpad usage would expand the participant pool available to researchers, something particularly relevant for MoTR studies recruiting within underrepresented demographic groups.

Conclusion

Our study provides a useful demonstration of the strengths and weaknesses of MoTR as an incremental processing paradigm. On one hand, the results back up Wilcox et al. (2024)'s assessment that MoTR is not as precise as ETR. The setup of MoTR may have reduced the impact of syntactic interference and limited the lasting effect of semantic interference. Conversely, MoTR succeeded in replicating, with roughly similar magnitudes, several effects observed in the ETR study. This replication was present even though we had longer, more complex stimuli than those studied by Wilcox et al. (2024) in their original study. Finally, the work presented here provides further support for cue-based retrieval theory and other feature-match accounts of memory retrieval during comprehension.

Acknowledgments

We would like to thank Iris Mulders and Rick Nouwen for discussion and help throughout the project. The research reported in this paper was supported by the European Research Council (ERC), grant 101088098 - MEMLANG. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

References

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, *23*(1), 45–93. https://doi.org/https://doi.org/10.1016/0010-0285(91)90003-7

Chromý, J., Brand, J. L., Laurinavichyute, A., & Lacina, R. (2023). Number agreement attraction in Czech and English comprehension: A direct experimental comparison. [Retrieved from https://escholarship.org/uc/item/4md8g417]. *Glossa Psycholinguistics*, *2*(1), 1–48. https://doi.org/http://dx.doi.org/10.5070/G6011235

Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research*, *32*(3), 453–467. https://doi.org/https://doi.org/10.1177/0267658316637401

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition*, *30*(1), 73–105. https://doi.org/https://doi.org/10.1016/0010-0277(88)90004-2

Cunnings, I., & Sturt, P. (2018). Retrieval interference and semantic interpretation. *Journal of Memory and Language*, *102*, 16–27. https://doi.org/https://doi.org/10.1016/j.jml.2018.05.001

Degen, J., Kursat, L., & Leigh, D. D. (2021). Seeing is believing: Testing an explicit linking assumption for visual world eye-tracking in psycholinguistics. [Retrieved from https://escholarship.org/uc/item/6182t9jb]. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*.

Forster, K. I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. *Behavior Research Methods*, *41*(1), 163–171. https://doi.org/https://doi.org/10.3758/BRM.41.1.163

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies.* [Doctoral dissertation, University of Connecticut].

Fuchs, Z., Parshina, O., Sekerina, I., Polinsky, M., & Sekerina, I. A. (2025). Processing of verbal versus adjectival agreement: Implications for syntax and psycholinguistics. *Glossa: A Journal of General Linguistics*, *10*(1). https://doi.org/https://doi.org/10.16995/glossa.11173

Hofmeister, J., Heller, D., & Radach, R. (1999). The return sweep in reading. In W. Becker, H. Deubel, & T. Mergner (Eds.), *Current oculomotor research* (pp. 349–357). Springer.

Jäger, L. A., Engelmann, F., & Vasishth, S. (2020). Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language*, *94*, 316–339. https://doi.org/http://dx.doi.org/10.1016/j.jml.2017.01.004

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology*, *135*(1), 12–35. https://doi.org/10.1037/0096-3445.135.1.12.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419. https://doi.org/https://doi.org/10.1207/s15516709cog0000_25

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, *81*(12), 899–917. https://doi.org/https://doi.org/10.1037/h0037368

Mertzen, D., Paape, D., Dillon, B., Engbert, R., & Vasishth, S. (2023). Syntactic and semantic interference in sentence comprehension: Support from English and German eye-tracking data. *Glossa Psycholinguistics*, *2*(1), 1–48. https://doi.org/http://dx.doi.org/10.5070/G60111266

Nicenboim, B., Vasishth, S., Engelmann, F., & Suckow, K. (2023). Exploratory and confirmatory analyses in sentence processing: A case study of number interference in German. *Cognitive Science*, *42*(4), 1075–1100. https://doi.org/https://doi.org/10.1111/cogs.12589

Oğuz, M., Ding, C., Wilcox, E. G., & Fuchs, Z. (2025). Using motr to probe agreement processing in russian. *Open Mind*, *9*, 1682–1710. https://doi.org/https://doi.org/10.1162/OPMI.a.35

Paape, D., & Vasishth, S. (2021). When nothing goes right, go left: A large-scale evaluation of bidirectional self-paced reading. [Retrieved from psyarxiv.com/d7pvz.].

Papoutsaki, A., Gokaslan, A., Tompkin, J., He, Y., & Huang, J. (2018). The eye of the typer: A benchmark and analysis of gaze behavior during typing. In *Etra '18 proceedings of the 2018 acm symposium on eye tracking research applications, june 14-17, 2018, warsaw, poland* (pp. 1–9). https://doi.org/https://doi.org/10.1145/3204493.3204552

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). Webgazer: Scalable webcam eye tracking using user interactions. *IJCAI'16: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 3839–3845.

Phillips, C., Wagers, M. W., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In J. Runner (Ed.), *Experiments at the interfaces* (pp. 147–180). Brill.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K., & Liversedge, S. P. (2004). Visual and linguistic processes during eye fixations in reading. In J. Henderson

& F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 59–104). Psychology Press.

Schmitz, T., Nouwen, R., & Dotlačil, J. (2025). Memory retrieval in discourse: Illusions of coherence during presupposition processing. *Journal of Memory and Language*, *143*, 104637. https://doi.org/https://doi.org/10.1016/j.jml.2025.104637

Schoknecht, P., Yadav, H., & Vasishth, S. (2025). Do syntactic and semantic similarity lead to interference effects? Evidence from self-paced reading and event-related potentials using German. *Journal of Memory and Language*, *141*, 104599. https://doi.org/https://psycnet.apa.org/doi/10.1037/0278-7393.33.2.407

Schotter, E. R., & Dillon, B. (2025). A beginner's guide to eye tracking for psycholinguistic studies of reading. *Behavior Research Methods*, *57*(68). https://doi.org/https://doi.org/10.3758/s13428-024-02572-4

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*, *55*, 3786–3804. https://doi.org/https://doi.org/10.3758/s13428-022-01989-z

Slim, M. S., Kandel, M., Yacovone, A., & Snedecker, J. (2024). Webcams as windows to the mind? A direct comparison between in-lab and web-based eye tracking methods. *Open Mind: Discoveries in Cognitive Science*, *8*, 1369–1424. https://doi.org/https://doi.org/10.1162/opmi_a_00171

van der Cruyssen, I., Ben-Shakhar, G., Pertzov, Y., Guy, N., Cabooter, Q., Gunschera, L. J., & Verschuere, B. (2024). The validation of online webcam-based eye-tracking: The replication of the cascade effect, the novelty preference, and the visual world paradigm. *Behavior Research Methods*, *56*, 4836–4849. https://doi.org/https://doi.org/10.3758/s13428-023-02221-2

Van Dyke, J. A. (2007). Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology*, *33*(2), 407–430. https://doi.org/https://psycnet.apa.org/doi/10.1037/0278-7393.33.2.407

Van Dyke, J. A., & McElree, B. (2011). Cue-dependent interference in comprehension. *Journal of Memory and Language*, *65*(3), 247–263. https://doi.org/https://doi.org/10.1016/j.jml.2011.05.002

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. https://doi.org/10.1016/j.jml.2009.04.002

Wilcox, E. G., Ding, C., Sachan, M., & Jäger, L. A. (2024). Mouse tracking for reading (MoTR): A new naturalistic incremental processing measurement tool. *Journal of Memory and Language*, *138*, 104534. https://doi.org/https://doi.org/10.1016/j.jml.2024.104534

Winkowski, J., Dotlacil, J., & Nouwen, R. (2024). Searching for semantic distance effects. *Glossa Psycholinguistics*, *3*(1).