GLOSSA psycholinguistics

Kondo, Y., & Yano, M. (2026). Readers can recognise multiple words at a glance but out of order: Evidence from transposed word effects in Japanese. *Glossa Psycholinguistics*, 5(1): 19, pp. 1–39. https://doi.org/10.5070/G6011.62009

eScholarship
University of California

Readers can recognise multiple words at a glance but out of order: Evidence from transposed word effects in Japanese

Yukiho Kondo, Department of Language Sciences, Tokyo Metropolitan University, g2590003@stu.musashino-u.ac.jp

Masataka Yano, Department of Language Sciences, Tokyo Metropolitan University; Advanced Language Science Research Centre, National Institute for Japanese Language and Linguistics, yanomasa@tmu.ac.jp

Readers find it difficult to reject ungrammatical sentences with two adjacent words transposed (e.g. *I met my friend kind at the station*), and they sometimes misclassify such sentences as grammatical – a phenomenon known as the *transposed word effect*. According to the parallel processing model, readers simultaneously process multiple words of a sentence, which creates uncertainty about the spatial position of words and leads them to perceive transposed sentences as if they were in the correct order. Recent studies, however, have challenged this interpretation and proposed that readers process words serially, and rational inference subsequently corrects transposition errors in the sentence representation. Using a rapid parallel visual presentation (RPVP) paradigm that minimises overt left-to-right serial reading, this study provides new behavioural and pupillometric evidence that transposed word effects can arise from fast parallel processing of words, under conditions in which words are unlikely to have been identified one at a time. Our findings contribute to the ongoing debate regarding serial versus parallel processing in reading, by elucidating the temporal dynamics of the transposed word effect.

1. Introduction

The ability to process written language represents one of the most remarkable capacities of the human mind. Despite the apparent ease with which skilled readers can process text, the mechanisms underlying this ability involve a complex interplay among perceptual, cognitive, and linguistic processes. Understanding these mechanisms is a central issue in the cognitive science of language. Converging evidence in this domain indicates that readers process sublexical information in parallel (Meade et al., 2021, 2022; Perea & Carreiras, 2006; Perea & Lupker, 2003, 2004). Evidence for this comes from transposed-letter effects, where pseudowords with two (adjacent) transposed letters are perceived as if they were real words. In a simple lexical decision task, for example, participants were slower and less accurate in judging non-words with transposed letters, such as *mohter*, as a non-word than orthographically legal, yet clearly non-word, items, such as *nordle*, due to activation of the lexical representation of the real word (O'Connor & Forster, 1981). The transposed-letter effect has also been observed in masked priming paradigms. Because a transposed-letter prime such as *shpae* is perceived as *shape*, it facilitates the recognition of the target word *shape*, yielding faster response times (RTs) in the lexical decision task and reduced N400 effects in event-related potential experiments, compared to a clearly non-word prime, such as *shgue* (Meade et al., 2022; Perea & Carreiras, 2006; Perea & Lupker, 2004). These results suggest that readers simultaneously process and encode multiple letters with only approximate positions in a word, and these good-enough representations, while generally efficient, become susceptible to transposition errors, as they are likely to activate a lexical representation of the corresponding real word (Davis, 2010; Grainger & Van Heuven, 2004).

1.1 Transposed word effects

In contrast to sublexical processing, debate continues regarding whether multiple words can be processed in parallel or must be processed serially (Hossain & White, 2023; Huang & Staub, 2021a, 2021b, 2023; Liu et al., 2020, 2022; Massol et al., 2021; Mirault et al., 2018, 2022a; Snell & Grainger, 2017, 2019b, 2019c; White et al., 2018, 2019). One influential line of evidence proposed for the parallel processing account comes from transposed *word* effects (TWEs). Applying this logic to word-level processing, Mirault et al. (2018) conducted an experiment to examine TWEs. In their study, participants were presented with transposed sentences such as *The white was cat big*, grammatical sentences such as *The white cat was big*, and clearly ungrammatical sentences such as *The white was cat slowly*, and were asked to judge their grammaticality as quickly as possible. The results revealed significantly longer RTs for the transposed sentences compared to the ungrammatical control sentences. Furthermore, the participants sometimes failed to classify the transposed sentences as ungrammatical, resulting in lower accuracy compared to the ungrammatical control sentences.

These results have been interpreted as evidence against the dominant serial processing models, including the E-Z Reader model (Reichle et al., 1998, 2003). If readers allocate their attention to only one word at a time, they should not confuse word order; this predicts that readers can readily detect transposition errors, such as *The white was cat big*.

By contrast, the TWE is consistent with parallel processing models, such as the OB1-reader model. In this model, parallel processing of multiple words with flexible word position coding creates a primitive syntactic representation, which helps constrain possible word orders in a sentence through top-down processes (Mirault et al., 2018; Snell & Grainger, 2019b; Snell et al., 2017, 2018). The combination of bottom-up flexible position coding and the top-down syntactic constraints explains why transposed sentences are misperceived as grammatical (Snell & Grainger, 2019c; Snell & Melo, 2024).

Such flexible word position coding introduces uncertainty regarding the location of words, thereby allowing the potential transposition of words in a sentence representation. This uncertainty is centred on the actual location of each word in a sentence, as evidenced by the larger TWE for transposed adjacent words (*The can man run*) compared to transposed non-adjacent words (*Run man can the*) in both RTs and accuracy (Snell & Grainger, 2019c).

Syntactic constraints function as a crucial factor in accounting for the TWE, because transposed sentences are perceived as grammatical, but grammatical sentences are not perceived as transposed sentences. Evidence for top-down syntactic constraints was provided by Wen et al. (2021b), who demonstrated that the transposed sentences in which two adjacent words were transposed *within* a syntactic phrase, such as *The girls gone have home*, induced a larger TWE than the sentences in which the words were transposed *across* a syntactic phrase, such as *The have girls gone home*. This finding indicates that the TWE is related not only to flexible word position coding, but also to syntactic constraints, such as phrase structure.

In sum, proponents of the parallel processing account argue that the combination of parallel processing with flexible word position coding and top-down syntactic constraints causes confusion for readers regarding what is actually seen, requiring readers to take some time to reach the correct conclusion that the transposed sentences are ungrammatical. When readers cannot overcome the interference from the activated sentence representations, they might incorrectly judge transposed sentences as grammatical. While flexible encoding allows skilled readers to ensure rapid language comprehension, it sometimes compromises the literal meaning of a sentence, as predicted by the OB1-reader model (Snell & Grainger, 2019c; Snell & Melo, 2024). However, as we discuss in the next section, this interpretation has been challenged on both methodological and theoretical grounds.

1.2 The role of rational inference

Existing evidence for parallel processing remains inconclusive, because the target sentences of these experiments were presented until participants made a response, or were presented for sufficiently long periods to allow them to read the sentences from left to right and even regress to earlier parts of the sentence, which undermines the role of parallel processing in accounting for the TWE (Mirault et al., 2018; Snell & Grainger, 2019c; Snell & Melo, 2024). More critically, the interpretation of the TWE has been challenged by recent studies that used rapid *serial* visual presentation (RSVP), in which a sentence is presented word-by-word and, thus, parallel processing is assumed to be limited (Hossain & White, 2023; Huang & Staub, 2023; Liu et al., 2022; Mirault et al., 2022b). For example, Hossain and White (2023) reported that the TWE occurred even with RSVP in RTs and accuracy, questioning the role of parallel processing in causing the TWE. Instead, they attributed the TWE to rational inference (Gibson et al., 2013). In the noisy-channel model of language communication, perceivers are considered rational agents who attempt to infer a likely intended message from potentially noisy input transferred from a speaker. As evidence for this model, Gibson et al. (2013) showed that in offline comprehension tasks, participants were less likely to literally interpret semantically implausible sentences, such as *The mother gave the candle the daughter*, compared to semantically plausible sentences, such as *The mother gave the daughter the candle*. Applying this idea to the TWE, Hossain and White (2023) argued that readers process words serially and hold them in a memory buffer until a plausible phrase, clause, or sentence boundary is reached. Readers then make a rational inference about them to obtain the most likely intended meaning of the sentence, using prior knowledge of the world, for example. As a result, readers might correct the transposed errors based on the plausible meaning of the sentences, leading to the judgement that the transposed sentences are grammatical. However, it is worth noting that findings from the RSVP studies are equivocal; some studies did not find any TWE in accuracy or RT, or found the TWE only in accuracy (see Section 3 for details).

Similarly, Huang and Staub (2023) compared TWEs in the RSVP and PVP (parallel visual presentation) paradigms. In the RSVP paradigm, participants read transposed sentences, such as *They hardly her text or call her,* word by word and were asked to judge their grammaticality at the end of the sentences (Experiment 1) or at any point (Experiment 2). In the PVP paradigm, participants read the sentences from left to right at their own pace and judged the grammaticality of the sentences. The results of both RSVP experiments showed significant TWEs in terms of accuracy, although the effects were smaller than those in the PVP (RTs were not reported). These results suggest that parallel processing is not necessary for the TWE to occur. Furthermore, the presence of the TWE in Experiment 2 ruled out the possibility that the incorrect acceptance was due to the transposed error being repaired in the sentence representation before the participants made their grammaticality judgements. In another study, Huang and Staub (2021a) analysed eye movements of participants reading transposed sentences such as *The fragile cup red shattered into pieces*. Interestingly, the transposed words did not exhibit a processing disruption when the participants incorrectly accepted the transposed sentences. They argued that although word recognition proceeds serially (as assumed in the E-Z Reader model), the integration of words into the unfolding sentence representation can be non-incremental, allowing the ungrammatical order of the transposed words, such as *cup red*, to be switched, based on a rational inference process. Consequently, this process suppresses the error signal, resulting in no processing difficulty. Thus, rapid rational inference can account for the TWEs without appealing to parallel processing (Huang & Staub, 2021b).

It is worth noting that Hossain and White's account and Huang and Staub's account differ in the timing of the operation of rational inference. Hossain and White propose that rational inference applies after words are held in a memory buffer until reaching a plausible phrase, clause, or sentence boundary. Huang and Staub, in contrast, argue that integration and the associated repair can occur over a more local window in which two, or a few, adjacent words are temporarily held in an unintegrated state, allowing rational inference to be brought to bear before the end of the sentence is reached. Despite this difference in timing, both accounts share the directional prediction that semantic plausibility should enhance TWEs (i.e. larger TWEs when the base sentence is semantically plausible), since the repair mechanism is in both cases driven by the plausibility of the resulting interpretation.

Most recently, McGowan et al. (2026) provided evidence for the role of temporal overlap in integration proposed by Huang and Staub (2021b, 2023). They reported that the TWE in accuracy was larger when sentences were presented at a faster rate (125 ms per word) compared to a standard rate (250 ms per word), and when the first, rather than the second, word of the transposed pair was presented at the faster rate. They argued that these findings support the view that TWEs arise from a temporal overlap in the integration stages of adjacent words within a serial architecture. We return to the relationship between their findings and ours in Section 3.

1.3 Summary of the two accounts

To summarise, the transposed sentences were judged more slowly and less accurately than the ungrammatical control sentences. The parallel processing model suggests that this effect arises from the combination of bottom-up and top-down effects, namely, parallel processing of multiple words with flexible position coding and a primitive sentence-level representation that corrects the transposed errors and makes participants perceive them as grammatical. In contrast, the rational inference account proposes that this effect arises from post-lexical plausibility judgements that lead to the misjudgement of grammaticality.

1.4 Overview of the present study

The present study aimed to test these two hypotheses by conducting a behavioural experiment in Japanese. Crucially, we used *rapid parallel* visual presentation (RPVP) to examine the role of parallel processing in TWEs. This paradigm differs from the two previously used paradigms in important ways. First, unlike the parallel visual presentation (PVP) paradigm used by Mirault et al. (2018), in which sentences were presented until participants made a response and, thus, participants read the sentences from left to right and even regressed back to the earlier part of the sentence, our RPVP paradigm presented sentences for only 400 ms. We predicted that participants would not be able to fixate on every single word in the five-word sentence serially, as fixation durations are 230 ms, on average, in Japanese, with 20–50 ms saccadic suppression between fixations (Gagl et al., 2022; Matin, 1974; Schotter et al., 2012) (see Appendix A for details; all the Appendices are provided as supplementary files in the OSF repository for this article). Unlike previous TWE studies, our study used an eye tracker to examine fixation positions before and during the sentence presentation. The analysis of eye movements, in fact, confirmed that participants did not read the sentences from the beginning to the end (see Appendix C for this analysis). Thus, this brief presentation duration minimised serial processing of multiple words. Second, unlike Hossain and White (2023), who used rapid serial visual presentation (RSVP) to enforce serial processing of words, our RPVP paradigm presents all words simultaneously, permitting parallel processing while minimising the opportunity for overt left-to-right serial reading. This allowed us to elucidate the role of parallel processing in the TWE, as originally proposed by Mirault et al. (2018).

One might think that RPVP does not completely preclude serial processing, because serial processing models, such as E-Z Reader, incorporate *covert* shifts of attention that do not involve eye movements. However, these shifts still impose strict temporal constraints, because they are contingent upon the completion of lexical processing (the second stage of word identification, L2) of the currently attended word, which takes approximately 150–300 ms, according to Reichle et al. (2003). Given the 400 ms presentation duration for the five-word sentences used in the present study, participants would need to complete the second stage of processing for each word in approximately 80 ms per word, far below the 150–300 ms estimated for the L2 stage in the E-Z Reader model. Therefore, under the E-Z Reader model, multiple serial shifts of attention would be severely constrained in our experiment.

One might further ask whether strategic adjustments to the E-Z Reader model, as proposed by Reichle et al. (2021), could allow serial processing under rapid display conditions. Specifically, Reichle et al. (2021) suggested that readers could adopt a lexical-processing deadline ($\theta = 92$ ms) that terminates the first stage of word identification (L_1) earlier than the model's default parameters allow. Even under this strategy, processing of each word requires several stages: visual propagation $t(V) = 60$ ms (incurred at least once, when the display first appears), the first stage of lexical access $t(L_1) = 92$ ms (under the deadline strategy), the second stage of lexical access $t(L_2) = \Delta \cdot t(L_1) \approx 20$ ms ($\Delta = 0.22$), and the attention shift $t(A) = 25$ ms.[1] Serial processing of five words would, therefore, require at least approximately $60 + 5 \times (92 + 20) + 4 \times 25 = 720$ ms, approximately 1.8 times our 400 ms presentation duration. We acknowledge that lexical and post-lexical processing may continue after stimulus offset; however, this does not lift the temporal constraint on serial attention shifts. Each shift in E-Z Reader is contingent upon the completion of L_2 for the currently attended word, and visual uptake of any word not yet attended ceases once the display offsets at 400 ms. The sequential dependency of attention shifts on lexical processing, thus, remains the bottleneck.

A further consideration concerns compensatory serial-reading strategies under brief presentation. Reichle et al. (2021) proposed, on the basis of E-Z Reader simulations, that readers may skip short function words and infer their identities from context when reading subtitles displayed at high rates, thereby maintaining comprehension while reducing the number of fixations required. This strategy, however, is not applicable to our stimuli. Japanese is an agglutinative language in which functional morphemes, such as case particles and verbal inflections, are bound to content words (see also the description of the stimuli in Section 2.2). The five-word sentences used in the present experiment, therefore, consist entirely of content words: there are no short, highly predictable function words (analogous to English *the*, *a*, or *is*) that could be selectively skipped to free up processing time for content words. This typological property of Japanese rules out a key compensatory serial-reading strategy that has been proposed for the rapid reading of English text.

Thus, using the RPVP paradigm, in which participants cannot read the sentences from left to right or regress back to the earlier part of the sentence, the present study aimed to elucidate the role of parallel processing in the TWE. If parallel processing of multiple words drives the TWE, we should observe this effect in the RPVP paradigm.

We also examined the role of rational inference by manipulating the plausibility of sentences. If rational inference drives the TWE, it should be enhanced when the transposed sentences are derived from more plausible base sentences (i.e. the non-transposed sentences are semantically plausible) than when they are derived from less plausible base sentences (i.e. the non-transposed sentences are semantically less plausible). This is because readers would be more likely to be misled by the semantic plausibility of the sentence, leading them to misjudge the transposed

1. This estimate does not incorporate the modulation of $t(L_1)$ by visual acuity, which is given by $t(L_1) \leftarrow t(L_1) \cdot \varepsilon^{\sum_{i=1}^{M} |\text{fixation}-\text{letter}_i|/M}$ ($\varepsilon = 1.15$), where the exponent indicates "the mean distance (in character spaces) between each of the M letters in a word being processed and the current fixation location (i.e. center of the vision)" (Reichle et al., 2021, p. 446). Because participants in our paradigm fixated centrally while the five-word sentence subtended five degrees of visual angle, peripheral words were several character spaces from the fovea and would, under serial processing, require considerably more time than $t(L_1) = 92$ ms. Including this factor would only strengthen our argument that 400 ms is insufficient for the serial processing of five words.

sentences as grammatical or to take longer to judge them as ungrammatical. Thus, a significant interaction between transposition and plausibility would support the rational inference account. For the parallel processing models, the top-down constraints discussed in previous studies (e.g. Wen et al., 2021a, 2021b) focus on syntactic constraints rather than semantic constraints. Thus, the effect of semantic plausibility on the TWE is not straightforwardly predicted by the parallel processing models. However, the parallel processing models might accommodate the effect of semantic plausibility, as they can, in theory, incorporate top-down constraints from semantic plausibility. Thus, the parallel processing models could potentially explain the interaction between transposition and plausibility. Importantly, the absence of the interaction would be less consistent with the rational inference accounts, as such a result suggests that semantic plausibility plays little role in TWEs.

We adopted a pupillometric technique to track the time course of cognitive effort before participants made their grammaticality judgements. Pupillometry is a physiological technique for measuring automatic pupil dilation and constriction, which are governed by the autonomic nervous system. In particular, pupil dilation is closely associated with the noradrenergic neuronal activity of the locus coeruleus (Aston-Jones & Cohen, 2005; Beatty, 1982; Winn et al., 2018). As pupil dilation can be used as an index of cognitive effort ranging from perceptual processes to higher-order processes under stable ambient conditions (e.g. light intensity is adequately controlled), pupillometry has been widely used in the field of cognitive science (e.g. Hess & Polt, 1964, for mathematical calculation complexity; Kahneman & Beatty, 1966, for digit-span tasks).

In psycholinguistic studies, pupillometry has been used to investigate word-level phenomena (e.g. effects of frequency and phonology-to-orthography consistency in word recognition) and sentence-level phenomena (e.g. effects of structural complexity on syntactic processing) in both comprehension and production (Engelhardt et al., 2010; Funasaki & Yano, 2025; Haro et al., 2017; Just & Carpenter, 1993; Mukai et al., 2023; Yano et al., 2025). Because pupillometry can continuously track time-varying cognitive load from sentence presentation to participant responses, this technique is useful for examining the time course of TWEs.

We used Japanese as the target language for the experiment. Japanese provides a critical test case for the universality of theories of parallel word processing, as most existing evidence comes from alphabetic languages with spaces between words. Japanese presents a unique case with its mixed writing system, which combines logographic (kanji) and syllabic (kana) scripts, and lacks explicit word boundaries (i.e. unspaced text) (see Liu et al., 2022, for TWEs in Chinese). If TWEs reflect a fundamental property of human sentence processing, they should be observed across diverse orthographic systems; conversely, if TWEs are specific to alphabetic writing systems with explicit word boundaries, this would indicate that the underlying mechanisms are more language-specific than previously assumed.

To preview our findings, behavioural data show that grammatical sentences and ungrammatical control sentences with morphosyntactic violations are correctly judged as such, while transposed sentences are sometimes misjudged as grammatical or take longer to be judged correctly, regardless of the plausibility of the base sentences. These results suggest that readers can recognise multiple words simultaneously but, crucially, out of order. The pupillometric data revealed that the transposed sentences required more cognitive effort than the ungrammatical control sentences when the base sentences were semantically *less* plausible. No such difference was observed when the sentences were semantically *more* plausible, contrary to the predictions of the rational inference account. These results support the parallel processing models. Importantly, we found that TWEs emerged early in the time course, which is difficult to reconcile with purely serial accounts, as these would predict that the TWE would emerge later, only after sequential processing of multiple words.[2]

2. Experiment

2.1 Participants

We recruited 30 native speakers of Japanese who were (under)graduate students at Tokyo Metropolitan University (mean age = 21.07 years, SD = 2.05). None of the participants reported any history of language or neurological disorders. The experimental protocol received approval from the Ethics Committee of Tokyo Metropolitan University (approval number: H7-008). Written informed consent was obtained from all participants before the experiment.

Because our design was quite similar to that of Hossain and White (2023), we adopted their reported power analysis (90% power for N = 28) as the basis for our sample size determination. We acknowledge that detecting interaction effects typically requires larger samples than detecting main effects does (Brysbaert, 2019; Staub, 2025). However, conventional power analysis is designed to detect effects that exist, not to provide evidence for null effects. Because our theoretical predictions included the possibility that there would be no meaningful Plausibility × Transposition interaction, we complemented our frequentist analyses with Bayesian analyses, which can quantify evidence for the null hypothesis (Wagenmakers et al., 2018). This approach allows us to distinguish between “no evidence for an effect” and “evidence for no effect.”

2. A further piece of evidence supporting parallel word processing models in reading has been reported: the sentence superiority effect (SSE). In SSE experiments, participants were briefly presented with grammatical sentences such as *Our fox can fly* and scrambled sentences such as *Our can fly fox* (Declerck et al., 2020; Massol et al., 2021; Snell & Grainger, 2017; Wen et al., 2019). Subsequently, they were asked to type a word that had appeared in a specific position in the stimulus. The results revealed higher word identification accuracy in the grammatical condition than in the scrambled condition, suggesting that when multiple words are processed in parallel, a sentence-level representation is activated in the grammatical condition, which facilitates word identification (Snell & Grainger, 2017). This interpretation is analogous to that of the classic word superiority effect, another piece of evidence for parallel letter processing (Reicher, 1969; Wheeler, 1970; Yokosawa & Umeda, 1988). We did not employ SSE as a diagnostic tool to test parallel processing in Japanese, because the task requires participants to engage in additional processing – specifically, more conscious recognition of lexical items, in order to report them (see Snell & Grainger, 2019a, for discussion).

2.2 Stimuli

The sentences in (1) show an example set of target sentences. Following Japanese conventional orthography, words were not separated by spaces, as shown in the first lines of each target sentence in (1), and they were written using a mixture of three writing systems: logographic characters, hiragana syllabary for native Japanese, and katakana syllabary for loan words. These stimulus characteristics differ from those in previous studies in English (Hossain & White, 2023), French (Mirault et al., 2018), and Dutch (Snell & Melo, 2024).

(1) a. the grammatical condition (+plausible/−plausible):

郊外の駅で{優しい/平凡な}恋人に会った。

kougai-no	eki-de	yasashii/heibonna	koibito-ni	at-ta.
suburb-GEN	station-LOC	kind/insipid	partner-DAT	meet-PST

'I met my kind/insipid partner at a suburban station.'

(Japanese order: 'at a suburban station my kind/insipid partner met')

b. the transposed condition (+plausible/−plausible):

郊外の駅で恋人に{優しい/平凡な}会った。

*kougai-no	eki-de	koibito-ni	yasashii/heibonna	at-ta.
suburb-GEN	station-LOC	partner-DAT	kind/insipid	meet-PST

(Lit.) 'I met my partner kind/insipid at a suburban station.'

(Japanese order: 'at a suburban station my partner kind/insipid met')

c. the ungrammatical control condition (+plausible/−plausible):

郊外の駅で恋人に{優しい/平凡な}飲んだ。

*kougai-no	eki-de	koibito-ni	yasashii/heibonna	non-da.
suburb-GEN	station-LOC	partner-DAT	kind/insipid	drink-PST

(Lit.) 'I drank my partner kind/insipid at a suburban station.'

(Japanese order: 'at a suburban station my partner kind/insipid drank')

The sentences of type (1a) were well-formed, grammatical sentences consisting of five words. The transposed sentences of type (1b) were created by changing the order of two adjacent words in the grammatical sentences (the underlines indicate transposed words). Crucially, to prevent participants from adopting a detection strategy, we varied the position of transposition across trials: half of the transposed sentences involved switching the 3rd and 4th words, while the other half involved switching the 2nd and 3rd words (see Appendix B for other examples). This design differs from most previous TWE studies, which created the transposed condition by always switching adjacent words at the same positions across all items (e.g. the 3rd and 4th words; see e.g. Hossain & White, 2023; Liu et al., 2022; Mirault et al., 2018, 2022b; Staub, 2021; Wen et al., 2021a). Such fixed-position designs adopted in previous studies may enable participants to develop a strategy for detecting transposed errors by focusing on particular positions within the sentences immediately upon sentence presentation. Our design makes it unlikely that participants attempted such strategies (see Appendix C for eye movement data during sentence presentation confirming this).

These transpositions created ungrammatical sentences, because Japanese permits adjective-noun order, but not noun-adjective order. The control condition was formed by replacing the verbs in the transposed conditions with verbs from different sentence sets to make the sentences semantically anomalous and syntactically ill-formed, as shown in (1c). This sentence is syntactically ill-formed, because the monotransitive verb *drink* requires an accusative object and cannot assign a dative case. As with the transposed sentences, the ungrammatical control sentences included grammatical errors in various positions, requiring the participants to examine the entire sentence holistically during sentence presentation to perform the task.

We also manipulated sentence plausibility to examine the role of rational inference in TWEs. We achieved this manipulation by replacing the semantically plausible adjectives in NPs such as *kind partner* (+plausible) with the less plausible adjectives in NPs such as *insipid partner* (−plausible). Note that the −plausible/grammatical conditions were made less plausible, yet were still acceptable in Japanese, because they were intended to be judged as grammatical. Prior to the main experiment, a norming survey confirmed that the +plausible/grammatical conditions were rated more plausible than the −plausible/grammatical conditions (+plausible: mean = 4.43 on the 5-point scale (SD = 0.39), −plausible: mean = 3.30 (SD = 0.63), $\beta = -1.11$, SE = 0.12, $t = -9.36$, $p < 0.01$) (see Appendix B for details).

Crossing +/−plausible and three types of sentences yielded six conditions. A total of 240 sets of sentences were created and distributed into three lists, using a Latin Square design. Each list contained 40 trials per condition, with 80 grammatical fillers included to balance the number of grammatical and ungrammatical sentences.

2.3 Apparatus

Eye movements and pupil sizes were recorded from both eyes at a sampling rate of 1000 Hz, using EyeLink 1000 Plus (SR Research, Ontario, Canada). The eye tracker was positioned approximately 54 cm in front of the participants. Head position was stabilised using a chinrest (SR Research) to fix the visual angle of the stimuli across participants. The calibration was performed at the beginning of the experiment and after short breaks. When necessary, the experimenter recalibrated the system to prevent tracking loss. Stimuli were presented on a monitor (BenQ EX2510S with a 165 Hz refresh rate) in a randomised order for each participant, using Experiment Builder ver. 2.4.1 (SR Research). Sentences were displayed in 14-point black monospaced Meiryo UI font against a grey background (RGB: 180, 180, 180).

Participants were instructed not to blink or look away from the centre of the monitor during each trial, from the fixation check until the end. Participants were asked to indicate whether the sentence was grammatical or not as quickly as possible, by pressing one of two keys. Behavioural data were recorded using a response pad (Cedrus RB-740; CA, USA). The "Yes" and "No" keys were counterbalanced across participants. During the experiment, the room was dimly lit to keep ambient illumination constant across participants.

2.4 Procedure

Figure 1 illustrates the trial sequence. After the eye tracker confirmed that participants were fixating on the centre of the monitor, the sentence was presented at the same position for 400 ms. This procedure was designed to minimise the opportunity for left-to-right serial reading.

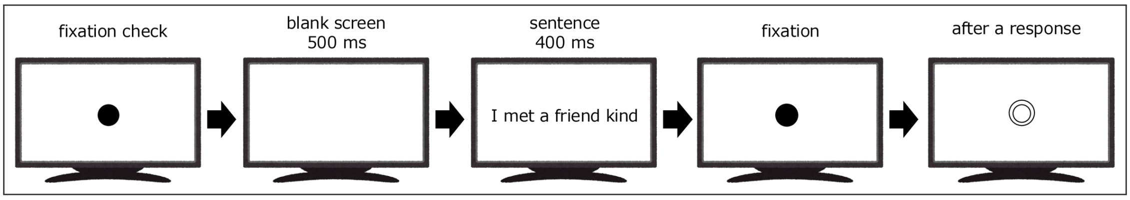

Figure 1: Illustration of the experimental trial sequence, showing the progression from fixation check to response recording.

Previous studies using the RPVP paradigm employed durations of 200–300 ms for three- or four-word sequences with several functional words such as *the* and *can* (Fallon & Pylkkänen, 2024; Snell & Grainger, 2017; Wen et al., 2021a). However, Japanese is an agglutinative language in which functional morphemes such as case particles and verbal inflections must be attached to content words (i.e. bound morphemes), and, thus, the transposed sentences cannot be created by switching content and functional words, unlike in previous studies in alphabetic languages. Furthermore, content words in Japanese use complex logographic characters, which might be difficult to recognise with such short durations.

Thus, we conducted a pilot experiment with grammatical and ungrammatical control sentences at three durations (200 ms, 300 ms, 400 ms) to determine an appropriate stimulus duration.[3] We calculated the sensitivity index, *d'*, from the hit rate (i.e. correctly judging ungrammatical sentences as ungrammatical) and the false-alarm rate (i.e. incorrectly judging grammatical sentences as ungrammatical) for each duration. *d'* is a standard measure in signal detection theory that quantifies how well participants can discriminate between two types of stimuli while accounting for response biases.

3. Six students from Tokyo Metropolitan University completed 20 trials per condition (i.e. grammatical/ungrammatical sentences × 200, 300, 400 ms). The grammatical sentences were the same as in the main experiment (i.e. of type (1a)), while the ungrammatical sentences involved case-assignment violations, but not transposition. The procedure and apparatus of the pilot experiment were identical to those of the main experiment.

The pilot experiment revealed that while all durations showed discrimination significantly above chance (all *ps* $<$ .05), only the 400-ms duration demonstrated sufficient discrimination (*d*’ = 1.79). The 200-ms (*d*’ = 0.35) and 300-ms (*d*’ = 1.08) durations showed substantially weaker discrimination, with the 200-ms condition close to chance level and the 300-ms condition’s confidence interval overlapping with 1.0 (**Figure 2**). Given that our main experiment required reliable detection of ungrammaticality in the ungrammatical control condition, which served as a baseline for TWEs, we selected 400 ms as the presentation duration, to ensure that participants could adequately perceive the sentence while still preventing left-to-right word-by-word reading.

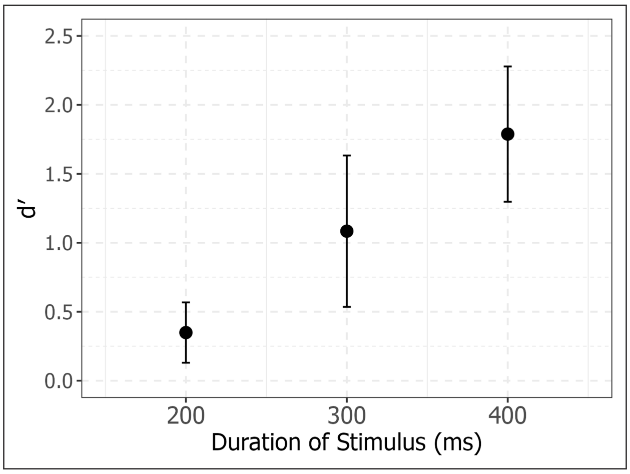

Figure 2: Mean sensitivity (*d*’) for grammaticality judgements as a function of presentation duration in the pilot experiment. Error bars represent 95% confidence intervals.

In the main experiment, each sentence was presented horizontally as a single line centred on the screen and subtended a visual angle of five degrees horizontally (i.e. within the parafoveal area; Schotter et al. (2012)). These sentences were then replaced with a black circle. Participants were asked to indicate whether the sentence was grammatical or not as soon as possible, by pressing one of two keys. Note that this grammaticality judgement task did not require participants to evaluate the plausibility of the sentence. After the response was made, the black circle was replaced by a white circle, which signalled to participants that the response was registered. Participants were instructed to look at the circle even after their response, to enable pupillometric data collection. The circle remained on the screen for 5000 ms, relative to the onset of the sentence presentation. The inter-trial interval (ITI) was substantially longer than 5000 ms, because the fixation check at the beginning of each trial took several seconds to confirm that participants were fixating on the centre of the monitor (**Figure 1**). This allowed pupil size to return to baseline before the next trial began.

After eight practice trials, each participant completed 320 trials, with breaks of 1–2 minutes after every 80 trials (i.e. three breaks in total). These breaks were intended to mitigate fatigue effects, which can attenuate pupillary responses (McLaughlin et al., 2023). Every participant completed the experiment within an hour. Pupillometric data from one participant were not recorded successfully, due to an error in the eye-tracking system, and, thus, were excluded from the analysis.

2.5 Data analysis

Behavioural and pupillometric analyses were based on the coding schema in **Table 1**. The ±Plausibility factor was deviation-coded such that +plausible conditions were compared with −plausible conditions. Sentence types were Helmert-coded such that the Ungrammaticality factor compared the two types of ungrammatical conditions (control and transposed conditions) with the grammatical condition, while the Transposition factor examined the difference between the transposed condition and the control condition. The Transposition factor is of primary interest in the present study, as the TWE has been defined as longer RTs and lower accuracy in the transposed condition, relative to the ungrammatical control condition (e.g. Hossain & White, 2023; Mirault et al., 2018).

Table 1: The coding schema of the fixed factors in the LME models.

	+Plausible			**−Plausible**		
	Grammatical	**Ungrammatical**	**Transposed**	**Grammatical**	**Ungrammatical**	**Transposed**
Plausibility	−0.5	−0.5	−0.5	0.5	0.5	0.5
Ungrammaticality	−2/3	1/3	1/3	−2/3	1/3	1/3
Transposition	0	−0.5	0.5	0	−0.5	0.5

For RT data, we analysed only the trials in which correct responses were given (following Hossain & White, 2023; Mirault et al., 2018). Response times shorter than 400 ms and longer than 5000 ms were excluded from the analysis. Then, a screening procedure was conducted to exclude response times exceeding 2.5 standard deviations from the mean; this affected 3.17% of the data. The response times were then log-transformed and used as a dependent variable in linear mixed-effects (LME) models. The accuracy and response times of the grammaticality judgement task were analysed using the *(g)lmer* function from the *lme4* package in R (Bates et al., 2015). The maximal model with participants and items as random intercepts and slopes was constructed ((g)lmer(RT/Accuracy ~Plausibility * SentenceType + (1 + Plausibility * SentenceType | Participant) + (1 + Plausibility * SentenceType | Item))), and then compared with more parsimonious models with simpler random effects, using the backward stepwise method via the *anova* function (Matuschek et al., 2017). For response time data, (LME) models included Item

Order (the *n*th trial) as a covariate (although its significance was not of interest), because RTs typically decrease over the course of the experiment. The *p*-values were calculated based on Satterthwaite's method, by submitting the final model to the *lmer* function from the *lmerTest* package (Kuznetsova et al., 2017). Final models are reported in Section 2.7. For pupillometric analyses, we interpolated the tracking loss due to blinks, using the *extend_blinks* function of the *gazer* package in R, assuming a linear function (Geller et al., 2020). The pupil size data were smoothed using the *moving_average_pupil* function. The pupil sizes were defined as the diameter (a.u.) relative to a 500-ms pre-stimulus baseline time window, using the standard subtractive baseline correction method (i.e. subtracting the mean baseline pupil size from each post-stimulus sample on a trial-by-trial basis). We chose the subtractive method over the proportional (divisive) method, because the latter has been shown to introduce systematic distortions in the estimation of pupillary response magnitudes, particularly when baseline pupil size varies systematically across conditions or participants (see van Rij et al., 2019; Winn et al., 2018; Yano et al., 2025, for detailed discussion). These baseline-corrected data were submitted to subsequent statistical analyses as the dependent variable. We retained trials with incorrect responses in the pupillometric analysis, since at least the early stages of the pupillary response reflect cognitive effort incurred by word transposition, regardless of whether the conflict arising from word transposition is subsequently resolved or not; incorrect trials, therefore, provide informative data about the cognitive effort underlying TWEs. An analysis without incorrect trials is reported in footnote 7.

The time window from 400 to 2000 ms was extracted in order to perform cluster-based permutation analysis (CPA), a statistical method suitable for time-series data (Maris & Oostenveld, 2007). We restricted the analysis time window to 400 ms post-stimulus onward, because pupillary responses to cognitive load typically emerge around 400 ms after stimulus onset (Hoeks & Levelt, 1993), and, thus, any effects observed before this time window might be due to artefacts rather than cognitive processing of the sentences. In CPA, the time-series data were divided into 50-ms bins, and the pupil sizes were averaged for each trial and eye. Subsequently, *t*-statistics (referred to as original statistics) were calculated for each bin, using the LME models (PupilSize ∼ Plausibility * Ungrammaticality * Transposition + (1| Participant) + (1| Item) + (1| Eye) + ItemOrder + Baseline).[4] These statistics were compared with the null hypothesis distribution created by shuffling the condition labels and running the LME models 1000 times. An effect of interest was considered significant in that bin if the proportion of *t*-statistics obtained from the null hypothesis distribution whose absolute value was at least as large as that of the original *t*-statistic was < 0.05. We assumed that an effect of interest was reliable when at least two consecutive bins reached significance. The CPA was performed using the *clusterperm.lmer* function from the permutes package (Voeten, 2022). Although model selection with random slopes/intercepts is a

4. Item Order and baseline pupil sizes were included as covariates in the LME models, to control for fatigue effects during the experiment and potential artefacts from the baseline time-window (Alday, 2019; McLaughlin et al., 2023).

standard practice in psycholinguistics, successful convergence of models is crucial in CPA, because the null distribution has to be generated by repeatedly running the same model as the original model. Complex models often fail to converge during this iterative process; therefore, we used the simplest model that converged reliably across all permutations.

The significant interaction between Plausibility and Ungrammaticality was further examined by comparing different types of sentences at each level of Plausibility by switching the deviation coding of Plausibility (-0.5 vs. $+0.5$) to treatment coding (0 vs. 1) (see Schad et al., 2020, pp. 29–30, for a coding tutorial). Specifically, we assigned 0 to the +plausible conditions and 1 to the −plausible conditions when examining effects within the +plausible condition, and conversely for the −plausible condition. The models were the same as those used for the main analysis, except for the coding of Plausibility. This approach avoids splitting the data and conducting separate analyses for each plausibility level.

Before turning to the results, we first present some control analyses to rule out the possibility that potential confounds could have influenced the results. First, we examined eye movements during sentence presentation, to check whether participants fixated on the centre of the screen, where the sentences were presented. This was important, because pupil size is overestimated or underestimated when eyes rotate greatly away from an eye-tracking camera, as they appear elliptical to the camera (Brisson et al., 2013; Gagl et al., 2011). The results showed that most fixations were in the centre of the screen (i.e. within 22 mm, which corresponds to 1.37 degrees in visual angle) (see Appendix C for more details). Furthermore, we did not find any significant effects of the factors (i.e. Plausibility, Transposition, and Ungrammaticality) on the number of fixations. These results suggest that eye movements cannot account for the pupillometric effects reported below.

Furthermore, the visual differences between sentences do not constitute a concern for our results, because pupillary responses were analysed at the condition level. Crucially, transposition only involved different word orders of the same lexical items. The ungrammatical control condition was created by shuffling verbs across the entire set of 240 sentences, ensuring that identical characters appeared across all conditions. Since we compared pupillary responses between conditions with the same set of characters overall, any effects of character-specific visual properties, such as letter complexity, were perfectly balanced across conditions, and cannot explain the systematic differences we observed.

Response-related activity cannot be completely isolated from the cognitive pupillary responses, because the speeded grammaticality judgement task, which is essential to confirm TWEs (i.e. longer RTs in the transposed condition than in the control condition), necessarily involves response preparation and execution. However, this does not pose a critical concern, for the following reason. If response-related activity were driving the pupillary effects, we would expect the ungrammatical control condition to elicit greater initial dilation, due to shorter RTs (cf. Mathôt & Vilotijević, 2022). In contrast, we predicted that the transposed sentences would initially elicit larger pupillary responses than the control condition, despite having longer RTs (not shorter). As reported below, this prediction was confirmed; the transposed sentences elicited larger pupillary responses than the control sentences, even though the former had longer RTs. Therefore, the pupillometric data reported below is unlikely to be explained solely by response-related activity.

2.6 Predictions

According to the parallel processing account, rapid parallel processing of multiple words, combined with noisy position coding, creates uncertainty about the precise word order in the transposed sentences. This would lead participants to take increased processing time to determine that the transposed sentences were actually ungrammatical. Thus, we predicted that the transposed sentences would take longer to be judged as ungrammatical than the control sentences. If top-down syntactic constraints make participants overlook the transposed errors, the transposed sentences would be judged as grammatical, resulting in lower accuracy. Since the processing of both plausible and less plausible transposed sentences is expected to activate a primitive sentence representation, no modulation of the transposed word effect by plausibility is predicted (i.e. no Plausibility × Transposition interaction). As for the pupillometry data, we expected that the TWE would be observed as larger pupil responses for transposed sentences than for control sentences, as this account predicts that parallel processing quickly activates a grammatical representation of the base sentence, requiring increased cognitive effort to resolve the conflict between this representation and the ungrammatical input. Note that under a parallel processing account, multiple words are processed quickly in parallel. However, given the inherent latency of pupillary responses to cognitive load, the predicted pupillometric TWEs would be observed at approximately 600 ms (peaking around 1000 ms) post-stimulus (Hoeks & Levelt, 1993).

In contrast, the rational inference account proposes that the TWEs are driven by an inference mechanism that corrects the transposed errors, based on the plausible meaning of the sentences. If this account is correct, the TWE should be more pronounced for transposed sentences with more plausible grammatical counterparts (+plausible) compared to those with less plausible grammatical counterparts (−plausible). This account predicts longer RTs for transposed sentences, compared to control sentences, when they are semantically plausible, but not when they are semantically less plausible. Furthermore, sentences in the +plausible/transposed condition are more likely to be judged as grammatical than sentences in the −plausible/transposed condition, due to the plausible adjective-noun relationship (e.g. *partner kind* versus *partner insipid*).

Importantly, if post-lexical inference underlies the TWE, pupillary effects would appear in a later time window, rather than an early time window. This is because rational inference is applied to correct the transposed errors, based on the plausible meaning of the sentences, only after each word is processed sequentially and stored in a memory buffer, without being integrated into a sentence representation (Hossain & White, 2023). Assuming that transposition errors cannot be noticed when words held in the memory buffer are unintegrated, the pupillometric TWEs, if any, should be observed in a later time window, when rational inference comes into play.

In sum, an interaction between plausibility and transposition would constitute critical evidence for the rational inference account, while the lack of such an interaction should be interpreted cautiously as evidence for the parallel processing account. Pupillometry data should reveal the temporal dynamics of TWEs, with an early effect favouring the parallel processing account and a later effect supporting the rational inference account.

2.7 Results

2.7.1 Accuracy

Figure 3 shows the mean accuracy of the grammaticality judgement task. Participants sometimes missed the ungrammaticality of the sentences, resulting in lower accuracy in the control and transposed conditions, compared to the grammatical conditions (**Table 2**). The main effect of Transposition was significant, suggesting that the transposed sentences were judged less accurately than the control sentences. Crucially, the effect of Transposition was numerically larger in the $+$plausible condition than in the $-$plausible condition, but the LME test of the Plausibility × Transposition interaction did not reach conventional significance (**Table 2**). This result should be interpreted alongside the Bayesian analysis reported below.

Because the interaction of Plausibility and Ungrammaticality was significant, we conducted further analyses to examine the effect of Ungrammaticality at each level of Plausibility (**Table 2**). The effect of Ungrammaticality was significant in the $+$plausible condition ($\beta = -1.58$, SE $=$ 0.10, $z = -15.82$, $p < 0.01$) and the $-$plausible condition ($\beta = -0.33$, SE $=$ 0.08, $z = -4.00$, $p < 0.01$).

We conducted pairwise comparisons for each sentence type to examine the effect of Plausibility. The models revealed a significant effect of Plausibility in the grammatical condition ($\beta = -0.84$, SE $=$ 0.12, $z = -6.53$, $p < 0.01$) and the transposed condition ($\beta = 0.50$, SE $=$ 0.10, $z = 4.73$, $p < 0.01$). The direction of the effect was opposite, as shown in **Figure 3**. Semantic plausibility of the grammatical sentences enhanced the correct identification of their grammaticality, whereas semantic plausibility of the transposed sentences caused participants to perceive them as grammatical. The effect of Plausibility was also significant in the control

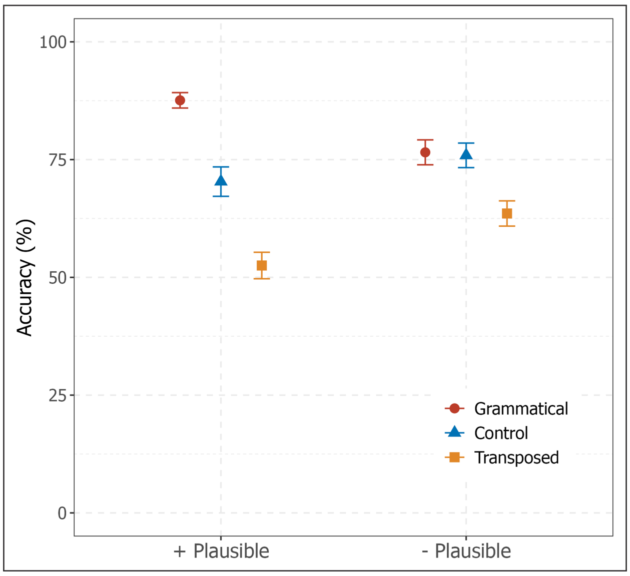

Figure 3: The mean accuracy of the grammaticality judgement task (%). Error bars represent the standard error of the mean. The Plausibility × Transposition interaction did not reach significance in the LME analysis ($p = 0.15$), the Bayesian analysis provided anecdotal evidence for the null hypothesis ($BF_{10} = 0.34$), and the TWE was descriptively larger in the +plausible condition.

condition, with participants being more likely to accept the control sentences in the +plausible condition than in the −plausible condition ($\beta = 0.31$, SE $= 0.11$, $z = 2.80$, $p < 0.05$).

We also conducted a Bayesian analysis, using the *brms* package in R (Bürkner, 2017).[5] The posterior distribution of the interaction effect showed considerable uncertainty, with the 95% credible interval including zero ($\beta = 0.19$, 95% CI $= [-0.07, 0.44]$, Rhat $= 1.00$). The Bayes factor ($BF_{10} = 0.34$) provided anecdotal evidence for the null hypothesis (Lee & Wagenmakers, 2013). Simple effects analyses confirmed that significant TWEs were observed in both the +plausible condition ($\beta = -0.85$, SE $= 0.14$, $z = -5.80$, $p < 0.01$) and the −plausible condition ($\beta = -0.65$, SE $= 0.15$, $z = -4.34$, $p < 0.01$), though the effect was descriptively larger in the +plausible condition.

5. For the accuracy data, we fitted a Bernoulli model with the same fixed-effect structure as the corresponding LME model, together with by-participant and by-item random intercepts. We specified weakly informative priors: Normal(1.0, 1.5) for the intercept, Normal(0, 1) for all regression coefficients, and Exponential(1) for the random-effect standard deviations. Four chains were run for 8,000 iterations each (2,000 warmup). Convergence was confirmed by $\hat{R} = 1.00$. Bayes factors for the Plausibility × Transposition interaction were computed as Savage–Dickey density ratios (Wagenmakers et al., 2010), that is, the ratio of the posterior to the prior density of the interaction coefficient at zero. We report BF_{10}, with values below 1 indicating evidence for the null hypothesis. Because Bayes factors are sensitive to prior specification, we repeated the computation with Normal(0, 0.5), Normal(0, 1.5), and Normal(0, 2) priors on the regression coefficients, yielding $BF_{10} = 0.69$, 0.22, and 0.18 ($BF_{01} = 1.46$, 4.45, and 5.44), respectively. Across all prior specifications, the data were more likely under the null hypothesis than under the alternative.

Table 2: Summary of the fixed effects in the generalised linear mixed-effects models for accuracy.

	Estimate	SE	*z*	*p*	
(Intercept)	1.05	0.10	10.44	< 0.01	
Plausibility	−0.01	0.08	−0.07	0.94	
Ungrammaticality	−0.96	0.06	−14.67	< 0.01	**
Transposition	−0.74	0.06	−11.37	< 0.01	**
Plausibility × Ungrammaticality	1.25	0.13	9.56	< 0.01	**
Plausibility × Transposition	0.18	0.12	1.42	0.15	

**$p < .01$

Final model: glmer(Accuracy ~Plausibility * SentenceType + (1 | Participant) + (1 | Item), family = "binomial")

Because accuracy could, in principle, be shaped by a tendency to respond "grammatical" in the absence of a salient cue to ungrammaticality, we also expressed these results in terms of signal detection theory. Grammatical sentences were treated as *noise* trials, and control and transposed sentences as *signal* trials, with an "ungrammatical" response counted as a signal response. Sensitivity (*d*') and the criterion (*c*) were computed for the control and the transposed conditions separately, against the shared grammatical baseline, for each participant and each level of plausibility, using the log-linear correction (Hautus, 1995). Full details are given in Appendix D.

Sensitivity was reliably above zero in every condition. In the +plausible condition, *d*' was 1.83 for the control condition ($t(29) = 13.33$, $p < 0.01$) and 1.29 for the transposed condition ($t(29) = 10.09$, $p < 0.01$, $BF_{10} = 1.78 \times 10^8$); in the −plausible condition, the corresponding values were 1.55 ($t(29) = 13.91$, $p < 0.01$) and 1.15 ($t(29) = 10.59$, $p < 0.01$, $BF_{10} = 5.18 \times 10^8$). Participants were, therefore, able to discriminate transposed sentences from grammatical ones, albeit less well than control sentences. The TWE was preserved in this bias-independent measure; *d*' was higher for the control condition than for the transposed condition, both in the +plausible condition (difference = 0.54, $t(29) = 5.83$, $p < 0.01$, $BF_{10} = 7.42 \times 10^3$) and in the −plausible condition (difference = 0.40, $t(29) = 5.04$, $p < 0.01$, $BF_{10} = 1.01 \times 10^3$).

The analysis also confirmed the presence of a response bias. The criterion *c* was reliably positive for the transposed condition at both levels of plausibility ($c = 0.58$, $t(29) = 12.51$, $p <$ 0.01 in the +plausible condition; $c = 0.21$, $t(29) = 3.21$, $p < 0.01$ in the −plausible condition), indicating that participants were conservative in responding "ungrammatical." This response bias cannot, however, by itself account for the difference between the control and transposed conditions; the two conditions share the same false-alarm rate, so the difference in *c* between them is algebraically determined by their difference in *d*' (see Appendix D).

2.7.2 Response times

Figure 4 shows the mean RTs of the grammaticality judgement task. The main effect of Plausibility was significant, indicating faster RTs in the +plausible conditions, compared to the −plausible conditions (**Table 3**). The results also showed a significant main effect of Ungrammaticality, indicating longer RTs for the control and transposed sentences than for the grammatical sentences (**Table 3**). These two factors significantly interacted, suggesting that the grammatical sentences were judged faster than the two types of ungrammatical sentences in the +plausible conditions ($\beta = 0.09$, SE $= 0.01$, $t = 8.75$, $p < 0.01$), but not in the −plausible conditions ($\beta < 0.01$, SE $= 0.01$, $t = -0.79$, $p > 0.10$).

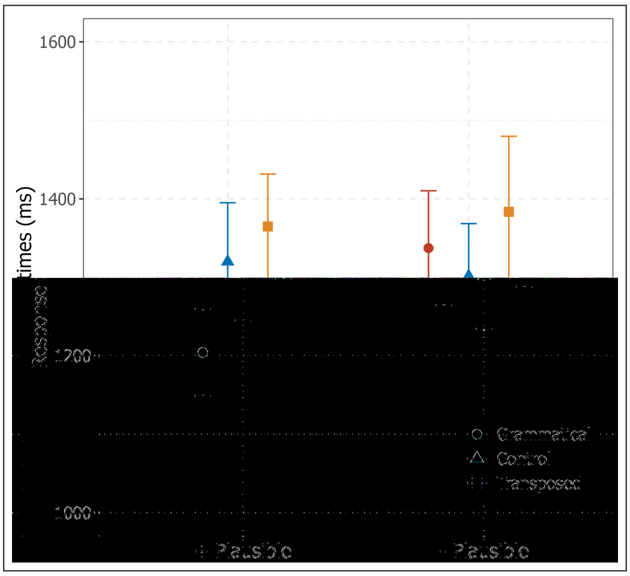

Figure 4: The mean response times of the grammaticality judgement task (ms). Error bars represent the standard error of the mean. The longer response times of the transposed conditions, relative to the control conditions, show the TWE. Pairwise comparison shows no significant effect of Plausibility in the transposed conditions.

Importantly, we observed a significant effect of Transposition, which did not interact with the Plausibility factor (**Table 3**). Pairwise comparisons did not reveal a significant difference between the +plausible/transposed and −plausible/transposed conditions ($\beta = -0.01$, SE $= 0.01$, $t = -0.80$, $p > 0.10$). Therefore, the participants took longer to classify the transposed sentences correctly as ungrammatical than to classify the control sentences, regardless of the Plausibility manipulation. The effect of Plausibility was significant in the grammatical conditions ($\beta = 0.09$, SE $= 0.01$, $t = 6.52$, $p < 0.01$), suggesting a facilitatory effect of semantic plausibility on grammaticality judgements. The effect of Plausibility was not significant in the control conditions ($\beta < 0.01$, SE $= 0.01$, $t = -0.21$, $p > 0.10$).

To address concerns about the absence of the interaction, we conducted a Bayesian analysis. The posterior distribution of the interaction effect was centred around zero, suggesting no meaningful interaction between Plausibility and Transposition ($\beta = -0.01$, 95% CI = [−0.05, 0.03], Rhat = 1.00). Critically, the Bayes factor strongly favoured the null hypothesis (BF_{10} = 0.07), providing strong evidence against the interaction effect. This converges with our theoretical expectation based on parallel processing models (see Appendix E for individual differences in TWEs).[6]

Table 3: Summary of the fixed effects in the linear mixed-effects models for response times.

	Estimate	SE	*t*	*p*	
(Intercept)	7.20	0.04	164.48	< 0.01	
Plausibility	0.02	0.01	2.32	0.02	*
Ungrammaticality	0.04	< 0.01	5.56	< 0.01	**
Transposition	0.03	< 0.01	3.76	< 0.01	**
Plausibility × Ungrammaticality	−0.10	0.01	−6.70	< 0.01	**
Plausibility × Transposition	−0.01	0.01	−0.53	0.59	
ItemOrder	< 0.01	< 0.01	−14.48	< 0.01	**

$^{**}p < .01$, $^{*}p < .05$

Final model: lmer(RTlog ~Plausibility * SentenceType + (1 + Plausibility | Participant) + (1 | Item) + ItemOrder)

2.7.3 Pupillometry

Figure 5 shows the mean pupillary responses from the target sentence onset (ms) to 2000 ms. Cluster-based permutation analyses revealed a significant cluster of the Ungrammaticality effect from 1050 to 1950 ms, suggesting that the two types of ungrammatical sentences were more difficult to process than the grammatical sentences (cluster mass statistic = 680, $t = 5.74$, $p <$ 0.01). The main effect of Plausibility did not yield a significant cluster in any time window. Further analyses were conducted for each level of Plausibility, because the interaction of Plausibility and Transposition was significant from 850–1950 ms (cluster mass statistic = 371, $t = -3.88$, $p <$

6. For the response time data, we fitted a Gaussian model on log-transformed RTs, with the same fixed- and random-effect structure as the corresponding LME model. Sampling settings and convergence diagnostics were the same as for the accuracy model. Priors were Normal(7.1, 0.5) for the intercept, Normal(0, 0.3) for all regression coefficients, and Exponential(1) for the random-effect and residual standard deviations. The prior on the regression coefficients is weakly informative on the log-RT scale, placing 95% of its mass within ±0.6 log units, which is generous relative to the largest observed effect (0.09). Bayes factors were computed as Savage–Dickey density ratios. Because Bayes factors are sensitive to prior specification, we repeated the computation with Normal(0, 0.1) and Normal(0, 0.5) priors on the regression coefficients, yielding BF_{10} = 0.22 and 0.04 (BF_{01} = 4.6 and 22.4), respectively. The evidence for the null hypothesis was, thus, moderate to strong across prior specifications.

0.01). Note that this negative t-statistic is in an opposing direction to the one expected from the rational inference account.

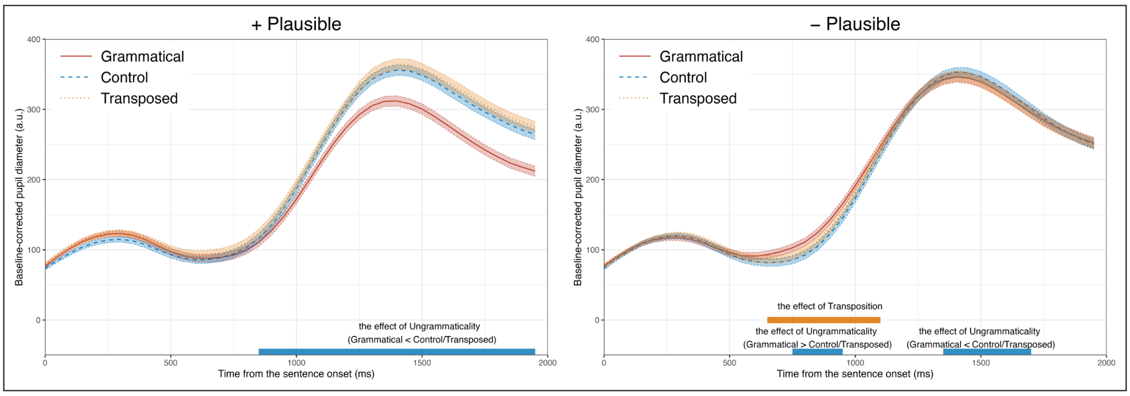

Figure 5: Changes in pupillary responses (as averaged over both eyes) in the +plausible (left) and −plausible (right) conditions. The x-axis indicates time relative to the target sentence onset (ms), and the y-axis indicates the relative change in pupil size. The shaded areas around the lines show standard errors calculated from participant means. The orange and blue lines on the bottom show time windows with significant clusters of the transposition and ungrammaticality effects, respectively. The larger pupillary responses in the −plausible/transposed condition, relative to the −plausible/control condition, illustrate transposed word effects (TWEs).

Crucially, in the −plausible conditions, a significant cluster of the Transposition factor was observed from 650 to 1100 ms (**Table 4**). In contrast, there was no significant effect of Transposition in the +plausible conditions.[7]

A significant cluster of the Ungrammaticality effect was observed from 750 to 950 ms (due to larger pupil responses in the grammatical condition) and from 1350 to 1700 ms (due to larger

Table 4: Summary of the significant clusters.

factors	**time windows**	**cluster mass**	***t***	***p***
Transposition in the −plausible conditions	650–1100 ms	64.3	1.77	< 0.001
Ungrammaticality in the −plausible conditions	750–950 ms	25.3	−1.58	< 0.01
Ungrammaticality in the −plausible conditions	1350–1700 ms	43.6	1.64	< 0.001
Ungrammaticality in the +plausible conditions	850–1950 ms	2015.0	6.15	< 0.001

7. We also conducted pupillometry analyses using only correct trials. The Transposition cluster in the −plausible condition was essentially unchanged (800–1150 ms, cluster mass statistic = 49.7, $t = 1.78$, $p < 0.001$). A further cluster emerged in the +plausible condition (1600–1750 ms, cluster mass statistic = 17.8, $t = 1.40$, $p < 0.01$), but it falls after the window in which responses were being made, and, thus, it is difficult to interpret.

pupil responses in the ungrammatical conditions) in the −plausible conditions (**Table 4**).[8] In the +plausible conditions, the Ungrammaticality effect was significant from 850 to 1950 ms.

3. Discussion

While it is well established that readers process multiple letters in a word simultaneously, controversy remains over whether they can process multiple words in parallel. One of the strongest pieces of evidence for parallel word processing comes from transposed word effects (TWEs), in which ungrammatical sentences with two adjacent words transposed, such as *I met my friend kind at the station*, are more difficult to judge as ungrammatical than control sentences. According to the parallel processing model, rapid parallel processing of multiple words, noisy position association of words with their location in a line of text, and top-down syntactic constraints conspire to create uncertainty in sentence representation, leading to TWEs. In contrast, the rational inference account proposes that TWEs are driven by a post-lexical inference mechanism that corrects transposed errors, based on the plausible meaning of the sentence.

This study examined the nature of TWEs by manipulating the plausibility of transposed sentences, in order to investigate the effect of rational inference on TWEs. In the experiment, grammatical, transposed, and control sentences were presented for 400 ms in a rapid parallel visual presentation (RPVP) paradigm, in which participants could not fixate on each word of the sentence sequentially within the presentation window. Participants were asked to judge the grammaticality of the sentences as quickly as possible, while pupillometric data were recorded as well as behavioural data, including response times (RTs) and accuracy.

3.1 Parallel processing of multiple words at a glance

The RT results replicated findings from previous studies that showed the transposed sentences were judged more slowly than the control sentences (Hossain & White, 2023; Mirault et al., 2018). Crucially, the replication was obtained under the RPVP paradigm, with sentences presented for only 400 ms. This experimental setting matters for the theoretical interpretation, as the TWE is not, by itself, diagnostic of parallel processing; the effect has been reported both when transposed sentences remain visible until a response is made (Mirault et al., 2018) and under RSVP (Hossain & White, 2023), and in both paradigms, words can be identified one at a time. In the present experiment, by contrast, serial fixation of all five words was not possible within the presentation window (see Section 1.4), so the TWE observed here arose under conditions in which words are unlikely to have been identified one at a time. This constrains accounts that assume serial word identification, such as the proposal that transposition errors are repaired by post-lexical inference over words held in a memory buffer (Hossain & White, 2023). Furthermore, the TWE

8. Within the −plausible condition, the early Ungrammaticality cluster (750–950 ms) showed the reverse pattern, compared with both the later cluster (1350–1700 ms) and the cluster in the +plausible condition (see Section 3.1).

was observed regardless of semantic plausibility, with no significant Plausibility × Transposition interaction. These results support the predictions made by the parallel processing model. This model proposes that multiple words are processed simultaneously with noisy position coding, creating temporary uncertainty about word order that requires additional processing time to resolve. The model predicts TWEs in both plausibility conditions, because a primitive syntactic representation is initially formed, regardless of its semantic plausibility, and provides top-down feedback to correct transposition errors. Our results confirmed this prediction.

The parallel processing account also explains the accuracy data, including the difference between the grammatical and the transposed conditions. Noisy position coding makes the encoded word order uncertain for both types of sentence, but the consequences differ, because the relationship between the bottom-up input and the top-down syntactic constraints differs between the two sentence types. For grammatical sentences, the two converge on the same representation, so that an occasional error in position coding is corrected by top-down feedback, and the sentence is accepted. For transposed sentences, by contrast, the bottom-up input is ungrammatical, whereas the top-down constraints favour the grammatical base sentence, and this conflict makes readers perceive the transposed sentences as if they were grammatical, leading to errors in the grammaticality judgement task. The asymmetry parallels the one found at the letter level, where *shpae* is sometimes perceived as *shape*, but *shape* is not misperceived as containing a transposition (Grainger & Van Heuven, 2004). The ungrammatical control sentences, by contrast, contain a violation that is independent of word order and can, therefore, be detected even when the encoded order is uncertain, which is why they were rejected more accurately than the transposed sentences. It should be stressed that it is this last comparison, between the transposed and the control conditions, that defines the TWE. The effect of Transposition on accuracy was significant and did not interact reliably with Plausibility, as expected under this account, on which a primitive syntactic representation is formed regardless of the semantic plausibility of the sentence.

Pupillometric data provided converging evidence for early parallel processing. In the −plausible condition, transposed sentences elicited larger pupil sizes than control sentences, with a significant Transposition cluster emerging at 650–1100 ms post-stimulus onset. Given that task-evoked pupillary responses typically peak around 900–1100 ms following the cognitive event (Hoeks & Levelt, 1993; Mathôt & Vilotijević, 2022), this latency suggests that the underlying cognitive conflict arose within the first few hundred milliseconds of sentence processing – well before response execution. This timing is consistent with rapid parallel word recognition combined with noisy position coding, suggesting that Japanese readers can process multiple words during initial sentence comprehension. Within a similar time window, the Ungrammaticality contrast in the −plausible condition ran in the opposite direction, with larger pupil sizes for the grammatical condition than for the two ungrammatical conditions between 750 and 950 ms, before the expected direction emerged at 1350–1700 ms (**Figure 5**). We have no account of the reversal. It should be noted, however, that the reversal does not bear on the TWE reported here; Ungrammaticality and Transposition were entered as orthogonal Helmert contrasts, with the grammatical condition coded 0 on the Transposition contrast, so the Transposition cluster at 650–1100 ms reflects the transposed–control difference alone.

A further finding that requires explanation is that the pupillometric TWE was observed only in the −plausible condition, and not in the +plausible condition. This pattern is not straightforwardly predicted by the parallel processing account, which expects TWEs regardless of semantic plausibility, and we acknowledge it as an unexpected finding that requires further investigation. Importantly, however, this pattern is also inconsistent with the rational inference account, which predicts the opposite directionality: larger TWEs for semantically plausible sentences. If rational inference were driving the pupillometric effects, we would expect greater cognitive effort for +plausible transposed sentences, as participants would be more strongly misled by semantic plausibility into perceiving these sentences as grammatical. The observed pattern, therefore, argues against the rational inference account on directional grounds, while leaving the mechanistic explanation of the plausibility modulation as an open question for future research. What remains clear from the pupillometric data is that the underlying cognitive conflict in the −plausible condition arose early in sentence processing (see Section 2.7.3 for the latency), which is consistent with rapid parallel word recognition, rather than with a post-lexical inference mechanism that operates after all words have been processed.

Relevant to the role of semantic plausibility in parallel processing, Massol et al. (2021) examined the effect of semantics on sentence superiority effects (SSEs). In their study, semantically natural (e.g. *Son amie danse bien* 'Her friend dances well') and anomalous yet grammatical (e.g. *Votre sac boit trop* 'Your bag drinks too much') sentences were presented for 300 ms in the RPVP paradigm and then native speakers of French were asked to identify a word that had appeared in a particular position in the sentence. The results showed superior word identification accuracy for semantically natural and anomalous yet grammatical sentences, compared to scrambled sentences (e.g. *Danse amie bien son* 'Dances friend well her'). Their findings suggest that the rapid partial processing of multiple words activates a sentence-level syntactic representation, which provides top-down feedback to ongoing word identification processes (Massol et al., 2021). In addition to this grammaticality effect, the interaction between grammaticality and semantic plausibility was significant, with semantically natural sentences showing an advantage for word identification over semantically anomalous sentences. The authors conclude that semantic plausibility plays some role in SSEs, but its contribution is relatively small compared to the syntactic contribution (i.e. a higher accuracy for the grammatical sentences, relative to the control sentences).

Parallel processing of multiple words has also been examined using the same-different matching task in recent behavioural and M/EEG studies (Dunagan et al., 2025; Flower & Pylkkänen, 2024; Pegado & Grainger, 2020; Pegado et al., 2021). In these studies, a 4- or 5-word sequence (referred to as *the reference*) was presented for a short period of time, followed by another sequence of words (e.g. grammatical, transposed, and replaced sentences), and then participants were asked to judge whether the second sequence was identical to the reference. Pegado and Grainger (2020) showed that the transposed sentences (i.e. *he wants these green apples → he these wants green apples*) made it more difficult to provide the "different" response than the replaced sentences (i.e. *he wants these green apples → he talks their green apples*). Interestingly, this TWE has been observed even when the transposed target sentences were created from the ungrammatical scrambled reference sentences such as *green wants these he apples → green these wants he apples*. Pegado and Grainger (2020) and Pegado et al. (2021) conclude, on the basis of these results, that the TWE in the same-different matching task is primarily driven by noisy bottom-up processes associating word identities with locations along a line of text, rather than by top-down syntactic processes. The rationale behind this interpretation is that if the top-down processes drive the TWE in the same-different matching task, a greater TWE should be observed in the grammatical reference condition than in the ungrammatical reference condition, due to a primitive sentence-level representation of the transposed target sentence misdirecting participants to make the "same" judgement. However, this is not consistent with their results.

Consistent with their view, Dunagan et al. (2025) propose that the same-different matching task allows readers to rapidly extract a rough sketch of syntactic information, but not to integrate lexical items into a detailed sentence-level representation. In their event-related potential (ERP) study, the processing of the reference sentences with a number agreement violation (*The dogs chases a ball*) was compared with that of the grammatical reference sentences (*The dogs chase a ball*). The ERP to the agreement error did not differ from the ERP to the grammatical sentences in any time window in their same-different matching experiment, despite the well-established observation that agreement errors elicit left anterior negativity and P600 effects in tasks that require participants to read sentences for comprehension (e.g. Molinaro et al., 2011; Osterhout & Mobley, 1995). Their results align with Pegado and Grainger's (2020) view that the same-different matching task does not require participants to engage in integrating lexical items into a sentence-level representation detailed enough to reject agreement violations.

Another possible source of the lack of agreement error effects is simply the low visual saliency of number affixes *-s*, which might be difficult to detect for a short presentation duration (Dunagan et al., 2025, p. 8). However, visual saliency is less likely to account for the results. In the present study, we found that the ungrammatical sentences with incorrect case particles (e.g. *-ni* 'DAT') induced more processing difficulty than the grammatical sentences, suggesting that Japanese readers were able to detect the visually less salient information and integrate it with a verb to form sentence-level representations even in the RPVP paradigm, although this detection was reliable rather than complete (see Appendix D).[9] The contrast between Dunagan et al.'s (2025) study and the present study might be related to different tasks tapping into different levels of sentence processing. In the same-different matching task, participants attempt to process sentences for better recall, but without a full-fledged syntactic analysis, resulting in failure to detect formal grammatical errors. On the other hand, the grammaticality judgement task (used in the present study) might require more detailed sentence-level processing.

It is worth noting one important difference between the present study and the previous study using MEG. Flower and Pylkkänen (2024) observed greater activity broadly over the left language area for *grammatical* sentences, such as *All cats are nice*, compared to transposed sentences, such as *All are cats nice*, in the early time window (210–230 ms post-stimulus onset), in contrast to the present study, in which transposed sentences were more difficult to process than grammatical sentences. Because these two studies differ in several dimensions, such as tasks (grammaticality judgement vs. same-different matching task), language (Japanese vs. English), and experimental methods (pupillometry vs. MEG), future research is necessary to investigate the underlying processing mechanisms responsible for these differences.

More recently, McGowan et al. (2026), through a series of behavioural experiments manipulating the temporal properties of word presentation, argued that parallel word recognition is not necessary to produce TWEs. They demonstrated that TWEs were observed under the RSVP paradigm, and that the magnitude of the effect is modulated by presentation rate. They interpreted these findings as supporting Huang and Staub's (2021b) proposal that TWEs may arise from a temporal overlap in the integration of adjacent words within a serial architecture. Our findings complement this line of work by addressing a related question. Whereas McGowan et al. (2026) demonstrated that parallel recognition of words is not necessary for TWEs, our RPVP design addresses whether parallel processing can produce TWEs. As discussed in Section 1.4, even under the most permissive strategic adjustments to the E-Z Reader model (Reichle et al., 2021), serial processing of five words within 400 ms remains implausible, requiring approximately 720 ms in total. Thus, even the covert serial attention shifts assumed by the E-Z Reader model would be severely constrained. Our observation of TWEs under these conditions, therefore, suggests that

9. One might question how Japanese readers can detect case-assignment violations in the control sentences if parallel processing creates imprecise word-order representations. This is possible because case assignment is independent of linear word order. Even if the word order of the control sentence 'suburb-GEN station-LOC partner-DAT kind drink-PST' is incorrectly represented (e.g. as 'partner-DAT kind suburb-GEN station-LOC drink-PST'), the sentence remains ungrammatical, because the monotransitive verb *drink* cannot assign a dative case to any argument, regardless of word order. Thus, while parallel processing with noisy position coding may cause readers to overlook word-order violations (as in the transposed sentences), it does not prevent them from detecting case-assignment violations (as in the control sentences). This distinction explains why the control sentences were reliably judged as ungrammatical, despite the RPVP presentation.

parallel processing can produce TWEs. Taken together with McGowan et al.'s (2026) findings, this suggests that noisy word order encoding may arise through multiple mechanisms, both rapid parallel word recognition under the RPVP paradigm and integration overlap under the RSVP paradigm.

3.2 Rational inference

Our study does not provide positive evidence for the rational inference account of TWEs. In the accuracy data, the LME analysis of the Plausibility × Transposition interaction did not reach conventional significance, and the Bayesian analysis provided anecdotal evidence for the null hypothesis ($BF_{10} = 0.34$). The direction of the interaction was numerically consistent with the rational inference account; participants were more likely to accept the transposed sentences when the base sentences were semantically plausible (Section 2.7.1). The effect was not, however, specific to transposition. The same effect was obtained for the ungrammatical control sentences. Plausibility increased the acceptance of ungrammatical sentences generally, rather than of transposed sentences specifically, which is also why the Plausibility × Transposition interaction was not reliable.

In the RT data, the Bayesian analysis provided strong evidence against the interaction ($BF_{10} = 0.07$), suggesting that plausibility did not modulate the time course of processing. Furthermore, the early time course of pupillometric TWEs suggests that TWEs do not primarily reflect a post-lexical memory effect. We note that the early latency of the pupillometric Transposition cluster is not, in itself, decisive against Huang and Staub's (2021b) version of the rational inference account, which allows inference to operate over a local window of unintegrated words, rather than only at the end of the sentence. However, this version of the account shares the directional prediction that TWEs should be enhanced in the +plausible condition, and our pupillometric data show the opposite directionality.

Nevertheless, our data do not exclude a contribution from rational inference. In the accuracy data the evidence against the interaction was only anecdotal ($BF_{10} = 0.34$), so the accuracy data might be better described as inconclusive than as evidence against the account. Note that the absence of a reliable interaction cannot be attributed to the possibility that the manipulation of plausibility was weak in the present experiment, because the effect of plausibility was significant in the grammatical sentences. Were the residual trend towards an interaction in the accuracy data reliable, it would suggest that rational inference operates at a late stage of processing, influencing the final grammaticality judgement, but not at the earlier stages reflected in response latencies and pupil size, as expected if inference acts on sentence representations held in memory after initial processing has occurred (Hossain & White, 2023).

This interpretation may account for why TWEs are often attenuated in the RSVP paradigm (Huang & Staub, 2023; Liu et al., 2022; Mirault et al., 2022b; Snell & Melo, 2024; Spinelli et al., 2024; Wen & Grainger, 2025). Snell and Melo (2024) found no TWE in RTs or accuracy (with significant TWEs in both RTs and accuracy in the parallel visual presentation (PVP) paradigm), while Liu et al. (2022), Mirault et al. (2022b) and Wen and Grainger (2025) found TWEs only in accuracy. Although Hossain and White (2023) and Spinelli et al. (2024) found significant TWEs in RTs, a recent meta-analysis by Wen and Grainger (2025, p. 5) demonstrated that TWEs in RTs were "statistically smaller in RSVP than the smallest effect estimated under parallel presentation" (Cohen's $d = 0.107$). Milledge et al. (2023), however, present a partial exception; while the accuracy effect was equivalent across PVP and RSVP presentations, the RT effect was significantly smaller under RSVP presentation than under PVP. The overall pattern of attenuation across these studies might arise because rational inference alone drives TWEs in the RSVP paradigm.

To date, only Hossain and White (2023) have reported comparable TWEs for RTs in the RSVP and PVP paradigms. The rationale for using the RSVP paradigm in their study is that if TWEs occurred even in the RSVP paradigm, which "enforces serial processing of words" (p. 2), it would provide evidence that parallel processing is not necessary for eliciting TWEs. However, this assumption is not warranted, because the RSVP paradigm does not necessarily force readers to process sentences word-by-word in a strictly serial manner; instead, readers can adopt a non-incremental integration strategy by remembering words, integrating them at the end of each sentence, and then judging their grammaticality (Hossain & White, 2023). Relevant to this discussion is a methodological difference between Hossain and White (2023) and the other previous studies cited above. In Hossain and White's (2023) experiment, the presentation speed of the RSVP was set to individuals' reading speed, determined in a pretest that used semantically coherent texts (approximately 203 ms per word on average, ranging from 140 to 322 ms). In contrast, other studies have presented words more slowly on average, i.e. 250–300 ms per word, fixed across participants (Huang & Staub, 2023; Liu et al., 2022; Mirault et al., 2022b). Thus, in Hossain and White's (2023) experiment, participants might have felt that the presentation speed of the RSVP paradigm was fast, especially because the target sentences were semantically more neutral, requiring more time for lexical processes compared to the semantically coherent text used in the pre-test. This could have encouraged them to adopt a non-incremental processing strategy. This conjecture is consistent with their result that faster readers (i.e. those for whom the presentation speed was faster in the main experiment) exhibited stronger TWEs in accuracy compared to slower readers. The integration of each word into the sentence representation might not have been completed before the subsequent word was encountered; as a consequence, all words were stored temporarily in memory. This temporary storage created some uncertainty about the temporal order of the words presented, making participants more susceptible to TWEs. Future studies are required to replicate the effect of presentation speed on TWEs in Japanese and other languages.

3.3 Limitations

The present study has several limitations. First, pupillometry does not allow us to directly determine which specific processing stages are reflected by task-evoked pupillary responses. Other techniques, such as EEG and MEG, should be considered. In these techniques, response-related activity requires special consideration, because the response execution generates large artefacts. In the same-different matching task, the response-related artefacts can be avoided by examining M/EEG responses to reference sentences, because they do not require overt responses (Pegado & Grainger, 2020; Pegado et al., 2021). However, this task might change the nature of sentence processing, as discussed above. Therefore, future studies should develop methods to investigate sentence-level representations without response-related artefacts.

Second, while our findings suggest that Japanese readers can process multiple words at a glance, as evidenced by TWEs in the RPVP paradigm, it remains unclear whether parallel processing is *necessary* to induce TWEs. To address this issue, strictly serial processing should be examined, but not in the RSVP paradigm, for the reasons discussed above. One possible approach is to use a moving window paradigm, in which the visual field is restricted to one word at a time, contingent on the participant's fixations (e.g. Osaka & Osaka, 2002).

Third, unlike Hossain and White (2023), we did not adjust the presentation duration of the target sentences to individual reading speeds. Thus, it is not clear to what extent the presentation duration of 400 ms was optimal for Japanese readers. Participants were reliably sensitive to word-order violations, though less so than to case-marking violations; *d*' was 1.29 for the transposed sentences and 1.83 for the control sentences in the +plausible condition, and 1.15 and 1.55, respectively, in the −plausible condition (see Appendix D). This pattern suggests that participants recognised individual words, but encoded their precise order less reliably, consistent with the parallel processing account with noisy position coding. However, we cannot rule out the possibility that a longer presentation duration might have allowed some participants to encode word order more accurately (see Appendix E for an exploratory analysis of this issue). Future studies should examine whether similar TWEs emerge with individualised presentation durations that allow for more complete sentence processing while still minimising serial word-by-word reading.

A further consideration concerns the possibility of a response bias in the grammaticality judgement task. The signal detection analysis reported in Section 2.7.1 confirms that such a bias was present; participants were conservative in responding "ungrammatical." The central finding of the present study cannot, however, be attributed to it. Crucially, sensitivity to the transposed sentences was reliably above zero (*d*' = 1.29 and 1.15), and the TWE was preserved when the data were expressed as sensitivity rather than as accuracy, with higher *d*' for the control condition than for the transposed condition at both levels of plausibility. Moreover, because the same grammatical baseline provides the false-alarm rate for both the control and transposed comparisons, the difference in *c* between the two conditions is algebraically determined by their difference in *d*' and, therefore, does not provide independent evidence for condition-specific response criteria. These results also speak against the more specific possibility that participants responded "ungrammatical" only when a violation involving the verb was detected; such violations occur only in the control condition, and this account predicts a sensitivity of 0 for the transposed condition.

What remains is the weaker possibility that the case-marking violation in the control sentences is simply easier to detect than the word-order violation in the transposed sentences, so that the difference between the two conditions reflects the detectability of the cue. Sensitivity to the case-marking violation was itself well below ceiling ($d' = 1.83$ in the +plausible condition and 1.55 in the −plausible condition), so the control sentences did not provide a cue that was detected on every trial, and the same asymmetry is inherent in the design of previous TWE studies, in which the control sentences also differ from the transposed sentences in the type of violation they contain (e.g. *The black was bear quietly*) (Hossain & White, 2023; Mirault et al., 2018). This possibility, nonetheless, cannot be excluded on the basis of the present data.

4. Conclusion

We examined the nature of transposed word effects (TWEs), with the aim of contributing to the parallel versus serial debate in reading. The behavioural results showed that Japanese readers found it difficult to reject transposed sentences and sometimes misclassified them as grammatical, which is consistent with previous studies in other languages. To our knowledge, this is the first demonstration of TWEs in Japanese. TWEs can, therefore, be obtained across orthographic systems, from alphabetic to mixed logographic/syllabary systems and from spaced to unspaced text, indicating that TWEs are not merely a surface-level phenomenon and that they call for a mechanistic, rather than an orthography-specific, explanation. Crucially, the RT and pupillometric data show that TWEs can be obtained under conditions in which the five words of a sentence are unlikely to have been identified one at a time, and that the conflict underlying the effect arises early in processing. Rapid parallel processing of multiple words is, therefore, sufficient to produce TWEs, although the present data do not establish that it is necessary. We found no positive evidence that semantic plausibility modulates the effect; the Plausibility × Transposition interaction was absent in the RT data, inconclusive in the accuracy data, and, in the pupillary data, ran opposite to the direction predicted by the rational inference account. These findings are difficult to reconcile with accounts on which the TWE depends on words having been identified one at a time, such as post-lexical inference over a memory buffer, although they do not exclude serial architectures in general.

Abbreviations

DAT = dative, GEN = genitive, LOC = locative, PST = past.

Data accessibility statement

Data, analysis scripts, and materials for this study are available at the Open Science Framework repository: https://osf.io/4q2tr/. This study is not preregistered. The Appendices are provided as supplementary files in the OSF repository.

Ethics and consent

The experimental protocol received approval from the Ethics Committee of Tokyo Metropolitan University (approval number: H7-008). The study was conducted in accordance with the Declaration of Helsinki, and written informed consent was obtained from all participants before the experiment.

Statement on the use of Generative AI

The authors used Claude Opus 5 (Anthropic, 2026) during the writing of the manuscript, solely for language editing purposes (grammar and spell checking).

Acknowledgments

We are very grateful to the anonymous reviewers for their insightful comments and suggestions. This article is based on the first author's undergraduate thesis submitted to the Department of Language Sciences at Tokyo Metropolitan University. This work was supported by JSPS KAKENHI Grant Numbers JP25K21849, JP25K00458, and JP23K20461 (PI: Masataka Yano).

Competing interests

The authors have no competing interests to declare.

Authors' contributions

Yukiho Kondo: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft; **Masataka Yano:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

ORCiD IDs

Masataka Yano: https://orcid.org/0000-0003-4465-8456

References

Alday, P. M. (2019). How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits. *Psychophysiology*, *56*(12), e13451. https://doi.org/10.1111/psyp.13451

Anthropic. (2026). Claude Opus 5 [Large language model]. https://www.anthropic.com/claude

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, *28*, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

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

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

Brisson, J., Mainville, M., Mailloux, D., Beaulieu, C., Serres, J., & Sirois, S. (2013). Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. *Behavior Research Methods*, *45*(4), 1322–1331. https://doi.org/10.3758/s13428-013-0327-0

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, *109*, 104047. https://doi.org/10.1016/j.jml.2019.104047

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review*, *117*(3), 713–758. https://doi.org/10.1037/a0019738

Declerck, M., Wen, Y., Snell, J., Meade, G., & Grainger, J. (2020). Unified syntax in the bilingual mind. *Psychonomic Bulletin & Review*, *27*(1), 149–154. https://doi.org/10.3758/s13423-019-01666-x

Dunagan, D., Jordan, T., Hale, J. T., Pylkkänen, L., & Chacón, D. A. (2025). Evaluating the timecourses of morpho-orthographic, lexical, and grammatical processing following rapid parallel visual presentation: An EEG investigation in English. *Cognition*, *257*, 106080. https://doi.org/10.1016/j.cognition.2025.106080

Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *Quarterly Journal of Experimental Psychology*, *63*(4), 639–645. https://doi.org/10.1080/17470210903469864

Fallon, J., & Pylkkänen, L. (2024). Language at a glance: How our brains grasp linguistic structure from parallel visual input. *Science Advances*, *10*(43), 9951. https://doi.org/10.1126/sciadv.adr9951

Flower, N., & Pylkkänen, L. (2024). The spatiotemporal dynamics of bottom-up and top-down processing during at-a-glance reading. *Journal of Neuroscience*, *44*(8), e0374242024. https://doi.org/10.1523/JNEUROSCI.0374-24.2024

Funasaki, N., & Yano, M. (2025). Role of prosody and word order in identifying focus: Evidence from pupillometry. *Language, Cognition and Neuroscience*, *40*(1), 23–40. https://doi.org/10.1080/23273798.2024.2396962

Gagl, B., Gregorova, K., Golch, J., Hawelka, S., Sassenhagen, J., Tavano, A., Poeppel, D., & Fiebach, C. J. (2022). Eye movements during text reading align with the rate of speech production. *Nature Human Behaviour*, *6*(3), 429–442. https://doi.org/10.1038/s41562-021-01215-4

Gagl, B., Hawelka, S., & Hutzler, F. (2011). Systematic influence of gaze position on pupil size measurement: Analysis and correction. *Behavior Research Methods*, *43*(4), 1171–1181. https://doi.org/10.3758/s13428-011-0109-5

Geller, J., Winn, M. B., Mahr, T., & Mirman, D. (2020). Gazer: A package for processing gaze position and pupil size data. *Behavior Research Methods*, *52*(5), 2232–2255. https://doi.org/10.3758/s13428-020-01374-8

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(20), 8051–8056. https://doi.org/10.1073/pnas.1216438110

Grainger, J., & Van Heuven, W. J. B. (2004). Modeling letter position coding in printed word perception. In *Mental lexicon: "Some words to talk about words"* (pp. 1–23). Nova Science Publishers.

Haro, J., Guasch, M., Vallès, B., & Ferré, P. (2017). Is pupillary response a reliable index of word recognition? Evidence from a delayed lexical decision task. *Behavior Research Methods*, *49*(5), 1930–1938. https://doi.org/10.3758/s13428-016-0835-9

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of *d*'. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. https://doi.org/10.3758/BF03203619

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*(3611), 1190–1192. https://doi.org/10.1126/science.143.3611.1190

Hoeks, B., & Levelt, W. J. (1993). Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research Methods, Instruments & Computers*, *25*(1), 16–26. https://doi.org/10.3758/BF03204445

Hossain, J., & White, A. L. (2023). The transposed word effect is consistent with serial word recognition and varies with reading speed. *Cognition*, *238*, 105512. https://doi.org/10.1016/j.cognition.2023.105512

Huang, K.-J., & Staub, A. (2021a). Using eye tracking to investigate failure to notice word transpositions in reading. *Cognition*, *216*, 104846. https://doi.org/10.1016/j.cognition.2021.104846

Huang, K.-J., & Staub, A. (2021b). Why do readers fail to notice word transpositions, omissions, and repetitions? A review of recent evidence and theory. *Language and Linguistics Compass*, *15*(7), e12434. https://doi.org/10.1111/lnc3.12434

Huang, K.-J., & Staub, A. (2023). The transposed-word effect does not require parallel word processing: Failure to notice transpositions with serial presentation of words. *Psychonomic Bulletin & Review*, *30*(1), 393–400. https://doi.org/10.3758/s13423-022-02150-9

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, *47*(2), 310–339. https://doi.org/10.1037/h0078820

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Liu, Z., Li, Y., Paterson, K. B., & Wang, J. (2020). A transposed-word effect in Chinese reading. *Attention, Perception, & Psychophysics*, *82*(8), 3788–3794. https://doi.org/10.3758/s13414-020-02114-y

Liu, Z., Li, Y., Cutter, M. G., Paterson, K. B., & Wang, J. (2022). A transposed-word effect across space and time: Evidence from Chinese. *Cognition*, *218*, 104922. https://doi.org/10.1016/j.cognition.2021.104922

Maris, E., & Oostenveld, R. (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

Massol, S., Mirault, J., & Grainger, J. (2021). The contribution of semantics to the sentence superiority effect. *Scientific Reports*, *11*(1), 20148. https://doi.org/10.1038/s41598-021-99565-6

Mathôt, S., & Vilotijević, A. (2022). Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis. *Behavior Research Methods*, *55*(6), 3055–3077. https://doi.org/10.3758/s13428-022-01957-7

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

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McGowan, V. A., Elsherif, M. M., Cutter, M. G., Liu, Z., Wang, J., & Paterson, K. B. (2026). Reading words of out order: Comparing serial and parallel accounts of the transposed word effect. *Cognition*, *269*, 106399. https://doi.org/10.1016/j.cognition.2025.106399

McLaughlin, D. J., Zink, M. E., Gaunt, L., Reilly, J., Sommers, M. S., Van Engen, K. J., & Peelle, J. E. (2023). Give me a break! Unavoidable fatigue effects in cognitive pupillometry. *Psychophysiology*, *60*(7), e14256. https://doi.org/10.1111/psyp.14256

Meade, G., Mahnich, C., Holcomb, P. J., & Grainger, J. (2021). Orthographic neighborhood density modulates the size of transposed-letter priming effects. *Cognitive, Affective, & Behavioral Neuroscience*, *21*(5), 948–959. https://doi.org/10.3758/s13415-021-00905-w

Meade, G., Grainger, J., & Holcomb, P. J. (2022). Language dominance modulates transposed-letter N400 priming effects in bilinguals. *Journal of Cognition*, *5*(1), 12. https://doi.org/10.5334/joc.203

Milledge, S. V., Bhatia, N., Mensah-Mcleod, L., Raghvani, P., A. McGowan, V., Elsherif, M. M., Cutter, M. G., Wang, J., Liu, Z., & Paterson, K. B. (2023). The transposed-word effect provides no unequivocal evidence for parallel processing. *Attention, Perception, & Psychophysics*, *85*(8), 2538–2546. https://doi.org/10.3758/s13414-023-02721-5

Mirault, J., Snell, J., & Grainger, J. (2018). You that read wrong again! A transposed-word effect in grammaticality judgments. *Psychological Science*, *29*(12), 1922–1929. https://doi.org/10.1177/0956797618806296

Mirault, J., Declerck, M., & Grainger, J. (2022a). Fast priming of grammatical decisions: Repetition and transposed-word priming effects. *Royal Society Open Science*, *9*(1), 211082. https://doi.org/10.1098/rsos.211082

Mirault, J., Vandendaele, A., Pegado, F., & Grainger, J. (2022b). Transposed-word effects when reading serially. *PLOS ONE*, *17*(11), e0277116. https://doi.org/10.1371/journal.pone.0277116

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*(8), 908–930. https://doi.org/10.1016/j.cortex.2011.02.019

Mukai, Y., Järvikivi, J., & Tucker, B. V. (2023). The role of phonology-to-orthography consistency in predicting the degree of pupil dilation induced in processing reduced and unreduced speech. *Applied Psycholinguistics*, *44*(5), 784–815. https://doi.org/10.1017/S0142716423000279

O'Connor, R. E., & Forster, K. I. (1981). Criterion bias and search sequence bias in word recognition. *Memory & Cognition*, *9*(1), 78–92. https://doi.org/10.3758/BF03196953

Osaka, N., & Osaka, M. (2002). Individual differences in working memory during reading with and without parafoveal information: A moving-window study. *The American Journal of Psychology*, *115*(4), 501–513. https://doi.org/10.2307/1423525

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*(6), 739–773. https://doi.org/10.1006/jmla.1995.1033

Pegado, F., & Grainger, J. (2020). A transposed-word effect in same-different judgments to sequences of words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(7), 1364–1371. https://doi.org/10.1037/xlm0000776

Pegado, F., Wen, Y., Mirault, J., Dufau, S., & Grainger, J. (2021). An ERP investigation of transposed-word effects in same-different matching. *Neuropsychologia*, *153*, 107753. https://doi.org/10.1016/j.neuropsychologia.2021.107753

Perea, M., & Carreiras, M. (2006). Do transposed-letter similarity effects occur at a syllable level? *Experimental Psychology*, *53*(4), 308–315. https://doi.org/10.1027/1618-3169.53.4.308

Perea, M., & Lupker, S. J. (2003). Transposed-letter confusability effects in masked form priming. In *Masked priming: The state of the art* (pp. 97–120). Psychology Press.

Perea, M., & Lupker, S. J. (2004). Can CANISO activate CASINO! Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language*, *51*(2), 231–246. https://doi.org/10.1016/j.jml.2004.05.005

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*(2), 275–280. https://doi.org/10.1037/h0027768

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*(1), 125–157. https://doi.org/10.1037/0033-295X.105.1.125

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*(4), 445–476, discussion 477–526. https://doi.org/10.1017/S0140525X03000104

Reichle, E. D., Yu, L., Liao, S., & Kruger, J.-L. (2021). Using simulations to understand the reading of rapidly displayed subtitles. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *43*. https://escholarship.org/uc/item/77f4h3qn

Schad, D. J., Vasishth, S., Hohenstein, S., & Kliegl, R. (2020). How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language*, *110*, 104038. https://doi.org/10.1016/j.jml.2019.104038

Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, and Psychophysics*, *74*(1), 5–35. https://doi.org/10.3758/s13414-011-0219-2

Snell, J., & Grainger, J. (2017). The sentence superiority effect revisited. *Cognition*, *168*, 217–221. https://doi.org/10.1016/j.cognition.2017.07.003

Snell, J., & Grainger, J. (2019a). Consciousness is not key in the serial-versus-parallel debate. *Trends in Cognitive Sciences*, *23*(10), 814–815. https://doi.org/10.1016/j.tics.2019.07.010

Snell, J., & Grainger, J. (2019b). Readers are parallel processors. *Trends in Cognitive Sciences*, *23*(7), 537–546. https://doi.org/10.1016/j.tics.2019.04.006

Snell, J., & Grainger, J. (2019c). Word position coding in reading is noisy. *Psychonomic Bulletin & Review*, *26*(2), 609–615. https://doi.org/10.3758/s13423-019-01574-0

Snell, J., Meeter, M., & Grainger, J. (2017). Evidence for simultaneous syntactic processing of multiple words during reading. *PLOS ONE*, *12*(3), e0173720. https://doi.org/10.1371/journal.pone.0173720

Snell, J., & Melo, A. N. (2024). Do love you me? Failure to notice word transpositions is induced by parallel word processing. *Journal of Cognition*, *7*(1), 21. https://doi.org/10.5334/joc.335

Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018). OB1-reader: A model of word recognition and eye movements in text reading. *Psychological Review*, *125*(6), 969–984. https://doi.org/10.1037/rev0000119

Spinelli, G., Yang, H., & Lupker, S. J. (2024). Rapid serial visual presentation of transposed-word sequences in the grammatical decision task: An examination of the roles of temporal and spatial cues to word order. *Language, Cognition and Neuroscience*, *39*(6), 773–792. https://doi.org/10.1080/23273798.2024.2360168

Staub, A. (2021). How reliable are individual differences in eye movements in reading? *Journal of Memory and Language*, *116*, 104190. https://doi.org/10.1016/j.jml.2020.104190

Staub, A. (2025). Sample size and its justification in the *Journal of Memory and Language*. *Journal of Memory and Language*, *142*, 104622. https://doi.org/10.1016/j.jml.2025.104622 van Rij, J., Hendriks, P., van Rijn, H., Baayen, R. H., & Wood, S. N. (2019). Analyzing the time course of pupillometric data. *Trends in Hearing*, *23*, 2331216519832483. https://doi.org/10.1177/2331216519832483

Voeten, C. C. (2022). Permutes: Permutation tests for time series data. https://cran.r-project.org/web/packages/permutes/index.html

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35–57. https://doi.org/10.3758/s13423-017-1343-3

Wen, Y., & Grainger, J. (2025). The transposed-word effect under serial presentation revisited: A replication study and meta-analysis. *Language, Cognition and Neuroscience*, *40*(8), 1156–1164. https://doi.org/10.1080/23273798.2025.2496910

Wen, Y., Mirault, J., & Grainger, J. (2021a). Fast syntax in the brain: Electrophysiological evidence from the rapid parallel visual presentation paradigm (RPVP). *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *47*(1), 99–112. https://doi.org/10.1037/xlm0000811

Wen, Y., Mirault, J., & Grainger, J. (2021b). The transposed-word effect revisited: The role of syntax in word position coding. *Language, Cognition and Neuroscience*, *36*(5), 668–673. https://doi.org/10.1080/23273798.2021.1880608

Wen, Y., Snell, J., & Grainger, J. (2019). Parallel, cascaded, interactive processing of words during sentence reading. *Cognition*, *189*, 221–226. https://doi.org/10.1016/j.cognition.2019.04.013

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, *1*(1), 59–85. https://doi.org/10.1016/0010-0285(70)90005-8

White, A. L., Palmer, J., & Boynton, G. M. (2018). Evidence of serial processing in visual word recognition. *Psychological Science*, *29*(7), 1062–1071. https://doi.org/10.1177/0956797617751898

White, A. L., Boynton, G. M., & Yeatman, J. D. (2019). You can't recognize two words simultaneously. *Trends in Cognitive Sciences*, *23*(10), 812–814. https://doi.org/10.1016/j.tics.2019.07.001

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends in Hearing*, *22*, 1–32. https://doi.org/10.1177/2331216518800869

Yano, M., Niikuni, K., Shimura, R., Funasaki, N., & Koizumi, M. (2025). Producing non-basic word orders in (in)felicitous contexts: Evidence from pupillometry and functional near-infrared spectroscopy (fNIRS). *Language, Cognition and Neuroscience*, *40*(1), 1–22. https://doi.org/10.1080/23273798.2024.2404178

Yokosawa, K., & Umeda, M. (1988). Processes in human kanji-word recognition. *Proceedings of the 1988 IEEE International Conference on Systems, Man, and Cybernetics*, *1*, 377–380. https://doi.org/10.1109/ICSMC.1988.754318