

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

Giacomo Spinelli<sup>1</sup>, Huilan Yang<sup>2</sup> & Stephen J. Lupker<sup>3</sup>

<sup>1</sup> University of Milano-Bicocca, Italy

<sup>2</sup> Zhejiang Gongshang University, China

<sup>3</sup> University of Western Ontario, Canada

Please cite as: Rapid Serial Visual Presentation of transposed-word sequences in the grammaticality judgment task: An examination of the roles of temporal and spatial cues to word order. *Language, Cognition, and Neuroscience*, 39, 773–792. <https://doi.org/10.1080/23273798.2024.2360168>

Running head: Temporal and spatial cues to word order

Corresponding Author:

Giacomo Spinelli

Department of Psychology, University of Milano-Bicocca

Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy

E-mail: giacomo.spinelli@unimib.it

Word count: 14312

## Abstract

Transposing two words in a sentence (e.g., “cat” and “was” in “the white cat was big”) creates a sequence that is harder to classify as ungrammatical than control sequences (e.g., “the white was cat slowly”), suggesting that word position coding is noisy and can be affected by syntactic expectations. In the present research, this transposed-word effect was examined more closely using Rapid Serial Visual Presentation (RSVP) formats which provided either clear temporal cues to word order but no spatial cues, or both types of cues. Compared to when all words were presented simultaneously, the two RSVP formats reduced the transposed-word effect to the same degree while having no parallel impact on another ungrammatical comparison condition involving no transposition. These results are discussed in the context of serial and parallel models of reading as well as models that propose a later processing stage for the locus of the transposed-word effect.

Keywords: serial processing; parallel processing; transposed-word effect; RSVP

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

A widely held opinion in psycholinguistics is that noise is an inherent aspect of language comprehension (e.g., Gibson et al., 2013; Norris & Kinoshita, 2012). For example, the finding that Transposed-Letter (TL) nonwords produced by transposing two letters in a word (e.g., “jugde” from “judge”) are harder to reject in a lexical decision task than are Substituted-Letter (SL) nonwords produced by substituting the same two letters with other letters (e.g., “jupte”) has been interpreted as evidence that there is noise in letter position coding, noise that would make “jugde” quite confusable with “judge” (e.g., Perea & Lupker, 2003; 2004).

A similar effect, one that suggests that there is noise in *word* position coding, is the Transposed-Word (TW) effect. This effect arises in grammatical decision tasks, tasks that require participants to judge the grammaticality of a visually presented sequence of words (Mirault et al., 2018; for a similar manipulation applied in the same-different task, see Pegado & Grainger, 2019). In Mirault et al.’s (2018) seminal studies, ungrammatical sentences were formed either by transposing the third and fourth words in a five-word grammatical sentence (e.g., “the white was cat big” created from “the white cat was big”), what we will call the Transposed-Word (TW) condition, or by transposing those words and additionally replacing the final word with a syntactically incompatible word (e.g., “the white was cat slowly”), what we will call the Transposed-Word, Substituted-Word (TWSW) condition. (The examples are in English, however, Mirault et al.’s experiments were conducted in French.) In two experiments, Mirault et al. found slower and less accurate responses to ungrammatical sentences in the TW condition than in the TWSW condition.

What this TW effect suggests is that, first and foremost, there is noise in word position coding, noise that would make “the white was cat big” confusable with “the white cat was big” (Snell & Grainger, 2019b). In addition, according to Mirault et al. (2018), this effect informs research on the long-standing debate in psycholinguistics concerning whether lexical processing of words in normal reading occurs serially or (to some extent) in parallel. Serial models (also known as serial-attention models) such as the E-Z Reader (Reichle et al., 1999, 2006, 2009b) assume a bottleneck in lexical processing limiting the allocation of attention for lexical processing to one word at a time (although lower-level processing, such as orthographic processing, would be possible for words ahead of the currently fixated word: Angele et al., 2013; Dare & Shillcock, 2013; see also Brothers et al., 2017). In contrast, parallel models (also known as attention-gradient models) such as SWIFT (Engbert et al., 2005), Glenmore (Reilly & Radach, 2006), or the more recent OB1-reader (Snell et al., 2018) reject this assumption, instead proposing that multiple words can be processed at a time, with a visuospatial attention gradient reducing the degree of processing for words at more eccentric locations than the fixated location.

The reason that the TW effect is potentially important in the serial vs. parallel processing debate is that it would only seem compatible with parallel models, particularly the OB1-reader (Snell et al., 2018). The reasoning is that, because words are processed sequentially in serial models, ungrammaticality should be detected as soon as the first of the transposed words is processed if that word is syntactically incorrect. As a result, judging sequences containing such transpositions as ungrammatical should produce no TW effect.

In contrast, the OB1-reader assumes that the multiple words being processed during reading would be mapped onto a spatiotopic representation in which their respective positions are coded. Position coding, however, is noisy, which can prevent the accurate determination of the positions of the words being read. Ultimately, the positions that the words are mapped onto are determined by 1) which word reaches the recognition threshold first (with words at the fixed location being more likely to “beat”

words at more eccentric locations), 2) bottom-up factors (e.g., word length) and, importantly, 3) top-down factors (e.g., syntactic constraints). The latter factors would be crucial in producing the TW effect because they could induce readers to misrepresent TW sequences as their grammatical base sentences, either causing errors in grammaticality judgments or slowing down the mapping process, delaying an “ungrammatical” response.

For example, because the expected syntactic sequence for the first four words in the TW sequence “the white was cat big” is 1) article, 2) adjective, 3) noun, and 4) verb, the system may attempt to map “was” and “cat” onto the syntactically expected positions (4 and 3, respectively). The mapping of “big” to the final position reinforces the possibility that “was” actually was in position 4 and “cat” actually was in position 3. The result would be either an incorrect “grammatical” decision or a correct, but slower, “ungrammatical” decision if the actual positions of the transposed words are ultimately correctly determined. Although a similar process may occur with TWSW sequences, the mapping of “slowly” to the final position indicates that, regardless of whether the sequence is interpreted as “the white was cat slowly” or “the white cat was slowly”, neither is a grammatical sentence (although the latter is a grammatical *sequence*). Hence, an “ungrammatical” decision can be made more easily even though the positions of “was” and “cat” were noisy.

As argued by Huang and Staub (2021b; see also Huang and Staub, 2021a, 2023), however, the TW effect really only poses a challenge to serial models if a strong assumption is made by those models. That assumption is that the process of syntactic and semantic integration of a word into the sequence that has been processed to that point does not start until the previous word has been fully integrated. Although the original E-Z Reader (Reichle et al., 2009b) did make this perfect incrementality assumption for the integration process, the assumption was not a fundamental one and could be relaxed.

Huang and Staub (2021b), in particular, proposed a modified version of the E-Z Reader in which, although word recognition would still proceed incrementally (i.e., the recognition process for a word would not start until the previous word had been recognized), the system would be able to hold on to multiple fully processed words before integrating them into the sequence processed up to that point. For example, there would be a point during processing of the TW sequence “the white was cat big” at which recognition of “was”, “cat”, and “big” (as well as that of the preceding words) was complete but those three words had not yet been integrated into the processed sequence. At that point, a rational-inference process would be engaged which might result in the conclusion that the order of those words was “cat was big”, the most likely continuation of “the white”, instead of “was cat big”, the actual order. According to this rational-inference idea, in general, comprehenders would rationally infer an intended message from a perceived message based on the assumption that various sources of noise might have corrupted the perceived message (Gibson et al., 2013). The rational inference, in the present case, would be influenced by readers’ knowledge of what is a typical grammatical sentence as well as the type of errors that might occur in the communication process, including transposition errors in generating a sentence or in perceiving it. The result is that, when a sequence such as “was cat big” is to be integrated with “the white”, readers may infer that the intended sequence was the more likely sequence “cat was big” assuming, for example, that an error was committed in perceiving the order of the words. This inference would not be made with “was cat slowly” because it is less likely that the full intended sequence was “the white cat was slowly”. Importantly, this inference would be applied even if the sequence was processed serially. The essential conclusion is that the TW effect is not actually incompatible with serial models in contrast to Mirault et al.’s (2018) claims.

In fact, the possibility cannot be excluded that the TW effect has actually nothing to do with the serial vs. parallel nature of lexical processing. On this point, it is worth noting that the distinction between serial and parallel models is blurrier than the terms suggest. On one hand, parallel models have a serial

component because they assume a limit for the words being processed simultaneously (e.g., five words in the OB1-reader; Snell et al., 2018), meaning that processing for longer sentences would occur in at least two batches, one after the other. On the other hand, serial models such as Huang and Staub's (2021b) seem to create a parallel component with the assumption of a buffer in which sequences of multiple words are evaluated before integration with the sequence processed up to that point.

More centrally, however, while both serial and parallel models have assumed that the TW effect would occur *during* lexical processing, thus potentially providing a window on the nature of the processes involved at that stage, the possibility exists that the effect may even occur *after* all the words in the sequence have been processed, that is, during the time when the constructed representation of the sequence is being evaluated. That is, a sequence that could be made grammatical (and meaningful) like "the white was cat big" may be perceived correctly but take longer to classify as ungrammatical due to the fact that the system is aware of its own inherent noisiness. The serial vs. parallel nature of lexical processing, in that case, would be completely irrelevant.

### *The present research*

Although, for the reasons noted above, it is debatable whether the TW effect is informative with respect to the serial vs. parallel nature of lexical processing in sentence reading, what does not appear to be debatable is that the effect reflects the fact that word position coding is noisy. As noted, noisy word position coding is a fundamental assumption of the OB1-reader, an assumption that, according to Snell et al. (2018), would allow parallel models to account for the problem of determining word order when multiple words are processed simultaneously. Word-order noise is also what makes rational inference possible in Huang and Staub's (2021b) serial model: If there was certainty that "was cat" instead of "cat was" was perceived when reading a TW sequence such as "the white was cat big", there would be no reason to assume otherwise. As noted, even a model assuming a decision-stage locus for the TW effect

would presumably want to make the assumption that the process that initially derives word order is noisy because a TW effect would not be produced if there were complete certainty about word order.

The present research was an attempt to examine more fully the idea of word-order noise being important in the TW effect using different presentation formats intended to manipulate the level and sources of that noise. The TW effect has typically been examined in “simultaneous” presentation formats, formats in which all the words composing the sentence are presented at the same time (Huang & Staub, 2021a; Liu et al., 2020, 2021; Mirault et al., 2018, 2020, 2022a, 2022b; Snell & Grainger, 2019b; Wen et al., 2021a, 2021b). In that format, word order can presumably be derived from spatial information because, in a left-to-right writing system, the word occupying the leftmost position is the first word in the sequence, the adjacent word to the right is the second word, etc. That is, there are clear spatial cues to word order. In contrast, word order cannot be derived as easily from temporal information because all the words composing the sequence are presented at the same time. Ambiguous temporal cues, rather than ambiguous spatial cues, would, therefore, seem to be the main potential source of word-order noise contributing to the TW effect in the simultaneous presentation format.

Temporal word-order noise may not be excessive in normal reading, as opposed to in a timed response task, given that processing one word at a time (i.e., essentially making reading a serial process as assumed by serial models) would allow readers to generate their own temporal cues (e.g., the first word is the one that was processed first, the second word is the one that was processed next, etc.). However, neither temporal nor spatial noise would likely ever be zero, explaining why TW effects can be produced. Temporal word-order noise would, presumably, be quite large in general according to parallel models such as the OB1-reader because processing of the words composing the sequence would largely overlap in time (e.g., the word in position  $n$  would reach the recognition threshold at approximately the same time as the words in positions  $n - 1$  and  $n + 1$ , making it difficult to determine their order).

However, because the task used in these types of experiments is a timed response task one would

assume that temporal word-order noise would be problematic for the experiment's participants according to any model attempting to explain TW effects.

A straightforward approach to determine the importance of temporal word-order noise to the TW effect is to use a presentation format that makes temporal word order clearer than in the simultaneous presentation format. Such a format is offered by the Rapid Serial Visual Presentation (RSVP) procedure which typically involves presenting the individual words composing a sequence one after another (centrally) for a short (100–400 ms) duration each. Because the words are presented one after another in this procedure, their order in the temporal sequence should be substantially clearer. If the procedure were to also involve the presentation of the words in the same positions as in the simultaneous presentation format (i.e., a spatial RSVP format, see the middle panel in Figure 1), there should also be clear spatial cues as well.

With clear temporal and spatial cues (as opposed to there only being reasonably clear spatial cues, as in the simultaneous presentation format), all relevant models would seem to predict a benefit (i.e., a reduced TW effect). For example, it can be presumed that serial models such as Huang and Staub's (2021) would need not store a word (e.g., "was") in a buffer while waiting for recognition of the subsequent words (e.g., "cat" and either "big" or "slowly") in order to proceed to integrating the first word (i.e., "was") into the sequence representation (Liu et al., 2022). That is, integration would occur in a fashion that is closer to the perfectly incremental fashion assumed by the original E-Z Reader, leaving little or no chance for rational-inference processes to (inappropriately) "correct" the word order from "was cat big" to "cat was big". Transpositions would thus be easier to detect and TW effects reduced. Concerning models that assume parallel lexical processing in normal reading, because the words composing the sentence are presented one after another in RSVP procedures instead of simultaneously (i.e., serial processing is enforced), determining their order would also be presumed to be less of a challenge. The implication is that in models such as the OB1-reader word position coding should be less

noisy and TW effects, again, reduced. The same conclusion would apply for models that assume a decision-stage locus for TW effects: If temporal cues to word order are important, having clearer temporal cues (as in the central RSVP procedure) or clearer temporal and spatial cues (as in the spatial RSVP procedure) would allow more confidence in making order-based grammaticality decisions, reducing the TW effect.

The presentation procedures used in the present research were designed to provide an evaluation of the impact of the two types of cues. That is, in the simultaneous presentation procedure in a timed responding task, temporal cues are weak and TW effects are large. In the central RSVP procedure the words composing the sequence are presented centrally, depriving them of any spatial cues to order (see the right panel in Figure 1) while at the same time providing clear temporal cues. The spatial RSVP procedure provides both types of cues. If temporal cues help reduce word-order noise, one would expect smaller TW effects in the RSVP procedures than in the simultaneous procedure, as has already been reported (Huang & Staub, 2023; Liu et al., 2022; Mirault et al., 2022b). Further, if spatial cues, despite being noisy, also contribute to reducing word-order noise, one would expect even smaller TW effects with the spatial RSVP procedure. In contrast, if the spatial cues are not used to any large extent due to their noisiness, one would expect that performance in the spatial RSVP procedure would match that in the central RSVP procedure. Most, if not all, models of sentence processing would then be able to adjust their assumptions to emphasize the role of temporal cues in sentence processing (although, as discussed in the General Discussion, some models may need no such adjustment).

In sum, an examination of the TW effect in RSVP formats appears to be informative in terms of the level of word-order noise associated with temporal and/or spatial information. In the present experiments, we conducted such an examination (for a similar examination, see Milledge et al., 2023). In Experiment 1, Mirault et al.'s (2018) TW manipulation was replicated, in an English grammatical decision task, using the simultaneous presentation format. In Experiment 2, the same materials were presented in a spatial

RSVP format in which the words were presented one after another in the same positions as in the simultaneous presentation format, thus maintaining the spatial relationships between the words, and in a central RSVP format in which the words were presented one by one at the same rate as in spatial RSVP format but centred on the screen, thus eliminating the spatial relationships between the words.

An additional novel aspect of the present experiments is the inclusion of a control condition that has not been used in previous experiments with the grammatical decision task. As noted, in their seminal studies, Mirault et al. (2018) used, as a control condition, a TWSW condition obtained by transposing the third and fourth words in a five-word grammatical sentence (as in the TW condition) and additionally replacing the final word with a syntactically incompatible word (e.g., “the white was cat slowly”). The reasoning is that, although “correcting” the transposition in the TW condition results in a grammatical sentence (“the white cat was big”), doing so in the TWSW condition, instead, results in an ungrammatical sentence differing from the grammatical sentence only by the nature of the final word (“the white cat was slowly”). Sentences of this sort (i.e., “the white cat was slowly”) were used to create what we will call a Substituted-Word (SW) condition. Although this SW condition involves ungrammatical sentences just like the TW and TWSW conditions, to our knowledge, grammaticality judgments for this condition have never been examined. Such an examination would prove informative for a couple reasons.

First, the SW condition would allow the examination of the contrast between transposing two words in a sequence (in the TW and TWSW conditions) vs. only substituting one word (in the SW condition). With respect to the TW-SW contrast, if TW sequences are inevitably misrepresented as their base grammatical sentences at the point that the fifth word in the sequence is processed, it would seem to follow that those sequences should be harder to reject as ungrammatical compared to SW sequences, sequences which should not be misrepresented as grammatical sentences once their fifth word has

been processed (but for a discussion of the idea that SW sequences may be represented as “incomplete” grammatical sentences, i.e., grammatical *sequences*, see below).

On the other hand, it seems unlikely that TW sequences would inevitably be misrepresented as their base grammatical sentences, especially in an experiment in which ungrammatical sequences often involve word transpositions and, therefore, precise word position coding would be important. That is, on some TW trials, the ungrammaticality may be clearly perceived at an earlier point. On those trials, TW sequences should be, if anything, easier to classify as ungrammatical compared to SW sequences because the point at which those sequences become ungrammatical occurs earlier in TW sequences (potentially, as early as with the third word, the first transposed word). Therefore, overall, the pattern of the TW-SW contrast would seem to give us some idea of the frequency with which the lack of grammaticality created by transposing the two words in TW sequences is readily perceived.

In contrast, with respect to the TWSW-SW contrast, any time a TWSW sequence is misperceived (e.g., when “the white was cat slowly” is perceived as “the white cat was slowly”), no difference between the two types of sequences is to be expected. However, given that, as with TW sequences, at least part of the time the ungrammaticality in TWSW sequences would be discovered when processing the words “cat” and “was”, it would seem to follow that those sequences should be easier to reject as ungrammatical than SW sequences. Therefore, overall, TWSW sequences should elicit, if anything, faster responses than SW sequences with the TWSW-SW contrast providing some information as to the frequency with which the lack of grammaticality created by transposing two words in TWSW sentences is readily perceived.

A second reason why the SW condition might be informative relates to the RSVP formats examined in the present experiments. As noted, the main impact of RSVP formats is to reduce word-order noise by providing clear temporal cues. The main impact of this reduction, in turn, can be presumed to be on

ungrammatical sentences containing transpositions, that is, for TW and TWSW sequences in which one of the grammatical violations was created by altering the order of two words in a base grammatical sentence. When word order is made clearer, detecting those violations would be generally easier. Hence, categorizing those sequences as ungrammatical would be easier (i.e., faster and more accurate) compared to simultaneous presentation formats in which word order is not as clear.

In contrast, for SW sequences, reducing word-order noise by providing clear temporal cues would be expected to have almost no impact. The reason is that the grammatical violation in those sequences was not created by altering the order of two words in a base sentence but rather by the substitution involving the fifth (i.e., the final) word. As a result, in RSVP formats, SW sequences would remain grammatical until the fifth word has been processed (because the first four words, “the white cat was”, are identical to those in the base sentence) or, in certain situations, until it is realized that there is no sixth word that would complete the sentence and make it grammatical (e.g., “the white cat was slowly *approaching*”). Therefore, categorizing those sequences as ungrammatical would be no easier in RSVP formats compared to simultaneous presentation formats even though word order is clearer in the former than in the latter.

If anything, it is the simultaneous presentation format the one in which the “ungrammatical” decision may be easier for the SW condition. That is, readers may manage to process the fifth word earlier in simultaneous presentation formats than would be possible in RSVP formats, because in RSVP formats the point in time at which each word can be processed is constrained by the presentation rate used. If, for example, a 250-ms presentation rate is used in an RSVP procedure, the fifth word in a sequence would not be available for processing until 1000 ms had elapsed. In contrast, the fifth word may be processed earlier in simultaneous presentation formats (according to parallel models such as the OB1-reader, much earlier, i.e., almost immediately), in which case the ungrammatical status of the sequence could be determined faster. An additional advantage in simultaneous presentation formats for the SW

condition is that it is clear which word is the final word in the sequence and, hence, a word such as “approaching” would not be following “The white cat was slowly”. The main point here is that the impact of RSVP formats for SW sequences can be presumed to be different than that for TW and TWSW sequences. The inclusion of the SW condition in the present experiments allowed us to examine these ideas.

As the present experiments were in progress, we became aware that a few similar experiments had recently been conducted in other laboratories. Huang and Staub (2023) found, in English, that grammatical decisions to sequences in which two adjacent words were transposed were more accurate in central RSVP formats than in simultaneous presentation formats. However, they did not either include a TWSW condition in their experiment or consider their RT results. In experiments in which the authors did do both of those things, Liu et al. (2022) reported that, in Chinese, the standard (central) RSVP format, compared to the simultaneous presentation format, eliminated TW effects in the RTs and reduced them in the error rates, although they were still significant in the error rate data. Mirault et al. (2022b) obtained similar results in a series of experiments with central RSVP formats in French. More recently, Milledge et al. (2023), using a central RSVP format, reported a somewhat complementary pattern in English, with a reduction (without complete elimination) of the TW effect emerging in RTs only and no difference with the simultaneous presentation format emerging in the error rates. Milledge et al. also included a spatial RSVP format similar to ours and found that although that format produced a smaller TW effect than the simultaneous presentation format, their TW effect was actually slightly larger than that produced by the standard (central) RSVP format (in the RTs; again, no difference emerged in the error rates). The general pattern, in any case, is that TW effects tend to be reduced, but not completely eliminated, in RSVP formats.

The present experiments provide an opportunity to corroborate and extend the prior conclusions, first, with the inclusion of an additional control condition, the SW condition, which has not been used so far

but which, for the reasons explained above, appears to have informative value. Second, what also needs to be noted is that the TWSW condition, the original control condition in Mirault et al.'s (2018) seminal experiments, one that was also included in the present experiments, is a condition which was not included in previous grammatical decision tasks in English except in Milledge et al.'s (2023). Huang and Staub (2021a) did not use a control condition, whereas Huang and Staub (2023) used a "scrambled" control condition which, as with their experimental condition, was also created by transposing words in a grammatical base sentence. However, in that control condition, the transpositions involved non-adjacent words and more words in general were transposed in that condition than in the experimental condition, e.g., "the was white big cat". Note that scrambled sequences of this sort are in fact more extreme versions of TW sequences, making the essence of the contrast between scrambled and TW sequences unclear.

Another generally novel aspect of the present experiments (although see Milledge et al., 2023), is combining the spatial RSVP task with the standard (central) RSVP task that has been used in all previous experiments, allowing a direct comparison between those two presentation formats. Further, because we included both of the RSVP formats in the same experiment (i.e., Experiment 2), that within-subject comparison is likely stronger than Milledge et al.'s, who employed spatial and central RSVP formats in separate experiments.

To sum up, in Experiment 1, the simultaneous presentation format was used in an attempt to replicate, in English, the pattern of results obtained by Mirault et al. (2018) in French, i.e., slower and less accurate responses to TW sequences compared to TWSW sequences, a TW effect (see also Milledge et al., 2023), with the results in the SW condition giving us some additional insight in interpreting those results. In Experiment 2, RSVP formats were used in an attempt to determine whether those formats would produce a reduction of the TW effect on one hand (due to there being clear temporal cues to word order) and, on the other hand, a general speed up of responses to TWSW and TW sequences compared

to SW sequences relative to what was observed in the simultaneous presentation format (due to the increased clarity of word order making it easier to detect the transpositions in TWSW and TW sequences). That experiment also included a contrast between spatial and central versions of the RSVP format that maintained vs. eliminated the spatial relationships between the words in the simultaneous presentation format. If spatial cues to word order, in addition to temporal cues, provide any help at all in determining word order, the TW effect should be smaller in the spatial RSVP format than in the central RSVP format.

### **Experiment 1 (simultaneous presentation format)**

#### *Method*

##### Participants

Sixty-three students (age 18 –20 years) at the University of Western Ontario participated in this experiment, which was conducted in the laboratory, for course credit. All had normal or corrected-to-normal vision and were native speakers of English. An equal number of participants was assigned to each counterbalancing list (see below). As described below, the number of observations in each cell of the design was 32, meaning that, with 63 participants, we had a total of 2016 observations for each condition, a number exceeding 1600, the number recommended by Brysbaert and Stevens (2018) for properly powered response time experiments with a repeated-measures design.

##### Materials

The ungrammatical sentences were created using a procedure similar to that used by Mirault et al. (2018). First, 48 pairs of grammatical five-word English sentences, the base sentences, were created. Each sentence was created so that transposing the third and fourth word produced an ungrammatical sentence (i.e., the TW condition). Further, the two sentences in each pair were arranged so that

substituting the final word in one sentence with the final word of the other sentence (and vice versa) also produced an ungrammatical sentence (i.e., the SW condition). Finally, applying both the transposition and the substitution produced another ungrammatical sentence (i.e., the TWSW condition). That is, for each pair of base sentences (e.g., “her skinny uncle criticizes everyone” and “a lonely grey wolf howled”), 6 ungrammatical sentences were created: 2 TW sequences (e.g., “her skinny criticizes uncle everyone” and “a lonely wolf grey howled”), 2 SW sequences (e.g., “her skinny uncle criticizes howled” and “a lonely grey wolf everyone”), and 2 TWSW sequences (e.g., “her skinny criticizes uncle howled” and “a lonely wolf grey everyone”). In total, 96 TW, 96 SW, and 96 TWSW sequences were created.

The assignment of base sentences to TW, SW, and TWSW conditions was counterbalanced across participants using 3 lists, each containing 32 TW sequences, 32 SW sequences, and 32 TWSW sequences, in order to present each participant with a single ungrammatical sentence (either TW, SW, or TWSW) for each base sentence. The words in the ungrammatical sentences were 1 to 13 letters long, had an average length of 5.14 and an average frequency of 3531.48 occurrences per million (median = 90.89 occurrences per million; Brysbaert & New, 2009), which is equivalent to 5.12 Zipf (median = 4.95 Zipf; van Heuven et al., 2014). Note that the three conditions did not differ in these average values because the same words were used to create them. Ninety-six grammatical sentences with similar characteristics as the ungrammatical sentences were created and included in all of the counterbalancing lists.

### Procedure

The procedure, illustrated in the left panel of Figure 1, was similar to Mirault et al.’s (2018). Each trial began with a fixation symbol (“+”) displayed in the centre of the screen for a random time period of between 500 and 700 ms, followed by the five-word sequence, which was also centred on the screen, with the initial letter of the first word in the sequence being capitalized. The sequence remained on the

screen until the response or until the time limit, which was 20 seconds, expired.<sup>1</sup> Participants were instructed to indicate as quickly and as accurately as possible as to whether the sequence formed a grammatical sentence or not by pressing the right shift key if it was a grammatical sentence or the left shift key if it was an ungrammatical sentence. After the response, a green/red “o” in bold font was presented for correct and incorrect responses, respectively, for 700 ms (in the rare case of a null response, the phrase “No response” was displayed). Immediately following the feedback message, the next trial began. All stimuli were presented in size 11 Courier New black font (except for the feedback “o”, which was either green or red in bold, as noted) against a white background. The order of presentation of the stimuli was randomized for each participant.

All participants received 12 practice trials involving a novel set of sentences (6 grammatical and 6 ungrammatical) prior to the 192 experimental trials. The entire session lasted approximately 12 minutes. DMDX software (Forster & Forster, 2003) was used to present the stimuli and collect responses. This and the following experiment were performed in line with the principles of the Declaration of Helsinki and were approved by the University of Western Ontario Research Ethics Board (Protocol # 108956). Informed consent was obtained from all individual participants included in the experiments.

(Figure 1 about here)

## *Results*

Data pre-processing and analysis were conducted in R version 3.6.3 (R Core Team, 2020) in a similar way as that used by Mirault et al. (2018). Because all participants had an accuracy above 70% ( $M = 92.15\%$ ,  $SD = 3.77$ ), no participant was excluded. There was one null response in the entire dataset, which was removed prior to all analyses. For the latency analysis, correct response times below or above 2.5 standard deviations from the grand mean (corresponding to 2.33% of the correct responses) were

removed as outliers. For the ungrammatical sentences, the remainder of the correct responses and the error rates were analysed using a generalized linear mixed-effects model (GLMM) in which Subjects and Items (the target stimuli) were random effects, and Condition (TWSW vs. SW vs. TW, within-subject and within-item) was a fixed effect.

In the latency analyses, a GLMM, a generalized linear mixed-effects model, was used instead of a linear mixed-effects model, the type of model that is more commonly used for latencies, because linear models, unlike generalized linear models, assume normally distributed residuals. Because the typically positively skewed distribution of raw RTs fails to accommodate that assumption, the common practice in linear model analyses is to normalize raw RTs with a reciprocal transformation (e.g.,  $\text{invRT} = 1000/\text{RT}$ ), which is indeed the procedure used by Mirault et al. (2018). However, nonlinear transformations such as a reciprocal transformation systematically alter the pattern and size of interaction effects, rendering such transformations inappropriate when the research interest lies in interactions (Balota et al., 2013). Although the analyses for Experiment 1 did not involve an interaction term, such terms were the focus of subsequent analyses reported in this paper. For this reason, consistent with recent practices (e.g., Cohen-Shikora et al., 2019; Lo & Andrews, 2015; Yang et al., 2019), we used a GLMM because those models, unlike linear models, do not assume normally distributed residuals and can, therefore, better accommodate the distribution of raw RT data without requiring a transformation of those data.<sup>2</sup>

A Gamma distribution was used to fit the raw RTs, with an identity link between fixed effects and the dependent variable (Lo & Andrews, 2015). In the current version of lme4 (the R package used for running the GLMMs), convergence failures for GLMMs are frequent (although many of those failures reflect false positives: Bolker, 2022). In order to limit the occurrence of those failures, the random structure of the model was kept as simple as possible by using only random intercepts for Subjects and

Items. For the same reason, the model was run specifying a maximum number of one million iterations and using the BOBYQA optimizer, an optimizer that typically returns estimates that are equivalent to those returned by lme4's default optimizer but that results in fewer convergence failures (see, e.g., Colombo et al., 2020; Lupker et al., 2020a, 2020b).<sup>3</sup>

Prior to running the model, R-default treatment contrasts were changed to sum-to-zero contrasts (i.e., `contr.sum`) in order to evaluate the main effect of Condition. The model was fit by maximum likelihood with the Laplace approximation technique. The lme4 package, version 1.1-23-1 (Bates et al., 2015), was used to run the GLMM. The function `Anova` in the car package, version 3.0-7 (Fox & Weisberg, 2016), was used to obtain estimates and probability values for the fixed effects specifying Type III Sums of Squares. Pairwise comparisons for the levels of the Condition factor were conducted using the emmeans package, version 1.4.6 (Lenth, 2018), with Tukey's HSD adjustment for multiple comparisons. Mean response latencies and error rates for each condition for the ungrammatical sentences are reported in Table 1. For this and the following experiments, the raw data and R files used for the analyses are publicly available at <https://osf.io/q7yd3/>. A spreadsheet with the word sequences used in the present experiments is also available at that link. The scripts for running the experiments are available upon request. The experiments were not preregistered.

(Table 1 about here)

### Ungrammatical sentences

*RTs.* The main effect of Condition was significant,  $\chi^2 = 23.16$ ,  $p < .001$ . Replicating Mirault et al. (2018), the TWSW condition (the control condition in their experiments) was significantly faster than the TW condition (a TW effect),  $\beta = -61.2$ ,  $SE = 12.9$ ,  $z = -4.75$ ,  $p < .001$ . The TWSW condition was also significantly faster than the SW condition (the new condition),  $\beta = -44.0$ ,  $SE = 13.4$ ,  $z = 3.29$ ,  $p = .003$ .

Although the TW condition was numerically (21 ms) slower than the SW condition, the difference was not significant,  $\beta = -17.2$ ,  $SE = 12.4$ ,  $z = -1.39$ ,  $p = .347$ .

*Error rates.* The main effect of Condition was significant,  $\chi^2 = 145.3$ ,  $p < .001$ . Similar to what was observed for the latencies, the TWSW condition produced significantly lower error rates than both the TW condition (a TW effect),  $\beta = 1.50$ ,  $SE = .13$ ,  $z = 11.69$ ,  $p < .001$ , and the SW condition,  $\beta = 1.39$ ,  $SE = .13$ ,  $z = 10.70$ ,  $p < .001$ . Paralleling the pattern in the latency data, although the TW condition produced numerically higher error rates than the SW condition (a 1.39% difference), this difference was not significant,  $\beta = .12$ ,  $SE = .09$ ,  $z = 1.25$ ,  $p = .421$ .

### Grammatical sentences

The mean RT and error rate were 1653 ms and 4.6%, respectively.

### *Discussion*

First, as expected, using the simultaneous presentation format, Experiment 1 replicated, in English, the TW effect reported by Mirault et al. (2018) in a grammatical decision task in French (see also Milledge et al., 2023): slower and less accurate responses to ungrammatical five-word sequences created by transposing the third and fourth word in a grammatical sentence (i.e., TW sequences) than to ungrammatical sequences created by transposing those two words as well as substituting the fifth word with a syntactically incompatible word (i.e., TWSW sequences). Second, in both latencies and error rates, responses to ungrammatical sequences created by substituting the fifth word only (i.e., SW sequences) stood between TWSW and TW sequences, although they were not statistically distinguishable from the latter.

In some sense, this positioning of the SW condition seems to be a bit puzzling. As noted previously, the grammaticality violation in TW sequences can occur as early as in position three while the grammaticality violation in SW sequences occurs no earlier than in position five. To the extent that participants notice the violation in TW sequences at the point that it arises, one would expect those sequences to be responded to somewhat faster than SW sequences. The fact that the TW condition was not faster than the SW condition, in fact the latencies on TW sequences were actually slightly longer than those on SW sequences, implies that the grammaticality violation in TW sequences was rarely noticed at the point that it arose. On the other hand, the large difference between the SW and TWSW condition seems to lead to the opposite conclusion. That is, both conditions involve a grammaticality violation arising at position five but the TWSW condition also involves a grammaticality violation occurring potentially as early as in position three. The large advantage of the TWSW condition over the SW condition suggests that the early grammaticality violation in the TWSW condition must have often been noticed at the point that it arose (e.g., at position three rather than at position five where the grammaticality violation in the SW condition would have been noticed).

In any case, overall, these results certainly support the claim that, in the simultaneous presentation format, accurate word order is not easily determined, or else TW sequences, sequences for which a misrepresentation of the word positions might result in grammatical sentences, would not be harder to reject than TWSW and (to some degree) SW sequences, sequences for which a misrepresentation of the word positions would not result in grammatical sentences. As discussed, this misrepresentation might occur in several ways, as exemplified by the various models. First, it might result from an interaction between noisy word position coding and syntactic expectations during parallel word processing in models such as the OB1-reader.<sup>4</sup> Second, in serial models, it might result from a rational-inference process also occurring during word processing at the stage in which the relevant words are held in a buffer before they are integrated with the sequence processed up to that point (Huang & Staub, 2021b).

Finally, the misrepresentation might occur after all words have been processed, either serially or in parallel, and the sequence is being evaluated for the grammatical decision.

Whatever model best represents the true nature of the TW effect, the effect appears to arise from there being incomplete clarity on word order in the simultaneous presentation format. The implication is that, if clarity is improved, the TW effect should be reduced as a result. As noted, in sentence reading, there appear to be two possible types of cues to word order: temporal cues and spatial cues. Temporal cues are not immediately clear in the simultaneous presentation format because all the words forming the sentence are presented at the same time in that format. However, the same would not be true in RSVP formats in which one word at a time is presented and, thus, clear temporal cues should be available. Spatial cues, on the other hand, are immediately clear in the simultaneous presentation format because the words forming the sentence are ordered on the line of text in a left-to-right manner. The fact that a sizeable TW effect typically emerges in the simultaneous presentation format indicates that spatial cues are not particularly effective at minimizing word-order noise, at least in a timed response task. However, spatial cues may still have a role in reducing that noise even when good temporal cues are available. If so, the TW effect in a spatial RSVP format that maintains those cues by presenting the words in the same positions as in the simultaneous presentation format may be smaller than in the (standard) central RSVP format that eliminates those cues by presenting all words centrally. These ideas were examined in Experiment 2 using both spatial RSVP and central RSVP formats.

## **Experiment 2 (central and spatial RSVP formats)**

### *Method*

#### Participants

One hundred and ten students (age 17–24 years) at the University of Western Ontario participated in this experiment, which was conducted on line, for course credit. All had normal or corrected-to-normal vision and were native speakers of English. None had participated in the previous experiment. As described below, in this experiment, compared to the previous experiment, the number of observations for each cell of the design was halved (16 from 32). However, with 110 participants (108 after the exclusions—see below), the total number of observations per condition, 1760 (1728 after the exclusions), still exceeded that recommended by Brysbaert and Stevens (2018).

### Materials

The materials were the same of Experiment 1, however, they were divided into two sets, each comprising 48 grammatical sentences and 16 TWSW, 16 SW, and 16 TW ungrammatical sentences. For each participant, one set was assigned to the central RSVP format, presented in one block, and the other set was assigned to the spatial RSVP format, presented in the other block. The assignment of sets to the two presentation formats and the order of the formats was counterbalanced across participants, as was the assignment of base sentences to TW, SW, and TWSW conditions (as in Experiment 1). Twelve lists were created for appropriate counterbalancing.

### Procedure

The experiment comprised two blocks of 96 trials each. In one block, the procedure, illustrated in the middle panel of Figure 1, was the same of Experiment 1 with the exception that, following the fixation symbol, the five words forming the sentence were presented individually for a duration of 250 ms each in the spatial locations that they occupied in the simultaneous presentation format. In the other block, the presentation format was the same except that the words forming the sentence were presented centred on the screen (see the right panel of Figure 1). The duration of the presentation of the entire

sequence was thus 1250 ms. Following the fifth word, the screen remained blank until a response was made. Participants could respond at any point during the presentation of the word sequence and response time was measured from the onset of the first word. If a response was made before the blank screen appeared, the feedback screen was displayed immediately following the word sequence.

Initially, participants were told that the five words forming the sentence would be presented one at a time. At the beginning of each block, participants were told about the presentation format used in the upcoming block and received 12 practice trials including 6 grammatical sentences and 6 ungrammatical sentences different from those used in the experiment proper. The instructions were otherwise identical to those of Experiment 1. The experiment was run using the jsPsych (de Leeuw, 2015) JavaScript library and participants completed it remotely in a full-screen browser window. Instead of using the right and left shift keys for “grammatical” and “ungrammatical” responses, participants were instructed to use the “A” and “L” keys, respectively. Further, we set no time limit for responses since this software allowed us to do so.

### *Results*

All participants except 2 had an accuracy above 70% ( $M = 91.72\%$ ,  $SD = 5.14$ ). Those two participants were excluded, leaving 108 participants equally distributed within each counterbalancing list. Prior to all analyses, the trials in which participants responded before the blank screen at the end of the word sequence (i.e., trials with RTs below 1250 ms, corresponding to .74% of all trials) were removed as invalid trials. For the latency analysis, correct response times below or above 2.5 standard deviations from the grand mean (corresponding to .60% of the correct responses) were removed as outliers. For the ungrammatical sentences, the remainder of the correct responses and the error rates were analysed in a similar way as in Experiment 1, with the exception that Format (spatial RSVP vs. central RSVP, within-subject and within-item) and Block Order (spatial RSVP first vs. central RSVP first, between-

subject and within-item) were included as additional fixed effects. The reason that Block Order was included was to account for potential practice effects emerging for the format presented in the second block. Mean response latencies and error rates for each condition for the ungrammatical sentences are reported in Table 2.

(Table 2 about here)

### Ungrammatical sentences

*RTs.* The main effect of Condition was significant,  $\chi^2 = 2084.1$ ,  $p < .001$ , however, the pattern of this effect was quite different from that of Experiment 1. As in Experiment 1, there was a significant TW effect overall, with faster responses in the TWSW condition than in the TW condition,  $\beta = -39.4$ ,  $SE = 4.73$ ,  $z = -8.32$ ,  $p < .001$ . However, different from Experiment 1, the SW condition was overall slower than both the TWSW condition,  $\beta = -189.3$ ,  $SE = 4.17$ ,  $z = -45.37$ ,  $p < .001$ , and the TW condition,  $\beta = 149.9$ ,  $SE = 5.33$ ,  $z = 28.14$ ,  $p < .001$ . The main effect of Block Order was also significant,  $\chi^2 = 8.54$ ,  $p = .004$ , reflecting the fact that participants who completed the central RSVP format first were overall faster than participants who completed the spatial RSVP format first. More relevantly, Block Order did interact with Format,  $\chi^2 = 60.01$ ,  $p < .001$ , reflecting the fact that responses in the central RSVP format were faster for participants who completed the spatial RSVP format first and, vice versa, responses in the spatial RSVP format were faster for participants who completed the central RSVP format first (i.e., there was a practice effect). Block Order also interacted with Condition,  $\chi^2 = 9.90$ ,  $p = .007$ , with follow-up analyses revealing that participants who completed the spatial RSVP format first produced, overall, a larger processing disadvantage for the SW condition compared to the TW condition (a 169-ms difference compared to a 138-ms difference for participants who completed the central RSVP format first),  $\beta = 28.3$ ,  $SE = 9.32$ ,  $z = 3.03$ ,  $p = .002$ . The overall pattern of results, however, was similar in the two block orders, with the SW condition being slowest, followed by the TW condition and then the TWSW

condition. No other effect was significant, all  $ps > .05$ , including the crucial interaction, that between Condition and Format, which was marginal,  $\chi^2 = 5.46, p = .065$ . This marginal interaction reflected the fact that the 205-ms difference between the SW condition and the TWSW condition in the spatial RSVP format tended to be larger than the 179-ms difference in the corresponding contrast in the central RSVP format. However, there was no hint of a presentation format difference for the key contrast between the TW condition and TWSW condition (i.e., the TW effect), and, in general, the pattern of results was quite similar in the two presentation formats, with, again, the SW condition being the slowest, followed by the TW condition and then the TWSW condition.

*Error rates.* The main effect of Condition was significant,  $\chi^2 = 183.03, p < .001$ . Similar to the latencies, overall, the TWSW condition produced significantly lower error rates than both the TW condition (a TW effect),  $\beta = .85, SE = .10, z = 8.88, p < .001$ , and the SW condition,  $\beta = 1.25, SE = .09, z = 13.52, p < .001$ . Further, the SW condition produced significantly higher error rates than the TW condition overall,  $\beta = -.40, SE = .08, z = -5.29, p < .001$ . The main effect of Format was also significant,  $\chi^2 = 24.35, p < .001$ , reflecting the fact that error rates were overall higher in the spatial RSVP format than in the central RSVP format. No other effect was significant, all  $ps > .150$ , including the crucial interaction, that between Condition and Format,  $\chi^2 = .08, p = .961$ .

### Grammatical sentences

The mean RT and error rate were 1761 ms and 5.4%, respectively, in the central RSVP format. The mean RT and error rate were 1762 ms and 5.8%, respectively, in the spatial RSVP format.

### *Combined analysis of Experiments 1 and 2*

In order to determine whether/how the results produced by the RSVP formats used in Experiment 2 differed from those produced by the simultaneous presentation format used in Experiment 1, the

latency and error rate data from the ungrammatical sentences in the two experiments were analysed in a combined analysis with Format (simultaneous vs. RSVP, between-subject and within-item) as an additional fixed effect.<sup>5</sup> The interest lay in how this factor might interact with the Condition factor. The interaction indeed emerged in both the latencies,  $\chi^2 = 389.46$ ,  $p < .001$ , and the error rates,  $\chi^2 = 25.19$ ,  $p < .001$ . (Main effects are challenging to interpret since the full sentence was available at the beginning of the trial in the simultaneous presentation format but not in the two RSVP formats in Experiment 2.)

For the latencies, follow-up analyses revealed that the RSVP formats significantly modified all the contrasts between the conditions examined. Specifically, the RSVP formats reduced the processing advantage for the TWSW condition compared to the TW condition (i.e., the TW effect, which diminished from 62 ms in the simultaneous presentation format to 39 ms in the RSVP formats),  $\beta = -26.3$ ,  $SE = 9.6$ ,  $z = -2.74$ ,  $p = .006$ . The reason is that, although the RSVP formats, compared to the simultaneous presentation format, sped up latencies for both of the conditions that involved a transposition, the speed-up was more pronounced for the TW condition,  $\beta = 90.4$ ,  $SE = 7.6$ ,  $z = 11.87$ ,  $p < .001$ , than for the TWSW condition,  $\beta = 64.1$ ,  $SE = 8.0$ ,  $z = 8.02$ ,  $p < .001$ .

While the TW effect was reduced in the RSVP formats, the processing advantage for the TWSW condition compared to the SW condition increased (from 41 ms in the simultaneous presentation format to 192 ms in the RSVP formats),  $\beta = 145.5$ ,  $SE = 8.4$ ,  $z = 17.31$ ,  $p < .001$ , and the contrast between the TW condition and the SW condition was also modified (changing from a 21-ms advantage for the SW condition in the simultaneous presentation format to a 153-ms *disadvantage* for that condition in the RSVP format),  $\beta = -171.8$ ,  $SE = 10.3$ ,  $z = -16.71$ ,  $p < .001$ . The reason for these changes is that while, as noted, latencies for the TW and TWSW conditions were shorter in the RSVP formats, latencies for the SW condition were longer in those formats,  $\beta = -81.4$ ,  $SE = 7.8$ ,  $z = -10.39$ ,  $p < .001$ .

For the error rates, follow-up analyses revealed that, similar to what was observed for the latencies, the RSVP formats reduced the processing advantage for the TWSW condition compared to the TW condition (i.e., the TW effect, which diminished from a 11.11% difference in the simultaneous presentation format to a 5.74% difference in the RSVP formats),  $\beta = .64$ ,  $SE = .16$ ,  $z = 4.02$ ,  $p < .001$ . Also similar to the latencies, the contrast between the TW condition and the SW condition was modified (changing from a 1.39% advantage for the SW condition in the simultaneous presentation format to a 4.13% disadvantage for that condition in the RSVP formats),  $\beta = .51$ ,  $SE = .12$ ,  $z = 4.31$ ,  $p < .001$ . In contrast, the processing advantage for the TWSW condition compared to the SW condition (a 9.72% difference in the simultaneous presentation format and a 9.87% difference in the RSVP formats) was not modified in the error rates,  $\beta = .13$ ,  $SE = .16$ ,  $z = .84$ ,  $p = .401$ . Overall, the reason for these changes is that the RSVP formats reduced error rates only for the TW condition,  $\beta = -.43$ ,  $SE = .13$ ,  $z = -3.27$ ,  $p = .001$ , whereas they had no impact on either the TWSW condition,  $\beta = .21$ ,  $SE = .17$ ,  $z = 1.23$ ,  $p = .219$ , or the SW condition,  $\beta = .08$ ,  $SE = .13$ ,  $z = .58$ ,  $p = .564$ .

### *Discussion*

Experiment 2 contrasted two RSVP formats, a central one in which (as is typically done) the words forming the sentence were presented centrally, and a spatial one in which the words were presented at the same rate as in the central format but in the spatial positions that they would have occupied in a simultaneous presentation. Both formats provided clear temporal cues to word order compared to the simultaneous presentation format used in Experiment 1. However, while the spatial RSVP format maintained the spatial relationships existing between the words in the sequence, providing clear spatial cues, in addition to clear temporal cues, to word order, the central RSVP format eliminated those relationships. As noted, both types of cues may be relevant in reducing word-order noise. If temporal cues are relevant, the expectation would be that RSVP formats, in general, would reduce the TW effect

compared to the simultaneous presentation format in which all the words forming the sequence are presented at the same time. Further, if spatial cues are relevant, the expectation would be that the TW effect would be more reduced in the spatial RSVP format which maintains those cues in comparison to the central RSVP format which eliminates them.

The results were only consistent with the first expectation, as in both RSVP formats, and in both the latencies and the error rates, the TW effect was reduced compared to that in the simultaneous presentation format used in Experiment 1, a similar pattern as that reported in previous studies that used an RSVP procedure (Huang & Staub, 2023; Liu et al., 2022; Milledge et al., 2023; Mirault et al., 2022b; for a similar pattern obtained using simultaneous but atypical presentation formats that were designed to encourage serial reading, see Mirault et al., 2023). The reason for this reduction is that although responses to both TW and TWSW sequences became faster and more accurate in the RSVP formats, the impact of the change in format was much larger in the TW condition. In contrast, SW sequences were unaffected by the format in the error rates, whereas, in the latencies, they even seemed to slow down in the RSVP formats. As a result, whereas performance for SW sequences was intermediate between TWSW and TW sequences in Experiment 1, responses to SW sequences were the slowest and least accurate in Experiment 2.

These results are consistent with the idea that the RSVP formats in Experiment 2 reduced word-order noise, noise that, in our experimental manipulations, is only relevant to sequences, specifically the TW and TWSW sequences, which contain a transposition, i.e., an alteration of the normal word order that makes word order uncertain when processing those sequences. Sequences that did not contain that sort of alteration, the SW sequences, would not be expected to be affected to any real degree by the presentation format. If anything, as noted above, RSVP formats may delay decisions to those sequences compared to the simultaneous presentation format because participants would always have to wait

until the fifth word is presented (i.e., 1000 ms after the initial word was presented) in order to realize that the sequence is ungrammatical. It seems possible that, in the simultaneous presentation format, processing of the fifth word may have started at an earlier point in time in some cases, allowing participants to begin processing the fifth word sooner. Note, further, that some of the SW sequences that we used, like “the white cat was slowly”, can be continued with a sixth word that would produce a grammatical sentence (e.g., “the white cat was slowly approaching”). What is also possible, then, is that the RSVP procedure might have suggested to participants to postpone their decision for those SW sequences (and, potentially, TWSW sequences if the transposition of the third and fourth word was not noticed) until they determined that no further words would appear that would make those sequences grammatical sentences. The finding that responses to SW sequences were approximately 100-ms slower in the RSVP formats than in the simultaneous presentation format while being slightly more accurate is consistent with these ideas.

What the present results are not consistent with is the idea that spatial cues, in addition to temporal cues, do much to help reduce word-order noise in a task like this. The reason is that the pattern of results was virtually identical in central and spatial RSVP formats (save for slight differences in the contrast between the TWSW condition and the SW condition in the latencies and an overall higher error rate in the latter format): In both formats, the TW effect (39 ms in both formats) was reduced compared with that in Experiment 1 (62 ms) and TWSW and TW sequences received faster and more accurate responses than SW sequences. Therefore, in the context of this experiment, having clear spatial cues to word order does not seem to do much to diminish noise in word position coding (and, hence, to diminish the potential for transposed-word confusability effects).

Note that this conclusion contrasts with Milledge et al.’s (2023), who also used central and spatial RSVP formats in an English grammatical decision task. As mentioned in the Introduction, they reported, in the latencies, a significantly larger TW effect in the spatial RSVP format (22 ms) than in the central RSVP

format (6 ms). This result is somewhat surprising based on our reasoning that, because a spatial RSVP format includes both temporal and spatial cues to word order, it should produce, if anything, a *smaller*, not a *larger* TW effect than a central RSVP format, a format which only includes temporal cues to word order. Indeed, a few considerations do invite caution in interpreting their result.

First, as noted, Milledge et al. used central and spatial RSVP formats in separate experiments (i.e., a between-subject manipulation) and one would expect that, if their pattern were real, it would replicate easily in a within-subject manipulation such as that used in the present experiment. Second, in Milledge et al.'s contrast, there was a numerical tendency in the opposite direction in the error rates, with the spatial RSVP experiment producing a 6.68% TW effect and the central RSVP experiment producing a 9.78% TW effect, although that difference was not significant. Finally, despite the fact that there were no timing differences in the procedure of the two experiments (i.e., the only difference was whether the words were presented centrally or in the position they occupied in the simultaneous presentation format), participants in the spatial RSVP experiment were over 100-ms faster than participants in the central RSVP experiment (a statistically significant difference). Although this overall latency speed-up in the spatial RSVP experiment was not accompanied by an overall error-rate increase in that experiment, it does not seem unreasonable that a speed-accuracy trade-off might have played some role in producing the different patterns observed with latencies and error rates in the contrast between the two experiments.

The present results, in any case, suggest that temporal codes, rather than spatial codes, would appear to be more relevant in allowing readers to deal with the problems associated with correctly perceiving word order. Overall, the main impact that both RSVP formats seem to have on the processing of ungrammatical sentences appears to be reducing uncertainty as to the point in time at which each word occurs in the sequence.

## General Discussion

The Transposed-Word (TW) effect in the grammatical decision task (Mirault et al., 2018), the finding that sequences such as “the white was cat big” are more likely to be classified as grammatical sentences than sequences of the sort “the white was cat slowly” (TWSW sequences), indicates that word-order coding must be noisy, whatever model of sentence reading is assumed. In serial models such as Huang and Staub’s (2021b) model, word-order noise for the words that are temporarily held in a buffer before being integrated with the sequence processed up to that point is what would allow rational-inference processes to “correct” unusual word orders (e.g., “was cat”) to more likely orders (e.g., “cat was”). Similarly, in parallel models such as the OB1-reader (Snell et al., 2018), word-order noise occurring as a result of processing multiple words in parallel is what would allow syntactic expectations to misrepresent some words (e.g., “was” and “cat”) as having occurred in the syntactically appropriate positions instead of their actual positions. Finally, word-order noise occurring at the decision stage, regardless of whether the words forming the sentence were initially processed serially or (to some extent) in parallel, is what would impair “ungrammatical” responses to TW sequences.

Word-order noise, however, can be reduced with spatial and temporal cues to word order, that is, cues that make the position and the point in time at which a word occurs in the sequence clearer. In normal reading, readers presumably have access to both type of cues (e.g., they can determine that “the” appeared to the left of “cat” and that processing of “the” was started before processing of “cat”). However, the presence of a TW effect in the present experimental paradigm suggests that those cues are not always sufficient to prevent TW sequences from being confused with their base grammatical sentences.

The present experiments allowed an examination of whether providing clearer cues would help reduce that confusion. Three types of presentation formats for sentences requiring a grammatical decision

were used. Experiment 1 involved the “simultaneous presentation” format in which all the words forming the sentence are presented at the same time, the standard format used in grammatical decision tasks (Huang & Staub, 2021a; Liu et al., 2020, 2021; Mirault et al., 2018, 2020, 2022a, 2022b; Snell & Grainger, 2019b; Wen et al., 2021a, 2021b). This format provides clear spatial cues because the words are presented next to one another on the screen, with the word occupying the leftmost position being the first word in the sentence, the adjacent word to the right being the second word, etc. In contrast, this format may not provide clear temporal cues because all the words forming the sentence are presented at the same time, making it difficult to determine which word was processed first unless specific actions are taken to create effective temporal cues. Experiment 2 involved two Rapid Serial Visual Presentation (RSVP) formats. In the “spatial” RSVP format, the words were presented in the same positions as in the simultaneous presentation format but one after another for 250 ms each. In the “central” RSVP format, the words were presented at the same rate as in the spatial RSVP format but centrally. Both RSVP formats should provide clearer temporal cues to word order than the simultaneous presentation format because the words forming the sentence are presented one after another rather than simultaneously. However, whereas the spatial RSVP format, retaining the positions occupied by the words in the simultaneous presentation format, should also provide clear spatial cues, the same is not true for the central RSVP format in which all words come with the same spatial code.

If temporal and spatial cues to word order help reduce word-order noise, the expectation would be that formats containing either or both of those cues would make accurate processing of sentences involving a word transposition easier, especially if the sentence is a TW sequence which could be misinterpreted as its base grammatical sentence. This expectation, however, would not hold for sentences that do not involve a word transposition, such as SW sequences created by substituting a word in a grammatical sentence with a syntactically incompatible word (e.g., “the white cat was slowly”).

With respect to temporal cues, the present experiments produced evidence consistent with these ideas. In Experiment 1, involving the simultaneous presentation format, responses were fastest and most accurate for TWSW sequences, followed by SW sequences, followed by TW sequences (although the contrast between TW and SW sequences was not statistically significant). In contrast, in Experiment 2, involving the RSVP formats, whereas responses were again fastest and most accurate for TWSW sequences, they were now followed by TW sequences and, finally, by SW sequences, with the contrast between the former two sequences (i.e., the TW effect) producing a smaller (but still significant) difference than that in Experiment 1.

A combined analysis of the two experiments revealed that the reason for the different patterns was that whereas responses to sentences involving transpositions (i.e., TW and TWSW sequences) were faster and, in the case of TW sequences, more accurate with RSVP formats than with the simultaneous presentation format, responses to SW sequences were actually slower with RSVP formats than with the simultaneous presentation format. As noted above, this result with the SW sequences may be due to the fact that, on some occasions, participants might realize more quickly that the SW sequence is not a complete grammatical sentence when the words comprising the sequence are presented simultaneously than when they are presented serially, with the final word, the word that makes the sequence ungrammatical, only appearing after 1000 ms have elapsed. The more important point, in any case, is that having clear temporal cues to word order only helps processing sentences involving word transpositions, especially TW sequences, resulting in reduced TW effects.

A similar conclusion would not seem to apply to spatial cues, however. The reason is that the contrast between the TW and TWSW conditions in Experiment 2 was virtually identical in the central and spatial RSVP formats despite the fact that the former format provided clear spatial cues whereas the latter did not. Therefore, although in normal reading the order in which words must be read is determined by their position in the sentence, spatial cues do not appear to be a type of cue that readers can use

effectively when dealing with problems associated with word order, at least when temporal cues are also available and rapid responding is required. In the following, we discuss the implications of these results for the most relevant models that have been used to explain the TW effect, i.e., Huang and Staub's (2021b) model and Snell et al.'s (2018) OB1-reader.

In Huang and Staub's (2021b) model, as noted, the TW effect is the result of a rational-inference process that is applied to words temporarily stored in a buffer to change their order in a way that is consistent with one's expectations about the intended message. Importantly, this process is assumed to be influenced by the type of errors that might occur in the communication process, including transposition errors, in generating a sentence or in perceiving it. The more likely that an error could have arisen, the higher the probability that the perceived sequence was corrupted by that error and, as a result, differs from the intended sequence. One possible way for this model to account for the present results, therefore, is simply to assume that the occurrence of transposition errors is estimated to be less likely in the presence of clear temporal cues, while being generally unaffected by clear spatial cues. That is, when a presentation format is used in which the point in time at which words have occurred in the sequence is clear (e.g., an RSVP format), the probability of two words being transposed in the perceived sequence will be estimated to be lower than when a simultaneous presentation format is used, a format which does not provide temporal cues that are as clear unless the reader is purposely reading slowly. Therefore, rational inference will be applied more rarely in that situation than in other situations, making it less likely that sentences involving transpositions, particularly TW sequences, are classified as grammatical. (Note, however, that rational inference would, presumably, be applied on a reasonable portion of trials, or else there would be no TW effect in RSVP formats.) In contrast, the presence of spatial cues would appear to have little effect on the estimated probability of transposition errors. Whether spatial cues are present (e.g., in the simultaneous presentation format and in the spatial RSVP

format) or not (e.g., in the central RSVP format), the probability of a transposition error having occurred would be estimated to be the same.

How the OB1-reader (Snell et al., 2018) would account for the RSVP format results is somewhat less clear. As noted, that model assumes that, in simultaneous presentation formats, processing of the words forming a sentence would largely overlap in time. The model also assumes that, as a result, each word is mapped onto a spatiotopic representation coding the word's position. That is, order information is represented in the model in a spatially coded mental representation. The model further assumes that the words' spatial representations in the full representation are noisy, which would allow syntactic expectations to influence them and TW effects to be produced.

What would be the impact of enforcing serial reading through RSVP formats according to this model? As just noted, in the normal reading of a sentence like "the white was cat big" (i.e., in simultaneous presentation formats), a spatiotopic representation is assumed to be constructed. That representation initially includes five "blobs" roughly matching the length of the words in the sentence (i.e., a 3-letter blob, followed by a 5-letter blob, followed by three more 3-letter blobs). Then, as each word is recognized, it is mapped to one of the blobs based on whether the word matches the blob's length and syntactic expectations (Snell et al., 2018). As suggested by an anonymous reviewer of the present paper, however, the spatiotopic representation would not be used in RSVP presentations because spatiotopic information is assumed (in the OB1-reader) to reside in visual short-term memory and the time constraints on that memory system would not allow the development of a representation of a serially presented word sequence. Therefore, the model, as currently formulated, does not actually have a mechanism for explaining reading behaviour in RSVP formats.

It is certainly the case, of course, that in RSVP formats (as well as in simultaneous but atypical presentation formats that were designed to encourage serial reading, such as unspaced text or right-to-

left reading in French: Mirault et al., 2023), some sort of memory representation must be constructed in order to carry out the task. If the creators of the OB1-reader do wish to extend the model to RSVP reading, they will, therefore, need to start by specifying what type of representation is used in RSVP reading. Considering the fact that our data indicate that spatial cues are not particularly effective at allowing readers to maintain word order accuracy, one possibility would be to suggest that the representation is temporally based. For example, each word would be given a time tag during its processing. From that point, the model's processes could generally unfold in the same way as they are presumed to unfold when reading simultaneously presented words. The main difference, of course, would be that the accuracy of the time tags would be assumed to be higher than the accuracy of the position codes created when reading simultaneously presented words. This set of assumptions would seem to be perfectly consistent with the OB1-reader model, allowing that model to explain the present results.

Note that, because the representation of the word sequence would be assumed to be more abstract in RSVP presentations, the factor of word length (i.e., as noted, the model assumes that words that are the same length are more likely to be incorrectly transposed) would not necessarily be assumed to play an important role. (As noted in footnote 4, Experiment 1 produced no evidence that word length similarity affected performance in the simultaneous presentation format in any case). However, syntax would presumably be able to play a role when evaluating the grammaticality of the sequence's representation (Mirault et al., 2022b). That is, based on knowledge of appropriate syntax and knowledge that word order coding is noisy, the expectation would be that when this representation was being processed, delays would be sometimes created when the temporal order was coded correctly but syntactic rules suggested that the coding might be incorrect, as in the TW condition.

Note further that this discussion does not concern the serial vs. parallel nature of Huang and Staub's (2021b) vs. Snell et al.'s (2018) models but, rather, the mechanisms those models assume (or in the case

of Snell et al.'s model for RSVP formats, *might* assume) to deal with word-order noise. Further, the processing stage at which the TW effect arises needs not be the lexical processing stage as assumed by those models. It may very well be during a decision stage, i.e., the stage at which a grammaticality evaluation is made of the code arisen from either serial or parallel processing. That is, according to this type of model, the TW effect may be explained by both serial and parallel models as the TW effect would mainly be a function of the decision/evaluation stage based on the assumption that, at that point in processing, word order would still be noisy. Thus, a TW sequence, a sequence which could be made grammatical with a transposition, would bias a "grammatical" decision more so than would a TWSW sequence, resulting in a TW effect. Further, this bias would vary as a function of the level of noise associated with word order, with temporal cues reducing noise (and with it, the TW effect) and spatial cues leaving the level of noise (and with it, the TW effect) essentially unaltered.

This type of model has actually received almost no attention in studies of the TW effect. One potential exception is represented by Huang and Staub (2023), who recently considered the hypothesis that the TW effect in RSVP formats results from a process of "late-stage redintegration". With this term, Huang and Staub refer to the possibility that, when dealing with TW sequences in RSVP formats, "participants might have, at the time of making the response, reconstructed their short-term memory into a grammatical sequence... even if they initially did detect the anomaly during incremental processing of the sentence" (p. 2).

In order to address this possibility, Huang and Staub examined two RSVP conditions in addition to a condition in which the words comprising the sequence were presented simultaneously. In one RSVP condition, called the "end-of-sentence" condition, participants were instructed to make the grammatical decision after the RSVP sequence was over. In the other RSVP condition, called the "self-terminating" condition, participants were instructed to make the grammatical decision at any time during the RSVP sequence. Huang and Staub assumed that while the processes involved in the former condition might be

based on a potentially reconstructed version of the sequence in short-term memory, the same would not be true for the latter condition, which would only involve processes occurring during processing of the RSVP sequence. A TW effect (defined as the error-rate difference between TW sequences and scrambled sequences) emerged in both conditions, suggesting that the TW effect is not the result of a late redintegration process.

Although this pattern of results would appear to challenge a decision-stage explanation for the TW effect, what it merely challenges is the idea that a TW effect would emerge at a decision stage starting only after the whole sequence has been processed. Admittedly, it is reasonable that, in some situations, participants would want to wait to process the whole sequence before making a decision. For example, in our Experiment 2, participants were given no specific instructions concerning whether they were allowed to respond during the RSVP sequence, although they were instructed to respond as quickly and as accurately as possible and could do so during the RSVP sequence if they so chose. Still, responses occurring during the RSVP sequence (i.e., before 1250 ms had elapsed) accounted for only .74% of all trials (1.08% considering only ungrammatical sequences), suggesting that, in the vast majority of cases, participants preferred to process the whole sequence before making a decision. Part of the reason for this preference might be due to the inclusion in our experiments of a type of ungrammatical sentence, the SW sequences, that only became ungrammatical on the fifth word and/or when it was realized that the sequence would not be continued with a sixth word that would make it grammatical.

Such, however, might not have been the case in Huang and Staub's (2023) "self-terminating" condition. In that condition, the sentences were composed of seven words (rather than five, as in the present experiments) and the transposition (in TW sequences) involved the third and fourth words, words that, in a seven-word sentence, occur much earlier than the end of the sentence (in our experiments, the third and fourth words were also transposed, but there was only one more word, the fifth word, before the end of the sentence). Further, there were no sentences in Huang and Staub's experiments that only

became ungrammatical toward the end. It is reasonable that in that situation, participants would have more than enough evidence to start the decision process before the whole sequence is processed, especially in a timed response situation. Thus, although Huang and Staub's finding of a TW effect in their "self-terminating" RSVP condition does suggest that there was not some quite late reconstructive process that contributed to the effect, the finding itself is not incompatible with the idea that the effect occurs at a decision stage, albeit one that can be started before the whole sequence has been processed.

Note, finally, that a decision-stage account would not deny the possibility that processing relevant to the sentence's grammaticality would occur before the grammaticality decision, or when there is no grammatical decision to be made at all. For example, Wen et al. (2019) used brief (200 ms) simultaneous presentations of four-word sequences with one of the word locations being subsequently cued. Replicating Snell and Grainger (2017), they found that participants were more accurate at reporting the word that had appeared in the cued location when the four-word sequence they were presented with formed a grammatical sentence (e.g., "the man can run") rather than an ungrammatical sentence obtained by scrambling the word order (e.g., "run the can man").

More important for the present discussion, they recorded Event-Related Potentials (ERPs) and found a reduced N400 component for grammatical than ungrammatical sentences, with the effect emerging as early as 274 ms post-sentence onset. According to Wen et al. (2019), this effect would reflect the interaction between rapidly activated sentence-level structures and word identities, with feedback from the former guiding the identification of the latter via syntactic and semantic constraints. Wen et al. (2021a) obtained a similar pattern of ERP results using the grammatical decision task, most notably for the comparison, in their Experiment 2, of TW and TWSW sequences, with the former producing, like grammatical sentences in Wen et al. (2019), a reduced N400. Overall, these results certainly seem to suggest that some syntactic/semantic processing relevant to the sentence's grammaticality can occur

quite rapidly and potentially before the stage at which, in grammatical decision tasks, the grammaticality of the sentence is evaluated. Therefore, it seems reasonable that some early processing contributes to the TW effect rather than decision-stage processing being responsible for all of it.

### *Conclusion*

Overall, the present experiments corroborate the idea that the position of words within a sentence is less easily confused when those words are presented one at a time (Experiment 2) than when they are presented simultaneously (Experiment 1), that is, when the presentation format provides clear temporal cues to word order. The position of words within a sentence, however, was not confused less easily when the presentation format provided additional clear spatial cues to word order (the spatial RSVP format used in Experiment 2, which produced results similar to those in the central RSVP format used in that experiment). Therefore, temporal codes would appear to be more relevant than spatial codes in allowing readers to deal with problems associated with word transpositions. This pattern of results can inform the development of models of sentence reading that assume that such problems can arise during serial word processing, during parallel word processing, or at the stage at which the sentence is evaluated for the relevant decision.

## Footnotes

1. We set such a long time limit in order to approximate Mirault et al.'s (2018) procedure, in which there was no time limit. The software we used, DMDX (Forster & Forster, 2003), to our knowledge, does require the specification of a time limit.
2. All analyses reported in this paper were repeated using a linear mixed-effects model with  $1000/RT$  as the dependent variable. The pattern of results was essentially unchanged.
3. All analyses in this paper were repeated specifying appropriate random slopes for Subjects and Items. None of the models with the maximal random structure (Barr et al., 2013) converged.
4. Following the suggestion of a reviewer of a previous version of this manuscript (Joshua Snell), we conducted post-hoc analyses in order to test one of the assumptions of the OB1-reader (Snell et al., 2018). As noted, word position coding in that model is influenced by bottom-up factors, among them, word length, with the positions of words with similar length (e.g., two 3-letter words) assumed to be more confusable with each other than the positions of words that differ noticeably in length (e.g., a three-letter word and a ten-letter word). The implication is that when the two transposed words have the same length TW and TWSW sequences should prove harder. The reason is that TWSW sequences would be easily confused with SW sequences, sequences that were harder than TWSW sequences overall in Experiment 1, and TW sequences would be easily confused with their grammatical base sentences. Further, it would seem that the latter would be likely the most impacted sequences because, being confused with grammatical sentences, a bias for the incorrect response would be produced. As a result, the TW effect would also increase. We examined this idea in two sets of analyses of the Experiment 1 data, one in which the length difference between the two transposed words was coded as a dichotomous variable (i.e., same length vs. different length) and one in which it was coded as a continuous variable (i.e., the absolute value of the difference). In neither analysis did the length

difference produce a main effect, all  $ps \geq .141$ , or an interaction with the TW effect, all  $ps \geq .119$ .

In sum, this assumption of the OB1-reader's gained no support from these analyses, although it must be noted that our experiment was not designed to examine that assumption (e.g., only 10 of our 96 ungrammatical sentences included same-length transposed words).

5. For this combined analysis, we used the data from both of the RSVP formats used in Experiment 2 because the analysis for that experiment did not reveal important differences between the two formats. When conducting the combined analysis contrasting the simultaneous presentation format of Experiment 1 only with the spatial RSVP format of Experiment 2 (the format that more closely matches the simultaneous presentation format), the pattern of results was virtually identical.

**Author note**

Correspondence concerning this article may be addressed to: Giacomo Spinelli, Department of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy (e-mail: giacomo.spinelli@unimib.it) or Stephen J. Lupker, Department of Psychology, University of Western Ontario, London, Ontario, N6A 5C2, Canada (e-mail: lupker@uwo.ca).

**Funding details**

This research was partially supported by Natural Sciences and Engineering Research Council of Canada Grant A6333 to Stephen J. Lupker.

**Disclosure statement**

The authors have no relevant financial or non-financial interests to disclose.

**Data availability statement**

The raw data and R files used for the analyses are publicly available at <https://osf.io/q7yd3/>. A spreadsheet with the word sequences used in the present experiments is also available at that link. The scripts for running the experiments are available upon request.

## References

- Angele, B., Tran, R., & Rayner, K. (2013). Parafoveal–foveal overlap can facilitate ongoing word identification during reading: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2), 526–538. <https://doi.org/10.1037/a0029492>
- Balota, D. A., Aschenbrenner, A. J., & Yap, M. J. (2013). Additive effects of word frequency and stimulus quality: the influence of trial history and data transformations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(5), 1563–1571. <https://doi.org/10.1037/a0032186>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bolker, B. (2022). GLMM FAQ. Retrieved from <https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html>
- Brothers, T., Hoversten, L. J., & Traxler, M. J. (2017). Looking back on reading ahead: No evidence for lexical parafoveal-on-foveal effects. *Journal of Memory and Language*, 96, 9–22. <https://doi.org/10.1016/j.jml.2017.04.001>
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41(4), 977–990. <https://doi.org/10.3758/BRM.41.4.977>

Brysbaert, M., & Stevens, M. (2018). Power Analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1(1), 9. <http://doi.org/10.5334/joc.10>

Cohen-Shikora, E. R., Suh, J., & Bugg, J. M. (2019). Assessing the temporal learning account of the list-wide proportion congruence effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(9), 1703–1723. <https://doi.org/10.1037/xlm0000670>

Colombo, L., Spinelli, G., & Lupker, S. J. (2020). The impact of consonant–vowel transpositions on masked priming effects in Italian and English. *Quarterly Journal of Experimental Psychology*, 73(2), 183–198. <https://doi.org/10.1177/1747021819867638>

Dare, N., & Shillcock, R. (2013). Serial and parallel processing in reading: Investigating the effects of parafoveal orthographic information on nonisolated word recognition. *Quarterly Journal of Experimental Psychology*, 66(3), 487–504. <https://doi.org/10.1080/17470218.2012.703212>

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12. <https://doi.org/10.3758/s13428-014-0458-y>

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A Dynamical Model of Saccade Generation During Reading. *Psychological Review*, 112(4), 777–813. <https://doi.org/10.1037/0033-295X.112.4.777>

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 35(1), 116–124. <https://doi.org/10.3758/BF03195503>

Fox, J. & Weisberg, S. (2016). car: Companion to applied regression. R Package Version 2.1-2. Retrieved from <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>

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*, *110*(20), 8051–8056. <https://doi.org/10.1073/pnas.1216438110>

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*, 393-400. <https://doi.org/10.3758/s13423-022-02150-9>

Lenth, R. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means. <https://CRAN.R-project.org/package=emmeans>

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., & Wang, J. (2021). Context but not reading speed modulates transposed-word effects in Chinese reading. *Acta Psychologica*, *215*, 103272. <https://doi.org/10.1016/j.actpsy.2021.103272>

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>

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, 6, 1171. <https://doi.org/10.3389/fpsyg.2015.01171>

Lupker, S. J., Spinelli, G., & Davis, C. J. (2020a). Is zjudge a better prime for JUDGE than zudge is?: A new evaluation of current orthographic coding models. *Journal of Experimental Psychology: Human Perception and Performance*, 46(11), 1252–1266. <https://doi.org/10.1037/xhp0000856>

Lupker, S. J., Spinelli, G., & Davis, C. J. (2020). Masked form priming as a function of letter position: An evaluation of current orthographic coding models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(12), 2349–2366. <https://doi.org/10.1037/xlm0000799>

Milledge, S. V., Bhatia, N., Mensah-Mcleod, L., Raghvani, P., A. McGowan, V., Elsherif, M. M., ... & Paterson, K. B. (2023). The transposed-word effect provides no unequivocal evidence for parallel processing. *Attention, Perception, & Psychophysics*. <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., Guerre-Genton, A., Dufau, S., & Grainger, J. (2020). Using virtual reality to study reading: An eye-tracking investigation of transposed-word effects. *Methods in Psychology*, 3, 100029.

<https://doi.org/10.1016/j.metip.2020.100029>

Mirault, J., Leflaëc, C., & Grainger, J. (2022a). Orthographic relatedness and transposed-word effects in the grammatical decision task. *Attention, Perception, & Psychophysics*, *84*(3), 1043–1051.

<https://doi.org/10.3758/s13414-021-02421-y>

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

Mirault, J., Vandendaele, A., Pegado, F., & Grainger, J. (2023). The impact of atypical text presentation on transposed-word effects. *Attention, Perception, & Psychophysics*, *85*(8), 2859–2868.

<https://doi.org/10.3758/s13414-023-02760-y>

Pegado, F., & Grainger, J. (2019). Dissociating lexical and sublexical contributions to transposed-word effects. *Acta Psychologica*, *201*, 102943. <https://doi.org/10.1016/j.actpsy.2019.102943>

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. & Lupker, S. J. (2003). Transposed-letter confusability effects in masked form priming. In S. Kinoshita & S. J. Lupker (Eds.), *Masked Priming: The State of the Art* (pp. 97–120). Hove, UK: Psychology Press. <https://doi.org/10.4324/9780203502846>

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>

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: Accounting for initial fixation locations and refixations within the EZ Reader model. *Vision Research*, *39*(26), 4403–4411.

[https://doi.org/10.1016/S0042-6989\(99\)00152-2](https://doi.org/10.1016/S0042-6989(99)00152-2)

Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E–Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research*, *7*(1), 4–22.

<https://doi.org/10.1016/j.cogsys.2005.07.002>

Reichle, E. D., Liversedge, S. P., Pollatsek, A., & Rayner, K. (2009a). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Sciences*, *13*(3), 115–119.

<https://doi.org/10.1016/j.tics.2008.12.002>

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using EZ Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, *16*(1), 1–21.

<https://doi.org/10.3758/PBR.16.1.1>

Reilly, R. G., & Radach, R. (2006). Some empirical tests of an interactive activation model of eye movement control in reading. *Cognitive Systems Research*, *7*(1), 34–55.

<https://doi.org/10.1016/j.cogsys.2005.07.006>

Schotter, E. R., & Payne, B. R. (2019). Eye movements and comprehension are important to reading.

*Trends in Cognitive Sciences*, 23(10), 811–812. <https://doi.org/10.1016/j.tics.2019.06.005>

Snell, J., & Grainger, J. (2019a). Readers are parallel processors. *Trends in Cognitive Sciences*, 23, 537-

546. <https://doi.org/10.1016/j.tics.2019.04.006>

Snell, J., & Grainger, J. (2019b). 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., 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>

van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67(6), 1176–

1190. <https://doi.org/10.1080/17470218.2013.850521>

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>

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>

White, A. L., Palmer, J., & Boynton, G. M. (2020). Visual word recognition: Evidence for a serial bottleneck in lexical access. *Attention, Perception, & Psychophysics*, 82(4), 2000–2017. <https://doi.org/10.3758/s13414-019-01916-z>

Yang, H., Chen, J., Spinelli, G., & Lupker, S. J. (2019). The impact of text orientation on form priming effects in four-character Chinese words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(8), 1511–1526. <https://doi.org/10.1037/xlm0000655>

Table 1

*Mean RTs and Percentage Error Rates (and Corresponding 95% Confidence Intervals) for the Ungrammatical Sentences in Experiment 1 (Simultaneous Presentation Format)*

Condition	Example	RTs	Error rates
TW	The white was cat big	1841 [1827, 1855]	15.28 [14.37, 16.18]
TWSW	The white was cat slowly	1779 [1764, 1793]	4.17 [3.67, 4.67]
SW	The white cat was slowly	1820 [1806, 1833]	13.89 [13.02, 14.76]
TW – TWSW		62	11.11
SW – TWSW		41	9.72
TW – SW		21	1.39

*Note.* TW = Transposed-Word condition. TWSW = Transposed-Word, Substituted-Word condition. SW =

Substituted-Word condition.

Table 2

*Mean RTs and Percentage Error Rates (and Corresponding 95% Confidence Intervals) for the Ungrammatical Sentences in Experiment 2 (Spatial and Central RSVP Formats)*

Condition	Example	Spatial RSVP		Central RSVP	
		RTs	Error rates	RTs	Error rates
TW	The white was cat big	1770 [1763, 1777]	12.69 [12.05, 13.34]	1779 [1771, 1786]	9.54 [8.97, 10.11]
TWSW	The white was cat slowly	1731 [1725, 1737]	6.25 [5.78, 6.71]	1740 [1733, 1746]	4.50 [4.10, 4.90]
SW	The white cat was slowly	1936 [1930, 1942]	17.52 [16.78, 18.26]	1919 [1914, 1925]	12.97 [12.31, 13.62]
TW – TWSW		39	6.44	39	5.04
SW – TWSW		205	11.27	179	8.47
TW – SW		-166	-4.83	-140	-3.43

*Note.* TW = Transposed-Word condition. TWSW = Transposed-Word, Substituted-Word condition. SW =

Substituted-Word condition.

## Figure 1

An Illustration of the Presentation Formats Used in the Present Experiments. The simultaneous presentation format was used in Experiment 1 and the spatial and central RSVP formats were used in Experiment 2. In Experiment 1, there was a 20-s limit for the response. In Experiment 2, each word was presented for 250 ms, there was no time limit for the response, and participants could respond at any point during the sequence presentation. If a response was made before the blank screen appeared, the feedback screen was displayed immediately following the word sequence. On the feedback screen, a green/red "o" in bold font was presented for correct and incorrect responses, respectively (please see the online version of this article in order to view colours).