Cognition 216 (2021) 104846

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Using eye tracking to investigate failure to notice word transpositions in reading

Kuan-Jung Huang *, Adrian Staub

Department of Psychological and Brain Sciences, University of Massachusetts, 426 Tobin Hall, Amherst, MA 01003, USA

ARTICLE INFO

Keywords:
Eye movements
Sentence comprehension
Word position coding
Incremental processing

ABSTRACT

Previous research (Mirault, Snell, & Grainger, 2018) has demonstrated that subjects sometimes incorrectly judge an ungrammatical sentence as grammatical when it is created by the transposition of two words in a grammatical sentence (e.g., *The white was cat big*). Here we present two eye-tracking experiments designed to assess the prevalence of this phenomenon in a more natural reading task, and to explore theoretical explanations. Readers failed to notice transpositions at about the same rate as in Mirault et al. (2018). Failure to notice the transposition was more common when both words were short, and when readers' eyes skipped, rather than directly fixated, one of the two words. The status of the transposed words as open- or closed-class did not have a reliable effect. The transposed words caused disruption in the eye movement record only on trials when participants ultimately judged the sentence to be ungrammatical, not when they judged the sentence to be grammatical. We argue that the results are not entirely consistent with the account offered by Mirault et al. (2018), which attributes failure to notice transpositions to parallel processing of adjacent words, or with a late, post-perceptual rational inference account (Gibson, Bergen, & Piantadosi, 2013). We propose that word recognition is serial, but post-lexical integration of each word into its context may not be perfectly incremental.

1. Introduction

The visual word recognition literature has demonstrated a range of 'transposed-letter effects' involving facilitatory priming provided by a non-word derived from a target word by transposing letters (e.g., Perea & Lupker, 2003), or inhibition in recognizing a target word when it is confusable with a transposed-letter neighbor (e.g., Andrews, 1996). These effects have been taken as evidence that letters are processed in parallel and that letter position coding is noisy (e.g., Davis, 2010). Recently, Mirault et al. (2018) demonstrated that when subjects are asked to rapidly judge the grammaticality of a sentence, they are relatively likely to incorrectly judge an ungrammatical sentence as grammatical when it is created by the transposition of two words in a grammatical sentence (e.g., *The white **was cat** big*); for related work, see also Pegado & Grainger, 2019a,b, Snell and Grainger (2019b), Wen, Mirault, and Grainger (2021), and Liu, Li, Paterson, and Wang (2020). Analogous to the argumentation from transposed-letter effects, they argued on the basis of this result that words are recognized in parallel, and that word position coding is noisy.

The claim that written words are processed in parallel has long been highly controversial (see Reichle, Liversedge, Pollatsek, & Rayner, 2009, and Snell & Grainger, 2019a, for extended discussion). One critical issue is whether, in eye movement experiments, properties of a parafoveal word – the word to the right of current fixation – can influence fixation times on the foveal word. With the exception of effects of low-level properties such as bigram frequency, such effects have been almost non-existent in controlled experiments (e.g., Brothers, Hoversten, & Traxler, 2017; Snell, Declerck, & Grainger, 2018; Veldre & Andrews, 2018). This lack of parafoveal-on-foveal effect has been regarded as evidence against parallel lexical processing. In addition, recent behavioral and imaging experiments have suggested that readers *cannot* simultaneously recognize two words (White, Palmer, & Boynton, 2018, 2020; White, Palmer, Boynton, & Yeatman, 2019). In behavioral experiments, subjects were presented with two briefly-presented words, preceded and followed by masks. The critical finding was that when subjects had to make lexical or semantic decisions for both words, their performance fit the predictions of a serial processing model in which only one word could be processed at a time, but was much lower than predicted by a limited-capacity parallel model. Further neuroimaging data identified the bottleneck for serial processing at the left anterior

* Corresponding author.
E-mail addresses: kuanjunghuan@umass.edu (K.-J. Huang), astaub@psych.umass.edu (A. Staub).

https://doi.org/10.1016/j.cognition.2021.104846
Received 23 April 2020; Received in revised form 10 May 2021; Accepted 11 July 2021
Available online 17 July 2021

visual word form area, whose BOLD signal was influenced only by the frequency of one attended word. In light of this controversy, the existence and interpretation of the transposed-word effect reported by Mirault et al. is of substantial interest.

Separate from this debate over serial versus parallel word processing, other research has investigated other misreading phenomena (e.g., Gibson et al., 2013; Staub, Dodge, & Cohen, 2019). One account of failure to notice various errors in sentences is the 'noisy-channel' framework (Gibson, Bergen, & Piantadosi, 2013; Levy, 2008; Ryskin, Futrell, Kiran, & Gibson, 2018). The central hypothesis is that readers, with their extensive reading and language experience, have knowledge of what types of errors might occur, and how likely they are, during the communication process. To optimize reading comprehension, readers incorporate this noise model, their prior linguistic knowledge (i.e., what a grammatical/plausible sentence should look like), and the bottom-up perceptual evidence to derive the intended message from a given string. Applying this account to the transposition phenomenon, readers might "correct" the word order and attain a nonliteral, grammatical interpretation by, for example, attributing the perceived ungrammatical sequence to eye movement control errors (e.g., mis-located saccades or skipping; see Staub et al., 2019).

Notably, this noisy-channel account, contrary to the parallel word processing account that posits re-ordering to occur at the encoding stage (see below), assumes that the veridical, ungrammatical string is initially perceived. Following from this assumption, and adopting the further assumption of highly incremental integration of each word with its context (e.g., Christiansen & Chater, 2016), the perception of the literal, ungrammatical input should be accompanied by immediate processing difficulty ()(Frazier & Rayner, 1982; Rayner, Warren, Juhasz, & Liversedge, 2004). However, the above-mentioned literature on misreading phenomena adopting the noisy-channel account (Gibson et al., 2013; Ryskin et al., 2018; Staub et al., 2019) has not explicitly tested this real-time prediction.

Here we report two eye-tracking experiments designed to further explore failure to notice word transpositions. One previous study (Rayner, Angele, Schotter, & Bicknell, 2013) has recorded eye movements when readers encounter transposed words. However, in Rayner et al. (2013), transposed words were presented only in the parafovea, with the words reverting to the correct order once directly fixated. In the present study, we present sentences like those used by Mirault et al. (2018), where a word transposition is fully visible throughout the course of reading.

The study addresses four empirical questions. First, we address the basic question of whether readers do sometimes fail to notice word transpositions in a fairly natural reading setting. Experiment 1 used a speeded grammaticality judgment task similar to Mirault et al. (2018), but presented sentences only after a fixation on the left edge of screen, prompting participants to read from left to right. Mirault et al., by contrast, presented sentences after central fixation, which resulted in subjects initially fixating near the middle of sentences, which may have artificially inflated error rates. Experiment 2 adopted a reading for comprehension task (Staub et al., 2019), embedding the critical grammaticality-detection sentences in a larger number of sentences that were followed by comprehension questions.

Second, we ask whether the word class (part of speech) and the length difference of the transposed words influence the rate at which readers fail to notice transpositions. These two factors were explored in post-hoc analyses by Mirault et al. (2018). Here we systematically manipulate the two factors, in order to evaluate the predictions of a reading model, OB1-Reader (Snell, van Leipsig, Grainger, & Meeter, 2018), which posits parallel processing of multiple words, and which is specifically motivated in part by the existence of the transposed-word effect. According to OB1-Reader, words are recognized in parallel, and their activation is distributed across multiple spatial positions (Snell & Grainger, 2019b). Thus, not only is it possible for *cat* to be recognized prior to *was* in the sentence *The white was cat big*, but because the activation of *cat* extends from its veridical position to its surroundings, *cat* can be encoded as the third word (W3) rather than the fourth word (W4). The model predicts that the easier it is to identify the second transposed word, relative to the first, the more likely it is that this word will be recognized first and be appended to the first position (see also Reichle, Liversedge, et al., 2009), and the more likely it is that a transposed-word sentence will be judged grammatical. As closed-class words are generally higher in frequency than open-class words, and are therefore likely to be recognized more quickly (e.g., Inhoff & Rayner, 1986),[1] it is predicted that readers should be especially likely to misidentify the order of an open-closed sequence as in *I walk fat my dog every morning*.

Another key aspect in which OB1-Reader differs from previous parallel processing models (e.g., Engbert, Nuthmann, Richter, & Kliegl, 2005; Reilly & Radach, 2006) is the introduction of a spatiotopic representation in working memory (Snell, Meeter, & Grainger, 2017). Snell et al. posited that the spatiotopic representation contains low-level visual information about the lengths of the to-be-recognized words. For a word to be recognized and mapped onto a slot in the spatiotopic representation, its length must match the information for that slot. For instance, even if *her* is very likely to be recognized before *favorite* in *She repeated favorite her song ten times*, it is unlikely to be mapped to position 3 because the spatiotopic representation dictates that W3 has approximately 8 letters. Thus, a second prediction is that readers should not misidentify the order of transposed words whose lengths are very different.

Our third goal is to determine whether specific patterns of eye movements play a role in detecting transposition errors. Does direct fixation on each of the two transposed words increase detection of the error? Specifically, we test two extreme hypotheses. First, in a serial-attention model like *E*-Z Reader 10 (Reichle, Warren, & McConnell, 2009) the word order representation is always veridical, since representation of word order by default follows the relative order of word recognition (Reichle, Liversedge, et al., 2009), and words are recognized serially even when skipped. Under this model, patterns of fixations and word skips should not influence the error rate. On the other hand, OB1-Reader, following previous parallel processing models (Engbert et al., 2005; Reilly & Radach, 2006), assumes that once a word is recognized it is unlikely to be selected as a saccade target. Thus, a reader should skip W4 when this word has been recognized before W3. On an extreme version of this view, readers should rarely if ever fail to notice a transposition when they do make sequential fixations on W3 and W4, as this pattern would indicate that W4 was not, in fact, recognized before W3.

Finally, we ask if there is evidence of registration of the erroneous order, as reflected in the moment-by-moment eye-tracking data, when readers fail to explicitly report the error on that trial. As noted earlier, it is possible that a reader may initially perceive two transposed words in their ungrammatical, veridical order, but a process of rational inference then leads the reader to recover a more plausible but non-literal message (e.g., Gibson et al., 2013). With eye movement data, we can compare incremental processing of the two critical transposed words on trials when readers fail to report the error in a transposed sentence, when they do report the error, and when they correctly accept a grammatical sentence. The most straightforward prediction under a noisy channel account positing that error-correction is post-perceptual is that some disruption should be present in the eye movement record even on those trials when participants ultimately accept the transposed-word sentence as grammatical, with a mitigation of this disruption in later processing,

[1] We note here that the hypothesis is about ease of recognition, which ideally should be affected by word frequency of the two transposed words with other factors controlled. We manipulate word class mainly to follow the exploratory analysis by Mirault et al. (2018), using word class as a proxy for word frequency. Future studies should directly test the roles of both word class and word frequency.

indicating a correction. That is, the incorrect word order is initially registered, and then corrected by inference.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Fifty-three undergraduates from the University of Massachusetts Amherst participated either for course credit or payment. Two participants whose accuracy on both the grammatical and filler sentences was below 75%, and one participant who accepted all the transposed sentences, were removed from analysis, leaving 50 in the final analysis. All participants were native speakers of English with normal or corrected-to-normal vision, and no self-reported history of reading or language disorders. Informed consent was obtained for each subject.

2.1.2. Materials

Experiment 1 manipulated the part of speech (open-class, e.g., adjective or noun, vs. closed-class, e.g., determiner or preposition) and word length of the transposed words, which were always the third and fourth words of the sentences (W3 and W4). The manipulation of word class resulted in three conditions, which we name based on the order in the transposed conditions: open-closed, closed-open, and open-open; see Table 1 for Examples. We manipulated the length of the open-class words, leading to items in which both words were short ('both short'), and items in which one word was long ('one long'). Short words were always 2–4 letters, and the length difference between words in the both-short sentences was never more than one letter. Long words were 8–11 letters. Table 2 further shows descriptive statistics for frequency of each transposed word in each condition: closed-class words were comparable in frequency across all the conditions; short open-class words were also comparable without significant differences. However, short open-class words were systematically more frequent than long open-class words.

We created 30 items for each cell of the design, with each having a grammatical version and a transposed counterpart. However, for the open-open, one-long condition, either the first or the second word could be long, so we created 60 items for this cell; in the analysis, we collapsed these sub-conditions, as post hoc analyses did not reveal significant differences in error rates between these two sub-conditions. We designed the stimuli so that after transposition, W3 itself was grammatical at the point where it appeared, resulting in ungrammaticality only at W4; this is least evident in the open-closed condition, but even here a grammatical continuation is possible, e.g., *I walk fat dogs to make some extra money*. Each participant saw either the grammatical or transposed version of each item. Fifty incomplete sentences served as ungrammatical fillers, which were included in order to encourage participants to read each sentence to the end. Each participant therefore read a total of 260 sentences, consisting of 105 transposed sentences, 105 grammatical sentences, and 50 incomplete sentence fillers.

Table 1
Example item in each transposed condition. Transposed words are in bold; in grammatical conditions these words were presented in the opposite order. Word class labels reflect the order in transposed conditions.

Word class	Both short	One long
Open-closed	I walk **fat my** dog every morning.	She repeated **favorite her** song ten times.
Closed-open	The boy **on sat** the school bus.	Her professor **on commented** the latest article.
Open-open	The fragile **cup red** shattered into pieces.	Our beloved **president new** represents us well. / The boring **task repetitive** made him quit.
Fillers	I see the incredible talent in your.	

Table 2
Descriptive statistics for mean frequency (*Zipf*, Brysbaert & New, 2009) of each transposed word in each condition. Standard deviation in parenthesis.

	Word 3 Frequency	Word 4 Frequency
Open-closed (both short)	5.01 (0.63)	6.86 (0.54)
Closed-open (both short)	6.78 (0.29)	4.92 (0.73)
Open-open (both short)	4.96 (0.65)	5.01 (0.62)
Open-closed (one long)	4.13 (0.60)	6.90 (0.56)
Closed-open (one long)	7.04 (0.27)	3.27 (0.60)
Open-open (one long)	4.53 (0.65)	4.39 (0.87)

2.1.3. Procedures

The movement of the subject' s right eye was recorded using an EyeLink 1000 (SR Research, Toronto, ON, Canada) eye-tracker. The sampling rate was 1000 Hz. Subjects were seated 55 cm from a CRT monitor, with 1024 × 768 resolution and a screen refresh rate of 120 Hz. Sentences were displayed on a single line in 11-point Monaco font, with between three and four characters subtending 1° of visual angle. Participants were asked to judge whether the sentence was a well-formed English sentence as quickly and accurately as possible, by pressing a button (left = good; right = bad) on a hand-held controller. A trial began with a central dot and then a fixation box on the left of the screen. After the box was gaze-triggered, the sentence appeared offset to the right so that participants needed to make a saccade to begin reading the sentence. Once the participant made a response, the sentence disappeared, and the next trial began. Six practice trials preceded the experimental trials. The experiment took approximately 40 min.

2.2. Analysis

We fit Bayesian (generalized) linear mixed-effect models using Stan, as implemented in the *brms* package (Bürkner, 2017) for the R statistical programming environment (version 3.5.3; R Core Team, 2019), with the default priors. For each model we ran four chains each with 5000 steps (warmup = 1000 steps). Rhat statistics in all models approached 1.00 and no warnings emerged (except for an extremely small number of divergent transitions in some cases). Although *brms* can fit highly complex models, as noted in Bates, Kliegl, Vasishth, and Baayen (2015), maximal models, even when they converge, might not be supported by data and could be uninterpretable. Specific procedures of model construction and comparison are elaborated in each results section, with this caveat, as well as practical running-time concerns, in mind.

3. Results

3.1. Analysis of accuracy

Mn initial model in which sentence type (transposed sentence, grammatical sentence, or filler; treatment-coded) was the only fixed effect, and with random intercepts for subjects and items, showed that transposed sentences elicited errors (13.3%) more than fillers (incomplete sentences, 6.3%, contrast 95% CI: [0.45, 1.21]; note that this CI expresses the difference in logit) and more than grammatical counterparts (6.7%, contrast 95% CI: [0.81, 1.71]). For comparison, in Mirault et al. (2018), transposed-word sentences elicited about 16% and 12% errors in a laboratory experiment and on-line experiment, respectively, while grammatical sentences and additional ungrammatical control sentences elicited about 3% errors.

Mean accuracy as a function of word length and word class for grammatical and transposed sentences is shown in Table 3.

Fig. 1 shows the distribution of error rates, in the transposed conditions, for participants in both experiments in Mirault et al. (2018), and the two experiments in the current study. It is notable that not only is the overall error rate similar in these studies, but there is also similarity at the distributional level. In both studies, many subjects made very few errors, while a few subjects made errors on a substantial proportion of

Table 3
Accuracy (%) in each condition; standard error (by subjects) in parenthesis.

	Grammatical	Transposed
Open-closed, both short	94.8 (1.2)	85.7 (2)
Open-closed, one long	94.3 (0.9)	90.8 (1.9)
Closed-open, both short	90.1 (1.2)	89.9 (1.6)
Closed-open, one long	97.6 (0.6)	91.3 (1.7)
Open-open, both short	90.9 (1.4)	75.9 (3)
Open-open, one long	92.5 (1.5)	86.8 (2.2)
Filler	93.7 (1.3)	

the transposed words: If the participant skipped at least one of the words on their first pass through the sentence, the trial is coded as a skipping trial, and if she fixated both words on first-pass reading, a fixating-both trial. This two-level factor was centered with skipping coded as 0.5 and fixating-both as −0.5.

Table 4 presents the percentage of trials on which both transposed words were fixated, contingent on response in each experimental condition.[2]

We started with full interactive (Word Class * Word Length * Fixation Pattern) models with subject and item random intercepts and all

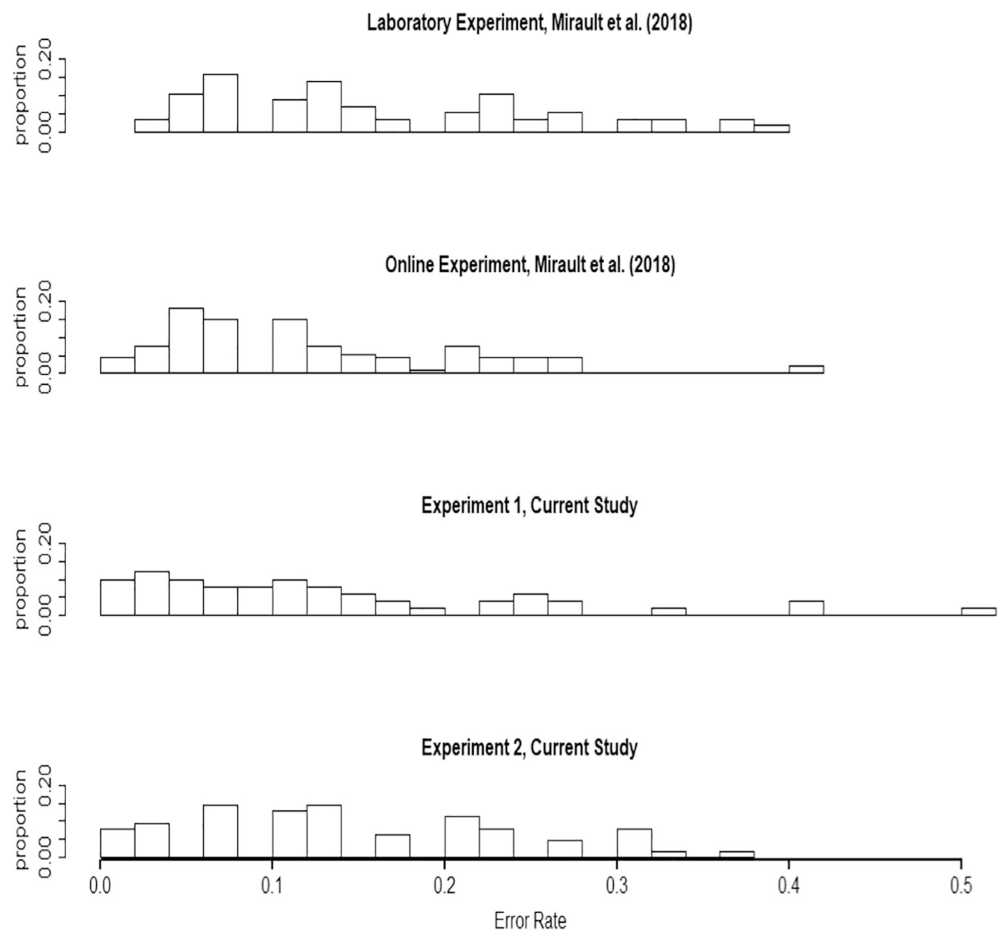

Fig. 1. Histogram of error rate by subject in each experiment.

trials. To explore this individual variation, we correlated subjects' accuracy for grammatical, transposed, and filler incomplete sentences in the present experiment. There was no significant correlation between the first two ($r = -0.13$) but a significant correlation between the latter two ($r = 0.37$, $p < .01$, Fig. 2), with subjects who more frequently failed to report the ungrammaticality in the fillers also more frequently failing to report the transposition sentences to be ungrammatical. It is also worth noting that after removal of a single subject with very low filler accuracy but high accuracy on the transposed sentences, this correlation increases to $r = 0.67$.

For the main statistical analyses, we included only data from the transposed sentences, including as fixed effects word class, word length, and fixation pattern (as described below), and their interactions. Word class compared open-closed, closed-open, and open-open sequences, using treatment contrasts with open-closed as the reference level. Word length was a two-level factor, centered with the both-short condition coded as 0.5 and the one-long condition coded as −0.5. Finally, the fixation pattern factor was defined by participants' eye-movement on

possible random slopes (i.e., if not a between-item effect), without correlation parameters. A random slope of an effect was then removed if its sd estimate was not reliably higher than zero and/or its sd was relatively small compared to variances of the other item/subject random effects (Bates et al., 2015). The reduced model was then compared to the all-slopes model by *loo_compare* using the *waic* criterion to confirm that the random slopes were not necessary; random slopes were retained otherwise (elpd difference < 1.96 * standard error of difference). The

[2] Since fixation pattern is not an experimental variable, one might be concerned with multicollinearity when treating it as an independent variable along with the other word class and length variables. While it is true that fixation rates for a single word (Word 3 or Word 4) were highly sensitive to the class and length of that word, when calculating fixation rates for the critical region (i.e., both Words 3 and 4 being fixated), the correlations were reduced, and when including all three variables into a model to explain transposition detection accuracy, all the variance inflation factors (VIFs) were below 3.87, which we believe warrants only little concern (Hair, Black, Babin, & Anderson, 2010).

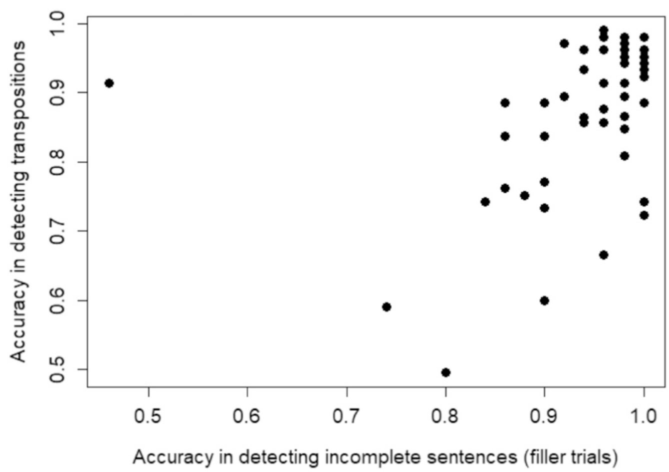

Fig. 2. Scatter plot of participants' accuracy in detecting the incomplete sentence in filler trials and their accuracy in detecting the transposition in critical trials.

Table 4
Percentage of trials on which both transposed words were fixated, as a function of response accuracy in Experiment 1; means and standard errors (in parenthesis) calculated by-subject. Note that not all subjects made errors in all transposed conditions.

	Transposed (correct response)	Transposed (incorrect response)	Grammatical (correct)
Open-closed, both short	0.46 (0.04)	0.37 (0.05)	0.36 (0.03)
Open-closed, one long	0.46 (0.04)	0.37 (0.06)	0.42 (0.03)
Closed-open, both short	0.41 (0.03)	0.30 (0.04)	0.34 (0.03)
Closed-open, one long	0.36 (0.03)	0.24 (0.04)	0.32 (0.03)
Open-open, both short	0.64 (0.04)	0.42 (0.05)	0.54 (0.03)
Open-open, one long	0.76 (0.03)	0.70 (0.05)	0.74 (0.03)
Combined	0.54 (0.03)	0.50 (0.04)	0.49 (0.03)

parsimonious full interactive model was subsequently compared with the parsimonious two-way interactive models, one-way interactive models, and additive-only models, using *loo_compare*. Correlation terms were added to the ultimately favored model, and comparison was made between the with- and without-correlation models.

Fig. 3 shows the mean accuracy broken down by the three factors. No

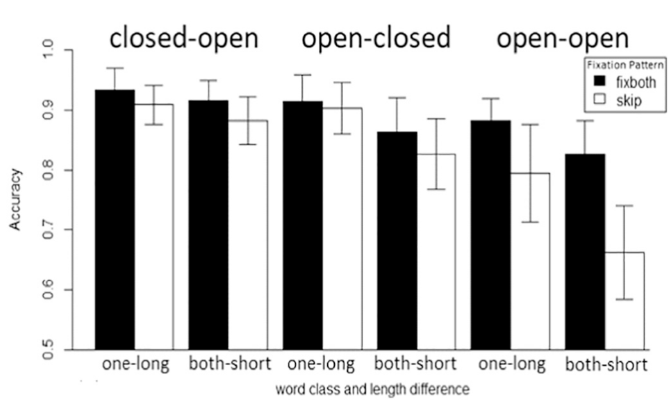

Fig. 3. Accuracy on the transposed sentences, grouped by word class, word length, and fixation pattern. Error bars reflect 95% confidence interval, by subjects.

models with interaction terms were favored. The resulting model was an additive model with random slopes for word class and word length and intercepts, by subjects, and random intercepts by items, without correlations. We found effects of all three factors—word length, word class, and fixation pattern (Table 5). The open-open condition elicited many more errors than the open-closed and closed-open conditions. There was a hint that the accuracy of the open-closed condition is lower than that of closed-open condition (contrast 95%CI: [−0.01, 0.96]). The both-short condition elicited more errors than the one-long condition. Direct fixation reduced the error rate to 11% compared to 16% when at least one of the two words were skipped; testing this contingency other way around, on 50% of trials when the transposition was not detected, both words were directly fixated, compared to 54% when the transposition was detected (Table 4).

3.2. Analysis of eye movements

Short fixations (<80 ms) within 1 character position of another preceding or following fixation were automatically combined. Eye-movement data were analyzed separately on each of Words 3, 4, and 5 (spillover), and on the final region of each sentence. Word 4 (W4) is the critical point at which a transposed sentence becomes ungrammatical. Eye movement measures were analyzed contingent on response type: We compared trials where participants Accepted a Transposed sentence (AT; i.e., failure to detect the transposition), trials where participants Accepted a Grammatical sentence (AG), and trials where participants Dismissed a Transposed sentence (DT). We omitted the relatively small number of trials on which participants dismissed a grammatical sentence. In our statistical models, this response type variable was the only fixed effect, treatment-coded with AT as the reference level. Because the number of AT responses was small and differed greatly across participants, further subdividing by word length and word class conditions would lead to many subjects having no AT observations in many cells of the design.[3]

We report four standard fixation time measures (Rayner, 1998): two early processing measures—first fixation duration (FFD; duration of the first fixation on a word) and gaze duration (GD, the sum of the durations of all fixations on a word before leaving it) —and two later processing measures—go-past time (GP, the sum of the durations of all fixations from the first on the word until the word is exited to the right, including any regressive re-reading) and total viewing time (TVT, the sum of the durations of all fixations on a word). All four measures included only observations of non-zero values; if a word was skipped on first-pass reading, that trial is not included in the calculation of FFD, GD, or GP, and if a word was not fixated at all, that trial is not included in the calculation of total time. We also analyzed two binary dependent measures: whether first pass reading of a word ended with a regression rather than a forward saccade, and whether there was a regression into a word from later in the sentence. Word 3 in AT and DT was compared with Word 4 in AG, and vice versa, so that the exact same words were compared.

Random slopes were first included for both subjects and items, without correlation parameters. A random slope of an effect was then removed if its sd estimate was not reliably higher than zero and/or its sd was relatively small compared to variances of the other item/subject random effects (Bates et al., 2015). The reduced model was then compared to the all-slopes model by *loo_compare* (criterion = "waic") to

[3] However, to account for the known effects of word length and word class on reading times and minimize the residual errors, we added the two factors as fixed main effects in the models. Effects of response type are qualitatively identical and quantitatively similar with or without the additional two factors. For the sake of exposition, here we present models with response type as the only fixed factor. Results of alternative analyses are available on OSF: https://osf.io/vhguf/

Table 5
Results of the Bayesian generalized linear mixed effect model (ACC ~ word class + word length + fixation pattern; linking function = logit). class 1: closed-open minus open-closed; class 2: open-open minus open-closed. Random effects are omitted for the ease of exposition; see full results on OSF.

Effect	Estimate	Error	l-95%CI	u-95%CI	Rhat	Bulk_ESS	Tail_ESS
intercept	2.64	0.21	2.23	3.07	1.00	9681	11,254
class1	0.45	0.25	−0.01	0.96	1.00	13,120	11,855
class2	−0.62	0.22	−1.05	−0.21	1.00	13,857	12,173
fixation	−0.48	0.11	−0.69	−0.27	1.00	32,310	11,903
length	−0.64	0.17	−0.97	−0.32	1.00	17,068	13,718

confirm that the random slopes were not necessary; the random slopes were retained otherwise. Correlation terms were finally added and compared to the no-correlation models. In this case, all no-correlation models were favored.

Tables 6, 7, and 8 present the models for all measures for each region. On the critical region (W4), participants showed more difficulty in DT than in AT across measures; i.e. on the trials when subjects did notice the transposition (DT), their eye movements showed disruption across the full range of measures, compared to when they did not notice the transposition (AT). However, there was no evidence of differences between AT and AG on W4. One less consistent finding is that the TVT on W3 was higher in AT than in AG, coupled with a higher regression-in rate. Furthermore, a potentially surprising finding was that FFD, GD, and GP were all longer for DT than the other conditions at W3 (recall that W3 itself was grammatical at that point). Finally, in the final region, AT again paralleled AG while reading times were significantly faster in DT.

4. Discussion

The findings of Experiment 1 can be summarized as follows. First, using a more natural mode of presentation in which subjects read from left to right, we replicated the basic transposition effect reported by Mirault et al. (2018): Subjects failed to notice that sentences contained transposed words more often than they made errors in judging grammatical sentences, and more often than they made errors in judging incomplete sentences. In the present study, the rate of failing to notice a transposition was comparable to the Mirault et al. study, but the error rate in both the grammatical and filler sentences was higher than in that study. The slightly worse performance in the present grammatical condition than in Mirault et al. might have been due to the greater sentence length of our stimuli, or to word length constraints on our target words, which might have made the sentences less natural, albeit grammatical. The incomplete sentence fillers in the present experiment and the ungrammatical control sentences in Mirault et al. are very different, so comparison of the error rates is of limited utility. Furthermore, participants varied substantially in their rate of transposition error detection, as in Mirault et al. Subjects who failed to notice the transposition errors were also less sensitive to the errors in the filler sentences.

We also found that the tendency to fail to notice a transposition was modulated by both word class and word length. Such failures were most common when the transposed words were both open-class words, and when both were short. There was only a slight tendency for open-closed sequences to lead to more judgment errors than closed-open sequences. In addition, errors were more likely when readers skipped one of the words on first pass reading than when they directly fixated both. Finally, while detecting a transposition corresponded to clear first-pass difficulty in the eye movement record, we found no evidence that reading of the ungrammatical word (W4) was disrupted on those trials on which the transposition was not explicitly reported.

We reserve detailed discussion of this complex pattern of results for the General Discussion. We first present Experiment 2, which was motivated by the possibility that an explicit error-detection task, as used both in our Experiment 1 and in Mirault et al. (2018), may not be ideal

Table 6
Response type effects on eye-movements on Word 3 in Exp 1. AG: accepting grammatical sentences; AT: accepting transposed sentences; DT: dismissed transposed sentences. Random effects are omitted for the ease of exposition; see full results on OSF.

Effect	Estimate	Error	l-95%CI	u-95%CI	Rhat	Bulk_ESS	Tail_ESS
FFD							
Intercept	227.91	5.48	217.17	238.61	1.00	3509	6676
AG-AT	−5.86	4.16	−14.02	2.34	1.00	11,451	10,494
DT-AT	9.30	4.56	0.34	18.16	1.00	10,028	11,497
Gaze							
Intercept	249.59	7.56	234.66	264.43	1.00	4241	6857
AG-AT	−4.87	5.67	−16.08	6.20	1.00	14,817	11,309
DT-AT	16.60	6.20	4.45	28.66	1.00	12,913	11,292
Go-past							
Intercept	294.16	11.69	271.48	317.37	1.00	5484	8324
AG-AT	−6.08	9.43	−24.42	12.33	1.00	14,321	11,965
DT-AT	43.30	11.40	20.70	65.82	1.00	10,442	11,455
Total Viewing							
Intercept	361.07	15.18	331.07	390.73	1.00	4586	8688
AG-AT	−40.61	10.33	−60.63	−19.96	1.00	16,422	10,862
DT-AT	45.56	13.10	19.70	71.24	1.00	10,230	11,788
Regress-in							
Intercept	−0.98	0.13	−1.24	−0.72	1.00	7903	10,716
AG-AT	−0.58	0.12	−0.82	−0.34	1.00	16,432	12,293
DT-AT	0.32	0.13	0.08	0.57	1.00	13,238	11,807
Regress-out							
Intercept	−2.20	0.19	−2.58	−1.84	1.00	12,429	11,843
AG-AT	−0.24	0.18	−0.59	0.12	1.00	17,223	11,937
DT-AT	0.22	0.18	−0.14	0.57	1.00	18,053	11,276

Table 7
Response type effects on eye-movements on Word 4 in Exp 1. AG: accepting grammatical sentences; AT: accepting transposed sentences; DT: dismissed transposed sentences. Random effects are omitted for the ease of exposition; see full results on OSF.

Effect	Estimate	Error	l-95%CI	u-95%CI	Rhat	Bulk_ESS	Tail_ESS
FFD							
Intercept	233.69	5.74	222.40	244.90	1.00	3028	6246
AG-AT	−5.83	4.38	−14.53	2.75	1.00	11,237	10,704
DT-AT	11.64	4.52	2.66	20.50	1.00	10,653	10,410
Gaze							
Intercept	260.20	8.54	243.49	276.88	1.00	2683	5957
AG-AT	−7.94	6.34	−20.25	4.33	1.00	12,334	11,651
DT-AT	21.53	6.51	8.73	34.48	1.00	12,464	11,833
Go-past							
Intercept	313.53	14.33	285.51	341.55	1.00	7591	9734
AG-AT	−3.97	11.78	−26.99	19.47	1.00	18,816	11,091
DT-AT	74.11	14.51	45.53	102.61	1.00	11,597	11,432
Total Viewing							
Intercept	362.64	16.52	330.73	395.22	1.00	3902	6985
AG-AT	−12.53	11.44	−34.88	10.14	1.00	12,750	10,145
DT-AT	48.15	12.60	23.32	72.74	1.00	11,148	12,154
Regress-in							
Intercept	−1.13	0.12	−1.36	−0.90	1.00	7986	10,083
AG-AT	−0.19	0.13	−0.44	0.06	1.00	8642	10,253
DT-AT	0.17	0.11	−0.04	0.39	1.00	12,656	11,555
Regress-out							
Intercept	−1.98	0.18	−2.33	−1.64	1.00	8635	10,455
AG-AT	−0.21	0.18	−0.56	0.15	1.00	11,254	11,271
DT-AT	0.56	0.18	0.21	0.90	1.00	10,281	11,091

Table 8
Response type effects on eye-movements on spillover and final regions in Exp 1. AG: accepting grammatical sentences; AT: accepting transposed sentences; DT: dismissed transposed sentences. Random effects are omitted for the ease of exposition; see full results on OSF.

Effect	Estimate	Error	l-95%CI	u-95%CI	Rhat	Bulk_ESS	Tail_ESS
Gaze (spillover)							
Intercept	259.97	7.38	245.70	274.46	1.00	10,534	11,668
AG-AT	−6.59	6.50	−19.39	6.00	1.00	20,371	12,176
DT-AT	8.88	7.12	−5.14	22.77	1.00	14,672	12,148
Total Viewing (spillover)							
Intercept	339.04	13.91	312.03	366.58	1.00	3936	7354
AG-AT	−5.79	11.57	−27.94	17.26	1.00	6848	9256
DT-AT	7.03	12.38	−17.50	31.00	1.00	5797	9736
Gaze (final)							
Intercept	403.28	18.05	368.46	439.29	1.00	4656	7740
AG-AT	17.59	13.01	−7.97	43.30	1.00	11,349	11,169
DT-AT	−119.90	12.83	−145.28	−94.68	1.00	11,230	10,895
Total Viewing (final)							
Intercept	485.11	25.49	435.55	536.38	1.00	4382	7255
AG-AT	34.96	15.75	4.85	66.84	1.00	10,980	11,700
DT-AT	−141.62	18.08	−177.95	−106.6	1.00	8687	10,326

for investigating the extent to which transpositions are noticed in normal reading. In Experiment 2 we adapted a paradigm used by Staub et al. (2019) in an experiment investigating readers' failures to notice repeated words. Critical trials that are followed by grammaticality judgment questions are intermixed with a larger number of filler trials that are followed by comprehension questions. Readers do not know, until a given sentence has been removed from the screen, which kind of question they will be asked. This paradigm at least somewhat reduces the demand to treat the task as proofreading, while at the same time still providing information as to whether the transposition was noticed on each trial.

5. Experiment 2

5.1. Methods

5.1.1. Participants

There were 65 participants in Experiment 2, from the same pool as in Experiment 1. One participant who self-reported ADHD after the experiment and one who did not complete the experiment were removed from the analysis, leaving 63 subjects. Informed consent was obtained for each subject. All participants had accuracy above 75% for both the grammatical and filler sentences.

5.1.2. Stimuli and procedures

Every sentence was followed by a question that appeared after the sentence was removed from the screen. Subjects were not aware, while reading a sentence, whether they would be tested for comprehension or for grammaticality detection. Of the total 180 sentences, two-thirds (120) were filler trials that were followed by a comprehension question, while one-third (60) were experimental trials followed by the question, 'Was there anything wrong with that sentence?' Half of the 60 experimental trials were grammatical and half contained a transposition error. Due to the large number of filler trials, the number of critical trials in each condition was reduced from 15 to 10, and we used only the both-short condition from Experiment 1. Manipulation of word class was preserved. There was no statistical difference in the mean frequency for the open-class words in each condition ($|t|$s < 0.46, all mean *Zipf* values around 5).

6. Results

6.1. Analysis of accuracy

Statistical analyses were performed in the same way as Experiment 1. Mean accuracy on the filler comprehension questions was 95.5% (Min = 87.5%), suggesting participants were paying attention to the task. Table 9 shows mean accuracy in each condition. The average error rate for grammatical sentences was 7.7%, with the error rate for the open-open condition being particularly high (13.2%). The average error rate for transposed sentences was 14.4%, reliably higher than their grammatical counterparts (contrast 95% CI: [1.04, 2.24]). Accuracy as a function of fixation pattern and word class is shown in Fig. 4.

Model specification procedures were the same as Experiment 1, with word class treatment-coded with open-closed as the reference level and with fixation pattern sum-coded. Table 10 further shows fixation rates on the critical region as a function of word class.[4]

As in Experiment 1, the additive model was favored, with only fixation pattern showing an effect (Table 11). On trials when the reader made a direct fixation on both words the mean error rate was reduced to 11% compared to 17% when at least one of the words were skipped; testing this contingency other way around, on 37% of trials when the transposition was not detected, both words were directly fixated, compared to 47% of trials when the transposition was detected (Table 10).

As Fig. 1 indicates, there was again substantial subject-level variability in the error rate in the transposed conditions. A correlational analysis by Staub et al. (2019) suggested that slower readers, as determined based on mean total sentence reading time for the filler sentences that were followed by comprehension questions, were somewhat more accurate in detecting a repeated *the*, though this relationship did not quite reach conventional statistical significance ($r = 0.26$, $p = .07$). We computed the same correlation for the present experiment. This correlation was of similar magnitude, but significant ($r = 0.34$, $p < .01$, Fig. 5).

6.1.1. Analysis of eye movements

Short fixations (<80 ms) within 1 character position of another preceding or following fixation were automatically combined. As in Experiment 1, including word class as another fixed effect did not alter the patterns of the results; here we report models including only response type as a fixed effect, treatment-coded with AT as the reference. Tables 12, 13, 14 show the effect estimates on different regions across measures. For some measures, models with correlation terms are favored; however, the response type effects were qualitatively identical and quantitatively similar with or without correlation terms. We therefore present the without-correlation models for the sake of exposition.[5] As in Experiment 1, there was evidence of early and sustained disruption in DT trials compared to AT trials on W4. However, there was very little evidence that readers in AT trials encountered processing difficulty at all throughout the sentences, based on comparison to AG trials; the only exception was a rather small difference in regress-in rates (contrast 95%CI: [−0.64, −0.01]) and a difference in regress-out rates on Word 3. Response time to the error-detection question in DT trials was about 800 ms faster than in the other conditions, suggesting that participants noticed the error ahead of the probe question, while in AT trials participants' response time was as long as that in AG. Finally, in contrast to Experiment 1, indication of early disruption on Word 3 in the transposed condition was evident only in the regressions-out and possibly go-past measures.

Table 9
Accuracy (%) in each condition; standard error (by-subjects) in parenthesis.

	Grammatical	Transposed
Open-closed	95.2 (1.4)	86.2 (1.6)
Closed-open	94.9 (1)	86.6 (1.8)
Open-open	86.8 (2)	84.1 (1.6)
Filler, Comp. Question	95.5 (2.6)	

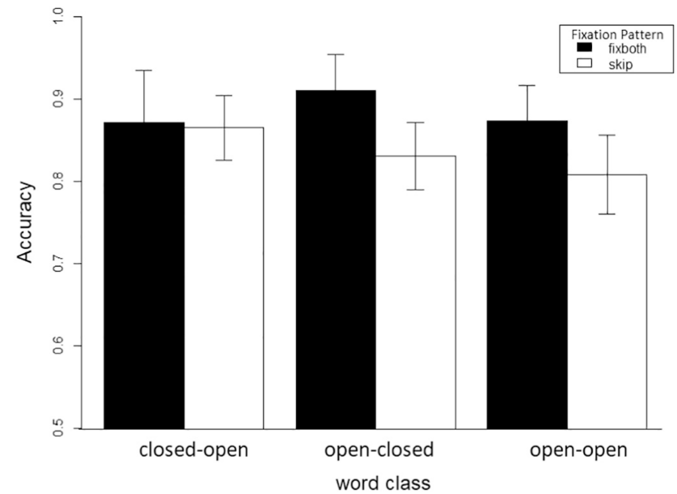

Fig. 4. Accuracy on the transposed sentences, grouped by word class and fixation pattern in Experiment 2. Error bars reflect 95% confidence interval (by subjects).

Table 10
Percentage of trials on which both transposed words were fixated, as a function of response accuracy in Experiment 2; means and standard errors (in parenthesis) calculated by-subject. Note that not all subjects made errors in all transposed conditions.

	Transposed (correct responses)	Transposed (incorrect responses)	Grammatical (correct)
Open-closed	0.44 (0.03)	0.24 (0.05)	0.39 (0.03)
Closed-open	0.38 (0.03)	0.33 (0.05)	0.31 (0.03)
Open-open	0.58 (0.03)	0.44 (0.05)	0.55 (0.03)
Combined	0.47 (0.02)	0.37 (0.04)	0.41 (0.02)

7. Discussion

In a task that emphasized reading for comprehension, we again found a transposed-word effect. Unlike in Experiment 1, however, there was no effect of word class, as all three kinds of sequences elicited

[4] As in Experiment 1 there was no strong indication of a multicollinearity issue; all VIFs <3.23 for a model with both word class and fixation pattern as independent variables.

[5] Results of those models, together with the models that include word class as another fixed factor, are available on OSF.

Table 11
Results of the Bayesian generalized linear mixed effect model (ACC ~ word class + fixation pattern; linking function = logit). class 1: closed-open minus open-closed; class 2: open-open minus open closed.

Effect	Estimate	Error	l-95%CI	u-95%CI	Rhat	Bulk_ESS	Tail_ESS
intercept	2.14	0.21	1.74	2.57	1.00	10,981	11,962
class1	0.19	0.29	−0.36	0.79	1.00	8959	9327
class2	−0.17	0.27	−0.70	0.38	1.00	11,299	12,025
fixation	−0.51	0.17	−0.84	−0.18	1.00	21,083	12,394

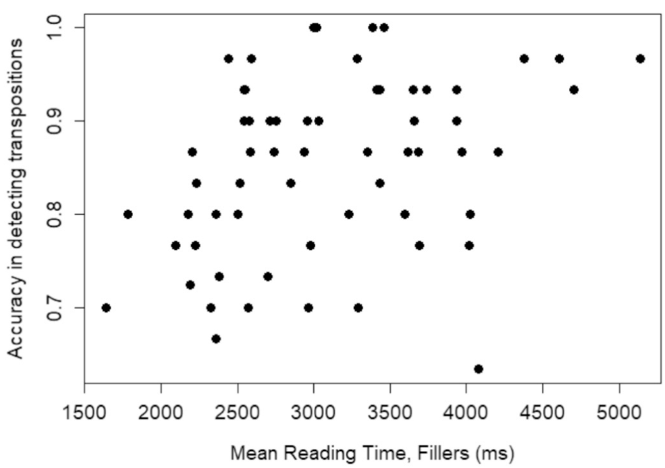

Fig. 5. Scatter plot of participants' overall reading time on the filler sentences and their accuracy in detecting the transposition.

similar error rates. Fixation pattern modulated the transposed-word effect as in Experiment 1, and the effect of fixation pattern did not differ across the three word-class conditions. Finally, we saw a consistent pattern whereby eye movements on sentences containing transposition errors were contingent on response type, as in Experiment 1, such that processing difficulty was very clearly present on trials when subjects did explicitly notice the transposition (DT), but was not present on trials when participants failed to report the transposition (AT).

7.1. Supplementary informed Bayesian analysis

Arguably, the lack of difference between AG and AT trials in the eye movement record could be due to limited power, as the number of AT trials is relatively small. To address this issue, we ran additional Bayesian analyses on each transposed word under an a priori rational inference plus incremental integration hypothesis according to which registration of an error from the perceived ungrammatical string is expected at the earliest fixation stage (see Appendix for full descriptions of the implementation and results; here we include results only for Experiment 2 but Experiment 1 results are very similar). The analyses showed that the patterns of AG-AT and DT-AT contrasts depend on the priors. That is, when we presume disruption from the ungrammaticality of the AT trials in our priors, the resulting posteriors indicate a reliable difference between AG and AT trials as early as FFD, different from our results under uninformative priors. While such a finding could, in theory, suggest that there was immediate disruption for the AT trials, a more appropriate conclusion is simply that the posterior is highly sensitive to the prior because we currently do not have enough data to strongly support either a null or non-null difference (Depaoli, Winter, & Visser, 2020).

But even if a difference in FFD between AG and AT trials exists, it is quite clear that there also exists a difference in FFD between DT and AT trials. Such a difference at the earliest fixation stage in the latter contrast, despite their identical perceptual input, is not compatible with the strong view that *all* transposed sentences, even those whose transposition is not explicitly reported, are first perceived as in their literal order (Gibson et al., 2013; Ryskin et al., 2018) and that such perception of an ungrammatical sequence should immediately elicit processing

Table 12
Response type effects on eye-movements on Word 3 in Exp 2.

Effect	Estimate	Error	l-95%CI	u-95%CI	Rhat	Bulk_ESS	Tail_ESS
FFD							
intercept	214.67	7.64	199.79	229.74	1.00	17,676	12,098
AG-AT	9.38	7.48	−5.32	23.98	1.00	21,519	12,654
DT-AT	13.57	8.32	−2.77	29.76	1.00	16,752	12,244
Gaze							
intercept	233.25	10.17	213.03	253.30	1.00	17,656	12,935
AG-AT	10.30	10.05	−9.38	30.28	1.00	20,476	12,523
DT-AT	19.62	11.19	−2.30	41.67	1.00	15,249	12,616
Go-past							
Intercept	356.58	20.91	316.09	397.79	1.00	13,647	10,910
AG-AT	−29.84	21.07	−71.44	11.62	1.00	14,595	11,080
DT-AT	38.43	25.17	−10.52	87.89	1.00	9451	10,329
Total Viewing							
Intercept	379.04	20.73	338.41	419.62	1.00	14,014	11,631
AG-AT	−28.97	19.20	−66.32	8.80	1.00	21,385	11,879
DT-AT	171.17	26.60	119.13	223.48	1.00	9046	10,884
Regress-in							
Intercept	−0.82	0.16	−1.13	−0.51	1.00	18,783	12,444
AG-AT	−0.33	0.16	−0.64	−0.01	1.00	21,293	12,279
DT-AT	1.32	0.15	1.02	1.63	1.00	23,052	12,235
Regress-out							
Intercept	−0.93	0.21	−1.35	−0.53	1.00	16,373	12,749
AG-AT	−0.60	0.21	−1.01	−0.18	1.00	18,700	12,511
DT-AT	−0.10	0.20	−0.51	0.30	1.00	18,481	12,668

Table 13
Response type effects on eye-movements on Word 4 in Exp 2.

Effect	Estimate	Error	l-95%CI	u-95%CI	Rhat	Bulk_ESS	Tail_ESS
FFD							
intercept	222.82	8.38	206.41	239.20	1.00	14,172	11,866
AG-AT	−1.68	8.15	−17.73	14.24	1.00	17,628	12,681
DT-AT	25.51	9.16	7.32	43.20	1.00	13,807	11,891
Gaze							
Intercept	242.05	10.91	220.70	263.30	1.00	18,881	13,068
AG-AT	−4.08	10.47	−24.57	16.42	1.00	21,346	12,375
DT-AT	40.01	12.79	15.19	65.21	1.00	13,158	12,106
Go-past							
Intercept	360.47	27.83	306.07	415.24	1.00	17,947	12,986
AG-AT	−11.35	27.56	−65.07	42.65	1.00	22,078	12,751
DT-AT	164.65	31.76	101.38	226.95	1.00	16,582	12,516
Total Viewing							
Intercept	372.95	21.44	330.82	414.49	1.00	10,429	11,445
AG-AT	−17.44	20.19	−56.83	22.25	1.00	14,168	11,938
DT-AT	143.70	24.54	95.64	192.42	1.00	7866	10,412
Regress-in							
Intercept	−0.97	0.15	−1.27	−0.67	1.00	13,124	10,840
AG-AT	−0.16	0.16	−0.48	0.17	1.00	11,917	11,468
DT-AT	0.66	0.17	0.33	1.00	1.00	10,524	10,874
Regress-out							
Intercept	−1.24	0.21	−1.66	−0.84	1.00	18,651	11,820
AG-AT	−0.15	0.26	−0.67	0.36	1.00	11,387	11,135
DT-AT	0.81	0.22	0.38	1.25	1.00	15,992	12,048

Table 14
Response type effects on eye-movements on spillover and final regions and decision time in Exp 2.

Effect	Estimate	Error	l-95%CI	u-95%CI	Rhat	Bult_ESS	Tail_ESS
Gaze (spillover)							
Intercept	263.08	11.29	241.18	285.23	1.00	11,104	11,997
AG-AT	1.42	11.36	−20.83	23.64	1.00	11,446	11,463
DT-AT	−3.47	11.09	−25.04	18.01	1.00	13,053	12,028
Total Viewing (spillover)							
Intercept	374.86	21.74	332.22	417.60	1.00	8493	10,740
AG-AT	−8.37	17.60	−42.62	26.16	1.00	21,332	12,526
DT-AT	63.88	20.12	24.51	103.63	1.00	15,992	11,737
Gaze (final)							
Intercept	373.72	19.16	336.34	411.39	1.00	13,009	12,290
AG-AT	42.43	19.80	3.61	81.00	1.00	16,483	12,496
DT-AT	−71.29	15.15	−101.36	−41.64	1.00	25,053	10,895
Total Viewing (final)							
Intercept	581.04	35.29	512.10	650.86	1.00	4028	6657
AG-AT	−7.39	26.33	−58.49	44.41	1.00	8183	11,249
DT-AT	−127.64	24.66	−177.16	−80.40	1.00	9290	10,679
Decision Time (in response to the detection question)							
Intercept	2135.09	63.02	2011.9	2259.2	1.00	18,057	12,964
AG-AT	57.02	79.49	−96.67	212.74	1.00	11,234	11,960
DT-AT	−798.43	57.23	−911.32	−685.8	1.00	31,557	12,268

difficulty (Frazier & Rayner, 1982; Rayner, Warren, Juhasz, & Liversedge, 2004) before being corrected by rational inference. In the following General Discussion, we discuss implications of these findings in relation to rational inference and incremental integration in reading.

8. General discussion

The present study had four main empirical goals. The first was simply to determine whether readers do sometimes fail to notice word transpositions when they read relatively normally, as opposed to making grammaticality judgments to sentences presented at central fixation, as in Mirault et al. (2018). The second was to assess how lexical variables (word length and word class) may modulate this phenomenon. The third was to determine whether a reader's specific pattern of fixations on the transposed words modulates this effect, i.e., fixating both words or skipping at least one. Finally, we used the eye movement record to assess whether there is evidence of incremental processing difficulty when a reader fails to explicitly report a transposition. We first discuss the conclusions from our experiments regarding each of these four issues, and then turn to the theoretical implications of our findings.

With respect to the first goal, we confirmed that readers sometimes fail to notice transpositions under more nearly normal reading circumstances. The rate of failure to notice transposition errors was quite similar in Experiments 1 and 2, and similar in these experiments and in Mirault et al. (2018). Subjects also showed a similar (and high) degree of variability in their error rate in the present study and in Mirault et al.

One might expect that the less explicit task demands emphasizing error detection in Experiment 2 would make readers even less likely to notice errors (Schotter, Bicknell, Howard, Levy, & Rayner, 2014). That such a difference did not occur might possibly be because Experiment 1 also emphasized speed, and the speed-accuracy-trade-off counteracted the task instruction effect. However, a pattern worth noting is that while accuracy was not different, the extent to which participants' eye movements were disrupted by the transpositions differed between the experiments, with disruption in natural reading being two to three times the magnitude as in the speeded-grammaticality judgment task (Tables 7 and 13); thus the task did modulate eye movement control (Schotter et al., 2014).

Regarding the potential modulation of the transposed-word error rate by the specific properties of the transposed words, first, in Experiment 1 the effect was reduced when the two transposed words were very different in length. The difference between the one-long and both-short conditions is consistent with the predictions of OB1-Reader, according to which the length of W4 must match the length of the slot for position 3 in the spatiotopic representation of the words. However, it should be noted that the one-long items – in which one word was four or fewer letters, and the other was eight or more letters – still elicited significantly more errors than control sentences (grammatical-transposed contrast 95% CI: [0.47, 1.34]; filler-transposed contrast 95% CI: [0.53, 1.42]). Furthermore, the natural confound of long words being less frequent hinders a strong conclusion about the effect being visual or lexical. Future work is needed to disentangle the two factors. Second, in both experiments, the difference between open-closed and closed-open conditions was not reliable. A prediction under OB1-Reader model, or under parallel processing models more generally, is that an easier word, such as a high-frequency closed-class word, at the second position of the transposed sequence should make it more likely that the words are encoded out of order (Reichle, Liversedge, et al., 2009). The present data do not support this prediction (but see footnote 1).

As for the role of fixation patterns in detecting a transposition, in both experiments we found that when the participants skipped at least one of the transposed words, they were less likely to detect the transposition. This effect of skipping resembles that found in Staub et al. (2019), where skipping was related to detection of repetition and omission errors.

Finally, from the analyses of eye movements we saw very little evidence that readers experienced incremental processing difficulty in AT trials. In Experiment 1, there were differences between AT and AG trials in the regress-in and total viewing time measures on Word 3, which is inconsistent with the overall picture, and could undermine the conclusion that there is no incremental processing difficulty at all when transposed words are followed by an explicit "no-error" response. However, we note that this disruption occurred at a relatively late stage, i.e., *after*, but not *during*, first-pass reading of W4, the point of ungrammaticality. The fact that in AT trials there was no processing difficulty on W4 but more re-processing on W3 suggests that the re-processing was not due to detection of the transposition error. Tentatively, we can only speculate that it is attributable to the unnatural speeded grammaticality judgment task used in Experiment 1. In Experiment 2, the no-cost pattern in AT trials was present across all measures and regions, except for regressions out of W3, in which case the interruption occurred unexpectedly early. Overall, there is no robust evidence of disruption in AT trials across the two experiments. Finally, it is also worth noting that we have recently reanalyzed data from Staub et al. (2019), focusing on the question of whether incremental processing difficulty is present in trials when readers failed to notice a repetition of *the*. Similar to the results of the present study, we find no evidence of disruption in the eye movement record in these trials compared to the reading of grammatical sentences with a single instance of *the*. By contrast, immediate disruption is evident at the second instance of *the* on those trials when readers did ultimately notice the repetition.

Admittedly, there remains a power issue in not having detected a difference between AG and AT trials, based on the informative Bayesian analyses. Future studies with greater power should be done to confirm the current findings. Importantly, however, the current data clearly showed a difference between DT and AT trials as early as FFD, hinting that for most, if not all, of the AT trials, there was no registration of the transposition error.

In the remainder of this discussion we consider how these results constrain theoretical accounts of the failure to notice transposed words. The results are informative both for models of eye movement control and for models of syntactic and semantic processing.

8.1. The role of fixation pattern in sentence comprehension

The *E*-Z Reader model dictates that, except for accidental skipping due to saccadic overshoot, skipping occurs when the skipped word is partially recognized, with the L1 familiarity-check stage completed during parafoveal processing (Choi & Gordon, 2014; Gordon, Plummer, & Choi, 2013). The model also assumes that skipped words go through the same word processing routine (i.e., Visual processing (V), L1, L2, and Integration (I) stages) as fixated words do (Reichle, Warren, et al. 2009). In terms of *word identification*, the consequences of skipping and fixation are no different. Thus, the most straightforward prediction from E-Z Reader is that skipping should play no role in failure to notice word transpositions. Our results, however, suggest that skipping is indeed detrimental to *word order representation*. Note that this result does not, by itself, override the assumption of serial word recognition. A possibility is that recognition of two words within one fixation event, rather than two separate fixation events, makes the representation of their relative order less definitive, even when the two words have been recognized serially, in the correct order. The strength of the perceptual evidence may come into play later, in making a rational inference about word order (Norris, 2006). That is, the perceived sequence is more likely to be discounted when not every word was directly fixated. Along these lines, Staub et al. (2019) found that fixating on both instances of *the* makes it more likely that the reader notices the repetition (though errors were still common), and suggested that this may be because separate fixation events make the two encodings more distinct.

On the other hand, under OB1-Reader (Snell, van Leipsig, et al., 2018), a relationship between skipping and failure to notice a transposition is expected: When the second word is recognized before the first, the complete recognition of the second word should lead to a skip of this word, and should make the mis-encoding of word order more likely. However, although we saw fewer fixating-both-words trials among AT trials than among DT trials, they still constituted a substantial proportion of AT trials (50% and 37% respectively in the two experiments). It will be informative to see whether such a high proportion of direct fixation trials can be formally modeled by OB1-Reader under the proposal that transposed-word effect arises because the second transposed word is recognized earlier than the first, and the assumption that once a word is recognized it does not receive a fixation. In sum, while the data are not consistent with the strongest prediction of *E*-Z Reader—no relationship between word skipping and failure to notice transpositions—they also appear to be inconsistent with the strongest prediction of OB1-Reader—failure to notice a transposition *only* when there is a word skip.

8.2. (Non)incremental integration and rational inference

While the transposed-word effect is a clear demonstration that the reader's sentence-level representation is not always faithful to the bottom-up input, it is also clear that any top-down influence that alters the representation of word order (i.e., inference) must operate rapidly enough that it prevents an initial error signal from disrupting the forward progress of the eyes, given that eye movements in AT trials were largely indistinguishable from AG trials and discrepant from DT trials. In addition, had integration of words been perfectly serial and incremental, in sequences like *The boy on sat*, before processing the fourth word *sat*, *on* would have been integrated as a preposition modifying *the boy*. Processing of *sat* should then have elicited disruption on the first fixation due to integration difficulty, whether or not it was later fixed by rational inference. We tentatively suggest the following account, which preserves the assumption that words are recognized serially (Reichle, Pollatsek, Fisher, & Rayner, 1998), but relaxes E-Z Reader's assumption that *integration* of each word into its context is entirely serial and incremental. The relaxation on perfect incrementality allows rational inference to prevent the literal, ungrammatical input from ever being perceived as an error.

The current version of E-Z Reader (Reichle, Liversedge, et al., 2009) posits that syntactic and semantic integration of word n into its context must be complete before integration of word $n+1$ can begin. However, we know of no evidence that requires this very strong assumption; while there is clear evidence that the syntactic or semantic fit of a word into its context often influences early eye movements on this word (e.g., Rayner et al., 2004; Staub, 2011), this does not establish that such integration *always* takes place in such an incremental fashion. Indeed, the sentence processing literature contains various hints that integration may not be perfectly incremental, at least in some circumstances; for example Frazier and Rayner (1987) suggested that the syntactic ambiguity in a string such as *desert trains* (where *desert* can be either a head noun or a modifier, and *trains* can be either a verb or a plural noun) is not resolved until subsequent disambiguating context is encountered. We propose that integration of word n may not always be complete when integration of word $n+1$ begins, especially if recognition of word $n+1$ follows rapidly after recognition of word n. In this case, the sentence processing mechanism may entertain multiple possible orderings of the two words, and a top-down influence of linguistic knowledge may result in a representation in which word $n+1$ precedes word n, especially when the veridical order – word n before $n+1$ – is ungrammatical. This is in essence a noisy-channel inference account (Gibson et al., 2013). The ungrammatical order will not induce processing difficulty, because though word n is recognized before word $n+1$, they are never represented/integrated as occurring in that order. Such an approach acknowledges the role of top-down constraints via inference, but revises the EZ-Reader model by assuming, first, that word order representation is not dictated by the order of serial recognition, but rather by the result of the integration process; and second, that integration is not perfectly incremental on a word-by-word basis (cf. Christiansen & Chater, 2016), resulting in a situation in which two words are sometimes integrated at the same time into a plausible representation.

8.3. Serial or parallel processing

As noted in the Introduction, the transposed word effect is of theoretical interest partially because it has been argued to support parallel lexical processing. Given the other evidence against parallel lexical processing listed in the Introduction, however, we believe that it is appropriate to pursue other explanations of this effect. This study set out to test specific predictions from the OB1-Reader model, a model of reading that assumes parallel word processing (Snell, Declerck, & Grainger, 2018). As summarized above, one prediction was born out in the current study: Transposed words with different lengths were more easily noticed. One the other hand, two findings failed to support the predictions of the model: transposed words of saliently different lengths still escaped detection more frequently than controls, and having a closed-class word at the second position of the transposed sequence did not make it harder to detect the transposition. Future studies that carefully manipulate length, class, and frequency, without confounding, are needed to further test these empirical predictions.

The subject-level correlations between detection rates of transposition and incomplete sentence errors (Exp 1) and between detection rates of transposition errors and reading speed on grammatical filler sentences (Exp 2) are also of potential relevance to the question of whether parallel processing is responsible for failure to notice transposition errors. The first correlation suggests that the substantial subject-level variance in detection rates for transposition errors might be due in part to individual differences in response bias: Some subjects are simply more inclined than others to report that a sentence is ungrammatical or unacceptable (see Hammerly, Staub, & Dillon, 2019, for discussion of the response bias in the context of grammaticality judgment tasks). The second correlation might reflect individual differences in readers' noise model. That is, readers who read fast throughout the experiment may be aware – at some level – that their perception of the sentences might have been rather unreliable, and thus the nonliteral interpretation would be considered more likely via inference. On the other hand, it is doubtful that the extreme subject variability in both experiments can be attributed to different tendencies to engage in parallel word processing among readers.

In light of the lack of full support for the predictions from the most relevant parallel processing model, we have proposed that a serial lexical processing model may account for the present data by relaxing the assumption that integration of each word into its sentence context is entirely serial and incorporating a new assumption that integration involves rational inference. Admittedly, this idea needs substantial further development. In particular, what is required is a detailed theory of the circumstances under which integrations of word n and $n+1$ may take place concurrently, giving rise to the possibility of representing word $n+1$ as preceding word n.

We note, however, some initial support for the hypothesis that readers may represent the order of two words incorrectly, even if they are encoded in the right order. Recently, we have conducted an experiment (Huang and Staub, in prep) in which materials similar to those in Experiment 1 were presented using rapid serial visual presentation (RSVP) at 250 ms per word. In this paradigm, words are presented serially, so that parallel processing is entirely impossible. Nevertheless, we have found that subjects fail to notice transpositions of successive words nearly 20% of the time.

9. Conclusion

This study examined readers' tendency to overlook transposition errors. In two eye movement experiments, participants' likelihood of failing to notice a transposition was comparable to the previous study of this phenomenon (Mirault et al., 2018). Further examination of the effects of word class and word length, however, failed to fully support the parallel processing account proposed by Mirault et al., although future experiments that carefully control frequency are needed to evaluate the model. With eye-tracking, we also found that readers were more likely to detect the transposition when they directly fixated both transposed words. Finally, comparing eye movements on the transposed words on trials where readers noticed the error and did not notice the error, along with eye movements on grammatical sentences, we did not find evidence of disruption in trials when participants read a transposed sentence but failed to report the error. This argues against a late rational inference account according to which the misjudgment is due to corrective inference occurring *after* the literal input is *perceived as an error*. In general, results in Experiment 2 paralleled the results in Staub et al. (2019), suggesting shared mechanisms underlying failure to notice a transposition error and a repetition error in sentence reading. We tentatively propose an account that assumes serial word recognition, but allows post-lexical integration to be delayed and to take into account alternative possible orderings, such that word $n+1$ is ultimately assigned to an earlier position even when word n is recognized first. Future studies that directly test such a mechanism will be needed.

Acknowledgment

The research was supported by a grant from the National Science Foundation BCS 1732008 to Adrian Staub. We thank Eunji Choi, Sabrina Elbach, Serene Elbach, Alex Green, Subin Jung, Jisun Kim, Hira Patel, and Leo Stern for participant recruitment and data collection. Thanks to audiences at the 2019 Psychonomic Society Conference and 2020 virtual CUNY conference on Human Sentence Processing for helpful feedback. Materials, data, and analysis scripts are available at: https://osf.io/vhguf/.

Credit author statement

Kuan-Jung Huang: Conceptualization, Methodology, Formal Analysis, Investigation, Writing-Original Draft, Visualization Adrian Staub: Conceptualization, Methodology, Resources, Data Curation, Writing-Review & Editing, Supervision, Project administration, Funding Acquisition.

Appendix A. Supplementary Bayesian analysis with informed priors

For the supplementary analyses, we implement more informative priors for both contrasts, for each transposed word.[6] The model structures were the same as the analyses reported in the main text. For Word 4—the point at which the ungrammaticality surfaces—the prior for AG-AT contrast was a Gaussian with a mean effect of −20 and − 40 ms for FFD and GD and a standard deviation of 10 and 20 ms respectively, and the AT-DT contrast was a Gaussian with a mean effect of 0 ms, with the same SDs. This reflects the hypothesis of a late post-perceptual rational inference account whereby the transposition error is first perceived at the earliest stages but corrected only at the later stages (Gibson et al., 2013). For GP and TVT, the priors for the AG-AT contrast remained the same as that of GD, reflecting the idea that rational inference kicks in at this stage and the transposition is "corrected", inducing no further increase in reading times. The difference between DT and AT starts to expand from this stage; the prior for GP had a mean of 80 ms and sd of 40 ms, and the prior for TVT had a mean of 160 ms and sd of 80 ms. For Word 3, the pre-ungrammaticality point, for both contrasts, the priors were a Gaussian with a mean effect of 0 ms, as the literature has mostly concluded that there is no high-level parafoveal-on-foveal effect (e.g., Brothers et al., 2017), and a standard deviation of 10, 20, and 40 ms for FFD, GD, and GP, respectively. For TVT, we assumed that inference has taken place by this late stage and therefore no difference between AG and AT would appear, but there would be a difference between DT and AT with a mean of 80 ms and standard deviation of 40 ms (this is half the size of the effect on Word 4 since it is the pre-ungrammatical region).

Results are shown in Tables A1 and A2. There was no hint of processing difference at Word 3 in any first-pass reading measure, but TVT was strongly disrupted on trials when participants answered correctly to the detection question. At Word 4, the patterns were less clear than the results with a uniform prior. First-fixation durations on AT trials were estimated to be longer than those on AG trials. First-fixation durations on DT trials were also longer than AT, although the 95% credible interval very slightly crossed the zero point. The difference between AG and AT and between DT and AT again showed a graded pattern on GD, although this time the difference between the latter trial types was highly reliable and the difference between the former was only marginal. On late measures, GP and TVT, the graded pattern remained, with both contrasts showing reliable differences. Notably, the magnitudes of the differences were substantially different, suggesting a clear correction by rational inference by these stages.

Table A1
Response type effects on eye-movements on Word 3 in Exp 2 with informed priors presuming no parafoveal on foveal effects.

Effect	Estimate	Error	l-95%CI	u-95%CI	Rhat	Bulk_ESS	Tail_ESS
FFD							
Intercept	219.99	6.08	208.11	231.79	1.00	11,230	10,831
AG-AT	3.81	5.50	−6.92	14.57	1.00	15,811	12,185
DT-AT	6.83	5.97	−4.90	18.57	1.00	12,771	11,911
Gaze							
Intercept	237.35	8.80	220.13	254.47	1.00	17,903	12,099
AG-AT	6.03	8.31	−10.17	22.33	1.00	19,934	13,126
DT-AT	14.05	9.15	−3.80	32.03	1.00	16,652	11,711
Go-past							
Intercept	356.09	18.65	319.43	392.63	1.00	12,632	11,872
AG-AT	−29.05	17.73	−63.77	5.43	1.00	16,068	11,941
DT-AT	32.80	20.65	−7.73	73.50	1.00	11,236	9981
Total Viewing							
Intercept	385.53	18.67	348.57	422.38	1.00	18,599	12,194
AG-AT	−34.10	16.47	−65.79	−1.83	1.00	30,788	11,153
DT-AT	148.19	21.65	106.00	190.72	1.00	15,191	12,978

Table A2
Response type effects on eye-movements on Word 4 in Exp 2 with informed priors presuming early interruption and later correction by rational inference.

Effect	Estimate	Error	l-95%CI	u-95%CI	Rhat	Bulk_ESS	Tail_ESS
FFD							
Intercept	233.54	6.59	220.70	246.57	1.00	11,555	11,508
AG-AT	−13.05	5.69	−24.12	−1.86	1.00	17,757	12,303
DT-AT	11.69	6.24	−0.53	23.94	1.00	13,875	11,986
Gaze							
Intercept	253.41	9.40	235.06	271.90	1.00	11,620	11,518

(*continued on next page*)

[6] We report results in Experiment 2 only; for similar results in Experiment 1 see OSF.

Table A2 (*continued*)

Effect	Estimate	Error	l-95%CI	u-95%CI	Rhat	Bulk_ESS	Tail_ESS
AG-AT	−16.26	8.63	−33.24	0.64	1.00	15,927	12,065
DT-AT	24.58	10.48	4.17	45.03	1.00	10,625	10,766
Go-past							
Intercept	385.36	19.82	346.36	423.98	1.00	16,498	12,841
AG-AT	−37.02	15.65	−67.61	−6.69	1.00	30,540	11,398
DT-AT	129.33	21.30	87.41	170.82	1.00	20,057	12,679
Total Viewing							
Intercept	381.04	18.02	345.54	416.04	1.00	18,723	12,966
AG-AT	−28.19	14.18	−55.91	−0.71	1.00	29,371	12,447
DT-AT	137.01	20.64	96.74	176.82	1.00	13,425	12,373

References

Andrews, S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. *Journal of Memory and Language, 35*(6), 775–800.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint*. arXiv:1506.04967.

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.

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.

Bürkner, P. C. (2017). brms: An R Package for Bayesian Multilevel Models using Stan. *Journal of Statistical Software, 80*(1), 1–28.

Choi, W., & Gordon, P. C. (2014). Word skipping during sentence reading: Effects of lexicality on parafoveal processing. *Attention, Perception, & Psychophysics, 76*(1), 201–213.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences, 39*.

Core Team, R. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review, 117*(3), 713.

Depaoli, S., Winter, S. D., & Visser, M. (2020). The importance of prior sensitivity analysis in Bayesian statistics: Demonstrations using an interactive shiny app. *Frontiers in Psychology, 11*.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*(4), 777.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*(2), 178–210.

Frazier, L., & Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language, 26*(5), 505–526.

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.

Gordon, P. C., Plummer, P., & Choi, W. (2013). See before you jump: Full recognition of parafoveal words precedes skips during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 633.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th ed.). New Jersey: Prentice Hall Upper Saddle River.

Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive Psychology, 110*, 70–104.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40*(6), 431–439.

Levy, R. (2008). A noisy-channel model of human sentence comprehension under uncertain input. *Proceedings of the 2008 conference on empirical methods in natural language processing*, 234–243.

Liu, Z., Li, Y., Paterson, K. B., & Wang, J. (2020). A transposed-word effect in Chinese reading. *Attention, Perception, & Psychophysics*, 1–7.

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.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113*(2), 327.

Pegado, F., & Grainger, J. (2019a). *A transposed-word effect in same-different judgments to sequences of words*. Journal of Experimental Psychology: Learning, Memory, and Cognition.

Pegado, F., & Grainger, J. (2019b). Dissociating lexical and sublexical contributions to transposed-word effects. *Acta Psychologica, 201*, 102943.

Perea, M., & Lupker, S. J. (2003). Does jugde activate COURT? Transposed-letter similarity effects in masked associative priming. *Memory & Cognition, 31*(6), 829–841.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Rayner, K., Angele, B., Schotter, E. R., & Bicknell, K. (2013). On the processing of canonical word order during eye fixations in reading: Do readers process transposed word previews? *Visual Cognition, 21*(3), 353–381.

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(6), 1290.

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

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125.

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.

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.

Ryskin, R., Futrell, R., Kiran, S., & Gibson, E. (2018). Comprehenders model the nature of noise in the environment. *Cognition, 181*, 141–150.

Schotter, E. R., Bicknell, K., Howard, I., Levy, R., & Rayner, K. (2014). Task effects reveal cognitive flexibility responding to frequency and predictability: Evidence from eye movements in reading and proofreading. *Cognition, 131*(1), 1–27.

Snell, J., Declerck, M., & Grainger, J. (2018). Parallel semantic processing in reading revisited: Effects of translation equivalents in bilingual readers. *Language, Cognition and Neuroscience, 33*(5), 563–574.

Snell, J., & Grainger, J. (2019a). Readers are parallel processors. *Trends in Cognitive Sciences, 23*(7) , 537–546.

Snell, J., & Grainger, J. (2019b). Word position coding in reading is noisy. *Psychonomic Bulletin & Review, 26*(2), 609–615.

Snell, J., Meeter, M., & Grainger, J. (2017). Evidence for simultaneous syntactic processing of multiple words during reading. *PLoS One, 12*(3).

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.

Staub, A. (2011). Word recognition and syntactic attachment in reading: Evidence for a staged architecture. *Journal of Experimental Psychology: General, 140*(3), 407.

Staub, A., Dodge, S., & Cohen, A. L. (2019). Failure to detect function word repetitions and omissions in reading: Are eye movements to blame? *Psychonomic Bulletin & Review, 26*(1), 340–346.

Veldre, A., & Andrews, S. (2018). Beyond cloze probability: Parafoveal processing of semantic and syntactic information during reading. *Journal of Memory and Language, 100*, 1–17.

Wen, Y., Mirault, J., & Grainger, J. (2021). *Fast syntax in the brain: Electrophysiological evidence from the rapid parallel visual presentation paradigm (RPVP)*. Journal of Experimental Psychology: Learning, Memory, and Cognition.

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

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.

White, A. L., Palmer, J., Boynton, G. M., & Yeatman, J. D. (2019). Parallel spatial channels converge at a bottleneck in anterior word-selective cortex. *Proceedings of the National Academy of Sciences, 116*(20), 10087–10096.