Limits on failure to notice word transpositions during sentence reading
Kuan-Jung Huang & Adrian Staub (University of Massachusetts Amherst)

Mirault et al. (2018) found that readers sometimes judge a sentence with transposed words to be grammatical (e.g., *The white* ***was cat*** *big*). They attributed these errors to noisy positional information resulting from parallel word processing. Their account predicts a higher error rate when the second transposed word is easier to recognize than the first, because this increases the probability that the second word will be identified before the first (Snell et al., 2018). Here we tested this prediction by manipulating the frequency of each of two transposed words; to avoid confounds with part of speech (as in Huang and Staub, 2020), both words were open-class.

Frequency was factorially manipulated for the first and second transposed word in sentences (Table 1); mean Zipf frequency was around 5 for high-frequency words and 3 for low-frequency words, based on the SUBTLEX Corpus (Brysbaert & New, 2009); the frequency distributions were non-overlapping. We used two sentence frames for each combination of levels of word frequency, one with a noun preceding a verb in the transposed order (e.g., *His sister* ***stuff drew***) and one with a noun preceding an adjective (*A really* ***fellow scary***), rendering 8 sub-conditions. Each subject read 7 transposed sentences and 7 un-transposed grammatical sentences in each of the 8 conditions. We also added a reference condition with an additional 7 transposed and 7 grammatical items, in which the transposition involved a pronoun (e.g., *It might* ***him cure*** *of the deadly disease*); this transposition was among the most frequently missed in our previous experiments. Finally, we also included twice as many grammatical fillers as critical items. Self-reported native English speakers participated on MTurk (N=69). For the critical items, the question to be answered after reading the sentence was an error-detection question, while for the fillers it was a comprehension question. Subjects could not predict which type of question would be asked until each sentence was removed from the screen (Fig. 1). Trials with RT to questions > 15s were discarded (0.4%).

Averaging across the critical conditions, subjects failed to detect transpositions only 9.1% of the time, while rejecting the corresponding grammatical versions 9.3% of the time. Thus, subjects failed to detect the transposition numerically less often than they rejected the grammatical counterparts; there was no transposition effect. In sharp contrast, in the reference condition, where one of the transposed words was a pronoun, they failed to detect transpositions 32% of the time, while rejecting the grammatical version 5.5% of the time (Fig. 2). Despite the apparent lack of a transposition effect in the target items, we assessed the prediction that word frequency should modulate the rate of failure to notice transpositions. We ran GLME models (Bates et al., 2015) testing effects of frequency, frame type, and their interaction (Table 2) on the probability of noticing the transposition. There was a main effect of frame type, with Frame 1 being more illusory, and a marginal effect of frequency, in the direction of less frequent failure to notice the error when the first transposed word was low frequency and the second word was high frequency (i.e., the low-high condition). This is the opposite direction from the prediction of the parallel processing account. This pattern was similar in a post-hoc analysis restricted to items that were highly acceptable in their grammatical version.

To explore the source of the difference in detectability of the transposition between the critical and reference conditions, we correlated (item-wise) error rate with word length, and with bigram frequency of the transposed words in their canonical order. While both factors explained between-condition variation, the latter also explained within-condition variation; the items in the reference condition that tended to elicit the highest rate of failure to notice the transposition were those that had high bigram frequency in the grammatical word order (Fig. 3).

In sum, we barely found a transposition effect with open-class words, while replicating a large effect in an additional reference condition in which one word was closed-class. Any effect of word frequency was in an unpredicted direction, with failure to notice errors being less common when the second word was higher frequency. Post-hoc analyses suggest that bigram frequency of the two critical words, in their grammatical order, may account for much of the item-level variability in the failure to notice transpositions. This finding aligns with a rational inference account emphasizing the role of the reader's prior for the underlying grammatical string (Gibson et al., 2013).

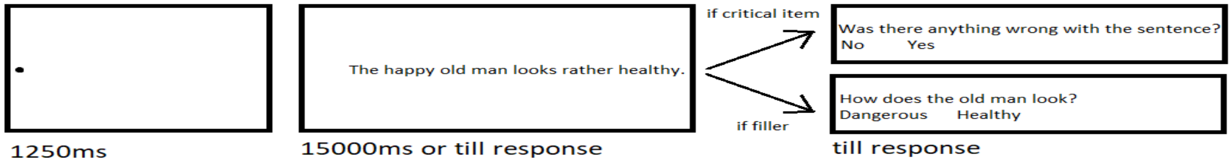

Figure 1. Procedure for each trial.

Freq condition	Frame 1	Frame 2
High-High	His sister **stuff drew** that was not recognizable.	A really **fellow scary** came into the room.
High-Low	The cells **water absorb** through their tiny pores.	A painfully **sound eerie** came from the woods.
Low-High	My nephew **cider stores** in the wine cabinet.	The particularly **jester short** will please the king.
Low-Low	The factories **alloy refine** using very high heat.	An especially **hobbit rugged** went into the cave.

Table 1. Critical items (grammatical not shown). All sentences were 8 words, and transposition always occurred between words 3 and 4. The transposed words were always shorter than 6 letters and with the length difference between them no greater than 1 letter.

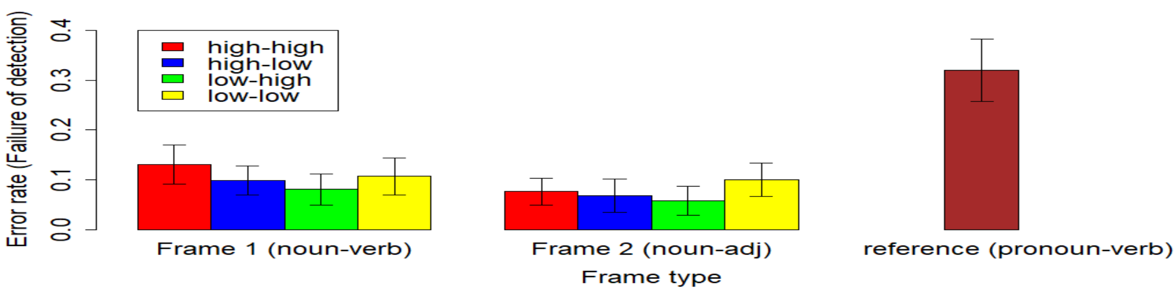

Figure 2. Error rate for transposed sentences, by condition (error bar = by-subject 95% CIs).

	Estimate	SE	Z	P
Intercept	3.32	0.253	13.13	<2e-16
Frame type (sum-coded)	0.45	0.186	2.39	.01
HH – LH (treatment-coded)	-0.52	0.268	-1.94	.05
HL – LH (treatment-coded)	-0.25	0.272	-0.93	.35
LL – LH (treatment-coded)	-0.50	0.27	-1.85	.06

Table 2. Estimation of effects of frame type and frequency on accuracy.

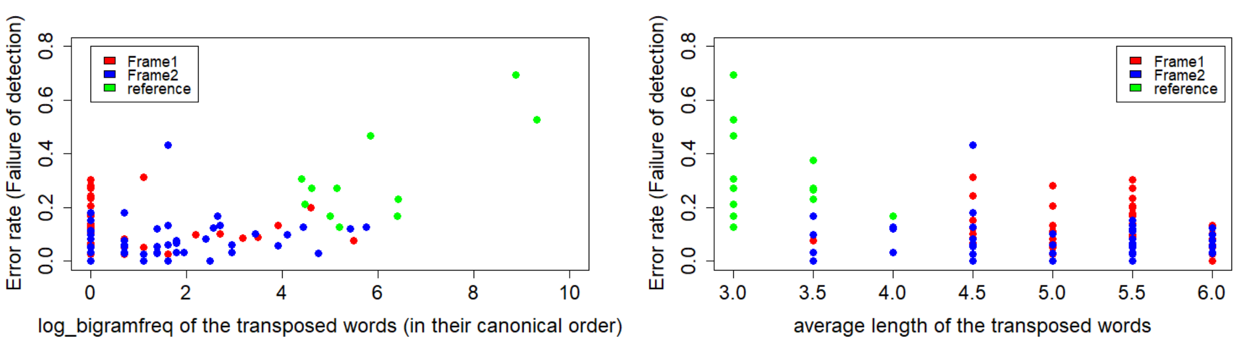

Figure 3. Scatterplots of error rate against bigram frequency (left) and word length (right).

References:

[1] Mirault, J., Snell, J., & Grainger, J. (2018). *PsychScience*. [2] Huang, K-J. & Staub, A. (2020). Poster at CUNY. [3] Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018). *PsychReview*. [4] Brysbaert, M. & New, B. (2009). *Behavioral Research Methods*. [5] Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *arXiv*. [6] Gibson, E., Bergen, L., & Piantadosi, S. (2013). *PNAS*.