Do you want / oloka/ on a /bist k/? On the scope of transposed-phoneme effects with non-adjacent phonemes

Sophie Dufour, Jonathan Mirault, Jonathan Grainger

To cite this version:

Sophie Dufour, Jonathan Mirault, Jonathan Grainger. Do you want / oloka/ on a /bist k/? On the scope of transposed-phoneme effects with non-adjacent phonemes. Psychonomic Bulletin and Review, Psychonomic Society, 2021, 28, pp.1668-1678. 10.3758/s13423-021-01926-9 . hal-03225295

HAL Id: hal-03225295

https://hal.archives-ouvertes.fr/hal-03225295

Submitted on 12 May 2021

Do you want /ʃoloka/ on a /bistɔk/?

On the scope of transposed-phoneme effects with non-adjacent phonemes

Sophie Dufour[1,3] Jonathan Mirault[2,3] and Jonathan Grainger[2,3]

1. Aix-Marseille Université, CNRS, LPL, UMR 7309, 13100 Aix-en-Provence, France

2. Laboratoire de Psychologie Cognitive, Aix-Marseille Université & CNRS

3. Institute for Language, Communication, and the Brain, Aix-Marseille Université, Aix-en-Provence, France

Sophie Dufour
Laboratoire Parole et Langage
Aix-Marseille Université, CNRS
5, avenue Pasteur
13604 Aix-en-Provence
France Email: sophie.dufour@lpl-aix.fr

Abstract: We conducted two lexical decision experiments and one replication study to examine the scope of transposed-phoneme effects when the transposition involves non-adjacent phonemes. The critical stimuli were nonwords derived from a real word (the base-word) either by transposing two phonemes or by substituting the same phonemes with different phonemes. In Experiment 1, the transposed phonemes belonged either to the same syllable (e.g. /bis.**t**ɔ**k**/ for the French base-word /*bis.**k**ɔ**t***/) or to a different syllable (e.g. /ʃo.**l**o.**k**a/ for the French base-word /*ʃo.**ko.la***/) and were located either at the beginning of the speech signal (e.g. /**s**i**b.**kɔt/ for /***bis.**kɔt*/; /**k**o.**ʃ**o.la/ for /***ʃ**o.**k**o.la*/) or at the end (e.g. /bis.**t**ɔ**k**/ for /*bis.**kɔt***/; /ʃo.**l**o.**k**a/ for /*ʃo.**ko.la***/). Experiment 2 compared within-syllable and between-syllable transpositions derived from the same set of bi-syllabic base-words (e.g., /**s**i**b.**kɔt/, /bi**k.s**ɔt/, /bis.**t**ɔ**k**/ for the base-word /*biskɔt*/). In both experiments, we found clear transposed-phoneme effects with longer "no" decisions for transposed-phoneme nonwords compared with the matched substituted-phoneme nonwords. The effect was of similar magnitude when the transposed phonemes occurred in the same syllable and across different syllables. Also, for both the within- and between-syllable transpositions, the size of the transposed-phoneme effect did not vary as a function of the position of the transposition. Overall, our results suggest that phonemes can migrate across their respective positions not only within a syllable, but also across syllables. More importantly, they also suggest that position-independent phonemes exert a continuous influence during the entire processing of the auditory stimulus to the extent that there is sufficient time for this influence to manifest itself.

Key words: spoken word recognition, transposed-phonemes, auditory lexical decision

Is the precise order of phonemes encoded during spoken word recognition? Since 2013, a growing body of research (Dufour & Grainger, 2019, 2020, 2021; Gregg et al., 2019; Toscano et al., 2013) indicates that the answer to that question is more likely to be negative, and thus goes against a strong and widely held assumption that the sounds that make up spoken words are precisely assigned to their correct positions in words. It has been shown, for example, that a speech input like [kat] not only provides support for the corresponding lexical representation *cat* but also for the lexical representation that contains the same phonemes in a different order *tack*. This finding counters the predictions of all current models of spoken word recognition (Gaskell & Marslen-Wilson, 1997; Marslen-Wilson & Warren, 1994; Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1990; McClelland & Elman, 1986; Norris, 1994) except one – the TISK model of Hannagan et al. (2013). The main reason is that TISK does not code for the precise order of phonemes. It replaces the position-dependent units postulated by most of models of spoken word recognition by both a set of position-independent phoneme units and a set of open-diphone units that represent ordered sequences of contiguous and non-contiguous phonemes. Within such a framework, both the position-independent phoneme units and the open-diphone representations trigger activation of words that share all their phonemes with a given target word but not necessarily in the same order[1].

One important prediction of the TISK model (Hannagan et al., 2013) is that a transposed nonword (e.g., /bi**ks**ɔt/) sharing all of the phonemes, but in different positions, with a base-word (*/bi**sk**ɔt/* BISCOTTE “toasted bread”) should activate to a greater degree the lexical representation corresponding to the base-word than does a phonological control non-word created by substituting two phonemes of the base-word at the same positions with different phonemes (/bi**pf**ɔt/). As a result, in TISK, transposed-phoneme nonwords should be perceived as being more similar to their base-words than substituted-phoneme nonwords. In accordance with this prediction, Dufour and Grainger (2021) have recently shown that transposed-phoneme nonword primes (/bi**ks**ɔt/) are more effective in facilitating the subsequent processing of the corresponding base-word target (/*bi**sk**ɔt*/) than substituted-phoneme nonword primes (/bi**pf**ɔt/). Moreover, using an unprimed lexical decision task, in the same study it was shown that transposed-phoneme nonwords (/bi**ks**ɔt/) took longer to classify as nonwords compared with substituted-phoneme nonwords (/bi**pf**ɔt/). In both experiments, the transposed-phoneme effect occurred when the critical transposed phonemes were adjacent (/bi**ks**ɔt/ derived from /*bi**sk**ɔt*/) but not when the transposition involved non-adjacent phonemes (e.g., /ʃo**l**o**k**a/ derived from /*ʃo**k**o**l**a*/ CHOCOLAT “chocolate”).

The present work was motivated by the null effect of non-adjacent phoneme transpositions in multisyllabic words reported by Dufour and Grainger (2021), a finding that stands in contradiction to the robust non-adjacent transposed-phoneme effects found with monosyllabic words in previous studies (e.g., BUS-SUB in Toscano et al., 2013 and in Gregg et al., 2019; and ROBE /ʀɔb/ "dress" - BORD /bɔʀ/ "side" in Dufour & Grainger, 2019, 2020). The most obvious explanation for this discrepancy is that the non-adjacent transposed phonemes belonged to different syllables (/ʃo.**l**o.**k**a/ for /*ʃo.**k**o.**l**a*/) in the Dufour and Grainger (2021) study, and this was not the case in prior investigations which were limited to monosyllabic words, and for which transposed phonemes inevitably belonged to the same syllable. Dufour and Grainger (2021) therefore pointed to a possible role for syllable boundaries in modulating transposed-phoneme effects. According to this account, consonants could migrate across their respective positions within a syllable, but not across syllables and thus transposed-phoneme effects would only occur when the transposed phonemes belong to the same syllable.

In the present study, we provide a more in-depth examination of transposed-phoneme effects when the transposition involves non-adjacent phonemes, with the primary aim being to test the role played by syllable boundaries. In Experiment 1, the transposed phonemes could either belong to the same syllable (the within-syllable manipulation) or to two different syllables (the across-syllable manipulation). If the syllable boundary account of the discrepancy between the Dufour and Grainger (2021) findings and findings prior to that is correct, then we expected to observe a greater transposed-phoneme effect when the transposed phonemes belonged to the same syllable than when they crossed a syllable boundary. We also manipulated the position of the transposed phonemes in the speech signal, such that they could be located either at the beginning or the end of the target nonword. This manipulation was included as part of our aim to explore the scope of non-adjacent transposed-phoneme effects.

Experiment 1

In Experiment 1, we manipulated both syllable boundary and phoneme position in a factorial design. Transposed phonemes belonged either to the same syllable (e.g., /bis.**t**ɔ**k**/ for /*bis.**k**ɔ**t***/) or to a different syllable (e.g., /ʃo.**l**o.**k**a/ for /*ʃo.**ko.la***/) and were located either at the beginning (e.g., /**s**i**b.**kɔt/ for /***bis.**kɔt*/; /**k**o.**ʃ**o.la/ for /***ʃ**o.**ko**.la*/) or at the end (e.g., /bis.**t**ɔ**k**/ for /*bis.**k**ɔ**t***/; /ʃo.**l**o.**k**a/ for /*ʃo.**ko.la***/) of the auditory input. In all cases the transposed phonemes were consonants.

Method

Participants: A total of 95 participants were recruited on-line for the experiment. All participants indicated that French was their native language. All were right-handed and aged 18- 49 years. They were paid for their participation. This sample size was determined a priori without formal power calculations. Since this was an on-line experiment, we aimed for a relatively large sample size in order to have enough participants after rejecting those with excessive error rates and/or RTs.

Materials: Fifty-two base-words with a CVC.CVC syllable structure were selected and were used in the same-syllable condition. 52 base-words with a CV.CV.CV syllable structure were also selected and were used in the different-syllable condition. The mean frequency of the base-words (from Vocolex: Dufour et al., 2002) was 17 and 14 occurrences per million in the same- and different- syllable conditions, respectively. From each of the 104 base-words, four nonwords were created: two by transposing two consonants and the two others by substituting two consonants with different consonants. For the transposed-phoneme nonwords, in the "initial" condition these were created by transposing the two initial consonants of the base-words (e.g., /**s**i**b.**kɔt/ for /***bis.****kɔt*/ BISCOTTE "toasted bread" in the same-syllable condition; /**k**o.**ʃ**o.la/ for /***ʃ**o.**k**o.la*/ CHOCOLAT "chocolate" in the different-syllable condition), and in the "final" condition by transposing the two final consonants (e.g., /bis.**t**ɔ**k**/ for /*bis.**k**ɔ**t***/ in the same-syllable condition; /ʃo.**l**o.**k**a/ for /*ʃo.**k**o.**l**a*/ in the different-syllable condition). For the substituted-phoneme nonwords, the same positions as the transposed-phonemes were respected, thus replacing the two initial consonants of the base-words in the initial condition (e.g., /**ʃip.**kɔt/ for /***bis.**kɔt*/ in the same-syllable condition; /**p**o.**f**o.la/ for /*ʃo.**ko.**la*/ in the different-syllable condition), and in the final condition by replacing the two final consonants (e.g., /bis.**p**ɔ**g**/ for /*bis.**k**ɔ**t***/ in the same-syllable condition; /ʃo.**r**o.**p**a/ for /*ʃo.**ko.la***/ in the different-syllable condition). In all cases, the substituted phonemes were phonetically similar to the transposed phonemes and shared three out of the four phonetic features generally used in French phonology (e.g., place, voice, manner, and nasality for consonants). This led to a 2 X 2 X 2 factorial design with Nonword Type (transposed vs. substituted), Syllable (same vs. different) and Position (initial vs. final) as factors.

The four categories of nonwords (substituted-initial, transposed-initial, substituted-final, transposed-final) derived from each of the 104 base-words (52 in the same syllable condition and 52 in the different syllable condition) were rotated across four experimental lists, and participants only saw one nonword from each quadruplet. Each list was thus composed of 104 nonwords, 52 in the same-syllable condition and 52 in the different-syllable condition. Within each syllable condition, there were 13 nonwords per condition (4 conditions from the Nonword Type X Position manipulation). For the purpose of the lexical decision task, 104 words of two and three syllables were added to each list. All of the stimuli were recorded by a female native speaker of French, in a sound attenuated room, and digitized at a sampling rate of 44 kHz with 16-bit analog to digital recording. The main characteristics of the critical nonword stimuli are given in Table 1. The complete set of nonwords is available at https://osf.io/gwk2j/ (Open Science Framework; Foster & Deardorff, 2017).

<Insert Table 1 about here>

Procedure:

The experiment was programmed using html and php protocols and then uploaded onto the lab server. We recruited participants via the Prolific platform (www.prolific.co). Instructions were presented in French on the first screen. Participants were asked to make a lexical decision as quickly and accurately by pressing the “m” key for the nonword response and the “q” key for the word response. A countdown of 3 seconds was displayed before the practice block (N = 8), followed by the experimental block (N = 208). Response times (RTs) were recorded from the onset of stimuli. An intertrial interval of 2,000 ms elapsed between the end of one trial and the beginning of the next. At the end of the experiment, a csv file was saved on the server.

Results

Seven participants were excluded from the analyses. One participant had an error rate of 46% and the six others had excessively long RTs (greater than 1800 ms on average and greater than 2.5 standard deviations above the general mean). One final-position transposed nonword of the same syllable condition that gave rise to an error rate of more than 70% was also removed. The mean RT and percentage of correct responses to nonwords in each condition are presented in Figure 1.

<Insert Figure 1 about here>

RT analysis:

RTs to nonwords (available at https://osf.io/gwk2j/; Open Science Framework; Foster & Deardorff, 2017) were analyzed using linear mixed effects models with participants and items as crossed random factors, using R software (R Development Core Team, 2016) and the lme4 package (Baayen et al., 2008; Bates and Sarkar, 2007). The RT analysis was performed on correct responses, thus removing 267 data points out of 9131 (2.92%). Three RTs <400 ms and 77 RTs >4000 ms were considered as outliers (less than 1%) and were also excluded from the analysis. Additionally, within each condition and for each participant RTs lying more than 3 standard deviations from the participant's mean were excluded (42 data points; less than 1%). For the model to meet the assumptions of normally-distributed residuals and homogeneity of variance, a log transformation was applied to the RTs (Baayen & Milin, 2010) prior to running the model. The model was run on 8742 data points. We tested a model with the variables Nonword Type (transposed, substituted), Syllable (same, different), Position (initial, final) and their interaction entered as fixed effects. The nonword deviation point, that is, the phoneme position at which the nonword is no longer compatible with any word, was also entered as a continuous variable. The model failed to converge when random participant slopes for the within-participant factors Nonword type, Syllable and Position were included (see Barr et al., 2013). Therefore the final model included only random intercepts for participants and items. We applied orthogonal contrast coding for the independent variables, namely 0.5 for one condition and -0.5 for the other condition, which allows an estimation of main effects. The continuous predictor (deviation point) was centered and scaled. The *b* values, standard errors (*SE*s) and *t* values (RTs) or *z* values (Errors) are reported, with *t* and *z* values beyond $|1.96|$ deemed significant (Baayen, 2008).

The model revealed no effect of Position (*b*=0.0055, *SE*=0.0073, *t*=0.76). A significant effect of the Syllable was observed (*b*=0.0243, *SE*=0.0071, *t*=3.42) with RTs being slower in the same syllable than in the different syllable condition. Crucially, a clear transposed-phoneme effect was observed with RTs being slower for the transposed-phoneme nonwords than for the substituted-phoneme nonwords (*b*=0.0199, *SE*=0.0071, *t*=2.82). This transposed-phoneme effect interacted neither with the Syllable factor (*b*=-0.0120, *SE*=0.0141, *t*=-0.85) nor with the Position factor (*b*=0.0098, *SE*=0.0141, *t*=0.69). Also, the Syllable X Position interaction (*b*=-0.0018, *SE*=0.0144, *t*=-0.13) and the three-way interaction Nonword Type X Syllable X Position were not significant (*b*=-0.0203, *SE*=0.0282, *t*=-0.72).

Accuracy analysis:

The percentage of correct responses was analyzed using a mixed-effects logit model (Jaeger, 2008) following the same procedure as for RTs. A significant effect of Nonword Type was found with more correct responses in the substituted than in the transposed condition (*b*=-0.4909, *SE*=0.1874, *z*=-2.62). The effects of Syllable and Position were also significant with more correct responses in the same than in the different syllable condition (*b*=0.7635, *SE*=0.1886, *z*=4.05), and more correct responses in the initial than in the final position condition (*b*=0.3965, *SE*=0.1905, *z*=2.08). No interactions were significant.

Discussion

Experiment 1 manipulated the position of transposed phonemes which could either be within the same syllable or belong to two different syllables and could either occur at the beginning or the end of the speech signal. Neither of these factors was found to modulate the main effect of transposed-phonemes. That is, it took longer to respond to transposed-phoneme nonwords than substituted-phoneme nonwords, and this effect was statistically equivalent independently of the position of the manipulated phonemes. Crucially, the effect was robust for both within- and across-syllable transpositions, hence allowing us to reject one account of the discrepancy between the findings of Dufour and Grainger (2021) and earlier work investigating transposed-phoneme effects. It therefore remains to be explained why Dufour and Grainger (2021) failed to observe a transposed-phoneme effect for non-adjacent phonemes. This was the goal of the following post-hoc analysis.

<u>Post-hoc analysis of Experiment 1</u>:

One possible explanation is related to the speed of participants' responses. In Dufour and Grainger (2021), the mean RT in the */ʃo**l**o**k**a/* condition of the unprimed lexical decision experiment was 1021 ms, while the mean RT in the same condition was more than 200 ms longer in the present study. We therefore performed post-hoc analyses in order to examine whether the magnitude of the */ʃo**l**o**k**a/* effect, that is, the difference between the transposed and substituted nonwords in the different syllable final position condition, correlates with the average RT of participants in that condition. As illustrated in Figure A1 of Appendix A, there is a positive correlation (*r*(86)=.32, *p*<.01), indicating that the size of the transposed-phoneme effect increased as RTs increased. Importantly, the same correlation concerning word-initial transpositions was not significant (*r* (86)= .11, *p*>.20).

Replication of Experiment 1 with CV.CV.CV type base-words:

Given the importance of this post-hoc analysis for understanding the contradictory patterns across Experiment 1 and Dufour and Grainger's (2021) findings, we decided to replicate the results obtained with the tri-syllabic stimuli in Experiment 1 with a larger sample size for the purpose of the planned correlation analysis. The results of this replication are reported in Appendix B. The replication was successful and, crucially, using a sample of participants almost twice that of Experiment 1, we again found that the size of transposed-phoneme effects with CV.CV.CV type base-words when the transposition occurs at the end of the auditory input depends on the speed of participants' responses, with effects becoming smaller as responses become faster (r(163)=.30; *p*<.001; see Figure B2 in Appendix B). Once again, it is important to note that the same correlation concerning word-initial transpositions was not significant (r(163)=.09 *p*>.20). Furthermore, a split-half analysis of fast and slow participants revealed that in the group of slow participants the position of the transposition did not interact with transposed-priming effects (*b*=0.0049, *SE*=0.0181, *t*=0.27), whereas in the fast participants the interaction was significant (*b*=-0.0410, *SE*=0.0179, *t*=-2.29) with fast participants showing an effect only in the word-initial transpositions (*b*=0.0393, *SE*=0.0132, *t*=2.98). The results of the replication experiment therefore confirm the results of the post-hoc analysis of Experiment 1 and point to differences in average speed of responding as a key factor determining the size of non-adjacent transposed-phoneme effects when these occur at the end of the speech signal.

Experiment 2

In Experiment 1, the influence of syllable boundaries on the transposed-phoneme effect was tested with stimuli that varied in both syllabic structure and number of syllables (i.e. CVC.CVC stimuli for the within-syllable condition and CV.CV.CV for the between-syllable condition). Experiment 2 was therefore designed to re-examine the potential role played by syllable boundaries in driving transposed-phoneme effects while using stimuli matched for number of syllables and syllable structure. To do so, we re-used the CVC.CVC non-words from Experiment 1 formed by a transposition of phonemes both within the first syllable (i.e., initial transposition condition /**sib.**kɔt/ for /***bis.**kɔt*/) and within the second syllable (i.e., final transposition condition /bis.**t**ɔ**k**/ for /*bis.**k**ɔ**t***/). A between-syllable condition was added that consisted of nonwords formed by the transposition of the two medial consonants of the CVC.CVC base-words of Experiment 1 (i.e., medial transposition condition /bi**k.s**ɔt/ for /*bi**s.k**ɔt*/). In this way we could compare effects of within-syllable and between-syllable transpositions derived from the same base-words.[2]

Method

<u>Participants</u>: 95 French-native participants were recruited on-line for the experiment. All were right-handed and aged 18-49 years. They were paid for their participation. The sample size was determined following the same criteria as in Experiment 1.

Materials: Among the 52 CVC.CVC base-words of Experiment 1, 48 were re-used in this experiment, with 4 being removed because the transposition of the two medial consonants led to another word. The mean frequency of the 48 words was 18 occurrences per million. The 48 corresponding substituted and transposed non-words of the initial and final transposition conditions of Experiment 1 were re-used. From each of these 48 base-words, 96 additional non-words were created and were used in the medial transposition condition. 48 consisted of transposed nonwords and were created by transposing the medial consonants (/bi**k.s**ɔt/ for /*bis.**k**ɔt*/). The 48 other consisted of substituted nonwords and were created by replacing the medial consonants (/bi**p.f**ɔt/ for /*bis.**k**ɔt*/). As in Experiment 1, the substituted phonemes were phonetically similar to the transposed phonemes and shared three out of the four phonetic features generally used in French phonology.

The six categories of nonwords (substituted-initial, transposed-initial, substituted-medial, transposed-medial, substituted-final, transposed-final) derived from each of the 48 base-words were rotated across six experimental lists, and participants only saw one nonword from each sextuplet. Each list was thus composed of 48 nonwords and there were 8 nonwords per condition. For the purpose of the lexical decision task, 48 words of two syllables were added to each list. All of the stimuli were recorded by a female native speaker of French, in a sound attenuated room, and digitized at a sampling rate of 44 kHz with 16-bit analog to digital recording. The main characteristics of the critical nonword stimuli are given in Table 2.

<Insert Table 2 about here>

<u>Procedure</u>: This was the same as in Experiment 1.

Results

Nine participants were excluded from the analyses. Among them 6 participants had an error rate superior to 50% and the 3 others had excessively long RTs (greater than 2000 ms on average and greater than 3 standard deviations above the general mean). Moreover, six nonwords that gave rise to an error rate of more than 60% were also removed. The mean RT and percentage of correct responses to nonwords in each condition are presented in Figure 2.

<Insert Figure 2 about here>

<u>RT analysis</u>:

As in Experiment 1, RTs to nonwords (available at https://osf.io/gwk2j/) were analyzed using linear mixed effects models following the same procedure as in Experiment 1. The RT analysis was performed on correct responses, thus removing 128 data points out of 4040 (3.17%). Two RTs <400 ms and 31 RTs >4000 ms were considered as outliers (less than 1%) and were also excluded from the analysis. For the model to meet the assumptions of normally-distributed residuals and homogeneity of variance, a log transformation was applied to the RTs (Baayen & Milin, 2010) prior to running the model. The model was run on 3879 data points and included the variable Nonword Type (transposed, substituted), Position (initial, medial, final) and their interaction as fixed effects. The nonword deviation point was also entered as a covariable. As in Experiment1, the model included only random intercepts for participants and items and we applied orthogonal contrast coding for the independent variables.

A clear main effect of Nonword Type was observed, with RTs being slower for the transposed-phoneme nonwords than for the substituted-phoneme nonwords (b =0.0262, SE = 0.0084, t = 3.12). To examine how the size of the transposed-phoneme effect varied as a function of syllable boundary (initial-within / final-within vs. medial-between) and within-syllable position (initial vs. final), separate 2 × 2 analyses were run with the different combinations of the Position factor (initial, medial, final) and Type of Nonword. They revealed longer RTs in the medial between-syllable condition in comparison to both the initial within-syllable transposition condition (b =0.0392, SE = 0.0104, t = 3.77) and the final within-syllable transposition condition (b =0.0285, SE = 0.0108, t = 2.65). Crucially, no difference in the magnitude of the transposed-phoneme effect was found between the between-syllable and the initial within-syllable conditions (b =0.0125, SE = 0.0208, t = 0.60), and between the between-syllable and the final within-syllable conditions (b =0.0008, SE = 0.0188, t = 0.04). Moreover, no difference in the magnitude of the transposed-phoneme effect was found between the two positions (initial, final) of the within-syllable condition (b =0.0115, SE = 0.0202, t = 0.57).

Accuracy analysis:

The percentage of correct responses was analyzed using a mixed-effects logit model (Jaeger, 2008) following the same procedure as for RTs. No significant effects were found.

Discussion

To sum up, using stimuli matched in terms of syllable structure and number of syllables, Experiment 2 replicated our finding that the transposed-phoneme effect is independent of whether or not the transposition occurs within or across a syllable boundary. These results suggest that non-words such as (/bi**k.s**ɔt/), created by transposing medial consonants belonging to different syllables, provide the same bottom-up support for the lexical representations corresponding to their base-words (/bis.kɔt/) as do nonwords created by transposing phonemes belonging to the same syllable (/**s**i**b.**kɔt/, /bis.**t**ɔ**k**/). Moreover as already observed in Experiment 1, Experiment 2 showed that when the transposed phonemes belong to the same syllable, the size of the transposed-phoneme effect was not affected by the position at which the transposition occurs, with effects of similar magnitude obtained for transpositions in the first and the second syllable.

General Discussion

A growing number of studies have shown that spoken word recognition is relatively insensitive to phoneme transpositions compared with phoneme substitutions. For example, it has been observed that a nonword such as /baksɛt/ is more readily confused with its base-word /baskɛt/ than a nonword such as /bapfsɛt/. In the continuing investigation of such transposed-phoneme effects, one key remaining issue concerns potential differences between adjacent and non-adjacent phoneme transpositions. In this respect, the present study was motivated by the conflicting results obtained by Dufour and Grainger (2021), who found effects for adjacent transpositions but failed to find effects for non-adjacent transpositions, a finding that contrasted with the results of several earlier studies demonstrating effects of non-adjacent transpositions (Toscano et al., 2013; Gregg et al., 2019; Dufour & Grainger, 2019, 2020). One obvious explanation for these contrasting findings was that Dufour and Grainger (2021) examined non-adjacent transposed-phoneme effects in bi-syllabic stimuli with the transpositions crossing the syllable boundary. On the other hand, all prior investigations had used monosyllabic words, and therefore the transposition involved phonemes belonging to the same syllable. Experiment 1 therefore examined non-adjacent transposed-phoneme effects in bi-syllabic stimuli where the transposition could occur within a syllable or across syllables. The results revealed a significant transposed-phoneme effect of comparable magnitude in the within- and across-syllable conditions. Syllable boundaries therefore do not appear to modulate transposed-phoneme effects, at least for adjacent syllables.

However, the syllable boundary manipulation in Experiment 1 was confounded with the number of syllables of the base-words. In other words, the within-syllable manipulation was operated on bi-syllabic stimuli, whereas the across-syllable manipulation was on tri-syllabic stimuli. This issue was addressed in Experiment 2, where we removed the confound between number of syllables and the syllable boundary manipulation. The results of Experiment 2 showed that transposing the two medial phonemes belonging to different syllables (/bi**k**.**s**ɔt/) in */bis.kɔt/*- type base-words did not diminish the magnitude of the transposed-phoneme effect neither when compared to a transposition that involved phonemes belonging to the same first syllable (/**s**i**b**.kɔt/) nor when compared to a transposition that involved phonemes belonging to the same final syllable (/bis.**t**ɔ**k**/). The results of Experiments 1 and 2 therefore allow us to reject the syllable boundary account of the divergent pattern of effects of non-adjacent transposed-phonemes seen in prior work.

It therefore remained to be explained why Dufour and Grainger (2021) failed to find a non-adjacent transposed-phoneme effect. In a post-hoc analysis of Experiment 1 we explored an alternative explanation based on differences in overall speed of responding in the Dufour and Grainger study and the present Experiment 1. We found a significant correlation between the size of non-adjacent word-final transposed-priming effects (the condition that failed to show an effect in Dufour & Grainger’s, 2021, experiment) and average speed of responding, with larger effects arising in the slower participants (see Appendix A). The results of this post-hoc analysis, and more generally the results of Experiment 1, were replicated in an additional experiment where we practically doubled the number of participants (see Appendix B). We are therefore confident in concluding that the null effect reported by Dufour and Grainger (2021) was due to the relatively fast RTs of the participants in that study.

How might speed of responding impact on the size of transposed-phoneme effects? Firstly, the impact of speed of responding must be considered in the light of the fact that we failed to find a correlation between this factor and the size of non-adjacent transposed-phoneme priming effects when the transposition occurred at the beginning of the stimulus. This is in line with the fact that the overall null effect of the position of transpositions in Experiment 1 was found to be modulated by speed of participants' responding, with a significant impact of this factor arising in the fastest participants (see replication of Experiment 1, Appendix B). That is, in the fastest participants, there was no transposed-phoneme effect when this occurred at word-final position, whereas a robust effect remained for word-initial transpositions. Thus, word-initial transpositions were influencing lexical decisions independently of the speed of responding of participants, thus pointing to an irrepressible impact of phoneme transpositions when they occur early enough or when the entire speech signal is short enough (e.g., the short monosyllabic words tested in the studies of Toscano et al., 2013; Gregg et al., 2019; and Dufour & Grainger, 2019, 2020). On the other hand, when the transposition arrives later in the speech signal and participants are trying to respond as rapidly as possible, it would appear that in this situation responses can be driven by information that was obtained prior to the transposition. Thus, upon hearing the initial part of the speech signal such as /*ʃo**l**o…*/, fast responders could make their decision on the basis of information that is deemed to be already incompatible with any real word.

This, of course, is only one possible explanation for the observed correlation between the size of transposed-phoneme effects and the speed of the participants' responses, and future experimentation could test this hypothesis using stimuli of various lengths in order to introduce more variation in the phoneme position from which the transposed nonwords diverge from their base-words (in the present study, all the */ʃo**l**o**k**a/* type of nonwords diverge from their base-words at the third phoneme). Note that our argumentation takes into account the phoneme position from which the nonword diverges from its base-word, and not the phoneme position from which the nonword diverges from all other words - the so-called deviation point. Here, it should be noted that even if an effect of nonword deviation point has not been clearly demonstrated in prior research (e.g., Goldinger, 1996), we controlled for a possible influence of this factor in our statistical analyses, and thus it cannot alone account for the present results. In our analyses, the deviation point was found to have a significant influence with longer *no* decisions for late deviation points. This is likely due to the fact that RTs were measured from the onset of nonwords. Indeed, prior research has shown that this factor no longer influences nonword decision time when RTs are time-locked to the deviation point (see O'Rourke & Holcomb, 2002, for a discussion).

In conclusion, in the present study we set-out to test a syllable boundary explanation for diverging results in prior studies investigating the effects of non-adjacent transposed-phonemes. Our results allow us to reject this syllable boundary interpretation, and show, on the contrary, that transposed-phoneme effects are not modulated by whether or not the transposition occurs within a syllable or across syllables. On the other hand, our analysis of differences in word-final phoneme transposition effects as a function of speed of participants' responding points to this as one key factor that can modulate the size of transposed-phoneme effects. We suggest that when the transposition occurs at the end of the speech signal, fast responders can avoid the negative impact of phoneme transpositions when making their lexical decision (i.e., longer “no' responses to transposed-phoneme nonwords) by responding prior to the complete processing of the transposed phonemes. Future work could fruitfully investigate effects of position of transposition in longer nonwords, while manipulating the deviation point of the nonword stimuli. More generally, the present study provides a further demonstration of a role for the flexible encoding of phoneme-order in speech perception, and aligns with studies conducted on written word perception (e.g. Grainger & van Heuven, 2004; Perea & Lupker, 2004) as well as a study on written and typed word production (Harrison et al., 2020). Hence, the flexible order-encoding of parts of a linguistic entity (e.g., phonemes and letters for spoken and written word comprehension and production) would appear to be modality-independent, and could thus constitute a general mechanism in language processing.

Footnotes

1: One possible way to reconcile transposed-phoneme effects with models that code for the precise order of segments is to incorporate the notion of noise in the order encoding process, hence mimicking certain models of orthographic processing (e.g., Gómez et al., 2008). However, such a model has not yet been developed for spoken-word recognition, and thus the TISK model that assumes position-independent phonemes is actually the sole model that can account for transposed-phoneme effects.

2: We acknowledge that controlling for syllable structure has introduced another confound in terms of the distance between the transposed phonemes, with non-adjacent transpositions in the within-syllable conditions and adjacent transpositions in the between-syllable condition.

Acknowledgements

This research was supported by ERC grant 742141.

References

Baayen, R. (2008). Analyzing linguistic data: A practical introduction to statistics. Cambridge, UK: Cambridge University Press.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390-412.

Baayen, R. H., and Milin, P. (2010). Analyzing reaction times. International Journal of *Psychological Research, 3*, 12–28.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bates, D. M. and Sarkar, D. (2007). lme4: Linear mixed-effects models using S4 classes, R package version 2.6, retrieved 20 November 2017 from http://lme4.r-forge.r-project.org/.

Dufour, S., & Grainger, J. (2021): When you hear /baksɛt/ do you think /*baskɛt*/? Evidence for transposed-phoneme effects with multi-syllabic words. *Journal of Experimental Psychology, Learning, Memory and Cognition*, in press.

Dufour, S., & Grainger, J. (2020). The influence of word frequency on the transposed-phoneme priming effect. *Attention, Perception, and Pyschophysics***,** *2060-2069.*

Dufour, S., & Grainger, J. (2019). Phoneme-order encoding during spoken word recognition: A priming investigation. *Cognitive Science, 43,* 1-16.

Dufour, S., Peereman, R., Pallier, C., & Radeau, M. (2002). VoCoLex: A lexical database on phonological similarity between French words. *L'Année Psychologique*, 102, 725–746.

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association, 105*(2), 203–206.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: a distributed model of speech perception. *Language and Cognitive Processes, 12,* 613–656.

Goldinger, S. D. (1996). Auditory lexical decision. *Language and Cognitive Processes,11,* 559-567.

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological Review, 115*, 577-601.

Grainger, J., & van Heuven, W. J. B. (2004). Modeling letter position coding in printed word perception. In P. Bonin (Ed.), *The Mental Lexicon* (pp. 1–23). New York: Nova Science.

Gregg, J., Inhoff, A.W. & Connine, C.M. (2019). Re (Re) re-considering the role of temporal order in spoken word recognition. *Quarterly Journal of Experimental Psychology, 72*, 2574-2583.

Hannagan, T., Magnuson, J. S., & Grainger, J. (2013). Spoken word recognition without a TRACE. *Frontiers in Psychology*, 4, 563.

Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language,* 59, 434-446.

Harrison, W., Hepner, C.R., J., & Nozari, N. (2020). Is segmental interference position-dependent. In Proceedings of the 42nd Annual Meeting of the Cognitive Science Society (pp. 681–687).

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interaction and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29-63.

Marslen-Wilson, W. D. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (pp. 148–172). Cambridge, MA: MIT Press.

Marslen-Wilson, W. D., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review, 101*, 653-675.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1– 86.

Norris, D. (1994). SHORTLIST: a connectionist model of continuous speech recognition. *Cognition*, *52*, 189– 234.

O'Rourke, T.B., Holcomb, P.J. (2002). Electrophysiological evidence for the efficiency of spoken word processing. *Biological Psychology, 60*, 121–150.

Perea, M., & Lupker, S. J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language, 51*, 231–246.

R Development Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available from http://www.R-project.org.

Toscano, J. C., Anderson, N. D., & McMurray, B. (2013). Reconsidering the role of temporal order in spoken word recognition. *Psychonomic Bulletin & Review, 20,* 981–987.

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J.,...Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin &Review, 25*, 35–57.

Table 1: Characteristics of the nonwords used in Experiment 1 (mean values)

	Initial position				**Final position**			
	Syllable number	Phoneme number	Deviation point[1]	Duration[2]	Syllable number	Phoneme number	Deviation point[1]	Duration[2]
Same-syllable								
Transposed (/**sib**.kɔt/;/bis.**tɔk**/)	2	6	3.96	734	2	6	4.73	704
Substituted (/**ʃip**.kɔt/; /bis.**pɔg**/)	2	6	3.83	731	2	6	4.54	709
Different-syllable								
Transposed (/**k**o.**ʃ**o.la/;/ʃo.**l**o.**k**a/)	3	6	4.56	611	3	6	4.48	600
Substituted (/**p**o.**f**o.la/; /ʃo.**r**o.**p**a/)	3	6	4.25	609	3	6	4.44	603

Note: [1] The phoneme position at which the nonword is no longer compatible with any word. [2] In milliseconds.

Table 2: Characteristics of the nonwords used in Experiment 2 (mean values)

	Syllable number	Phoneme number	Deviation point[1]	Duration[2]
Initial condition				
Transposed (/**s**i**b.**kɔt)	2	6	3.92	734
Substituted (/**ʃ**i**p.**kɔt/)	2	6	3.79	733
Medial Condition				
Transposed (/bi**k**.**s**ɔt/)	2	6	3.94	750
Substituted (/bi**p**.**f**ɔt)	2	6	3.92	748
Final Condition				
Transposed (/bis.**t**ɔ**k**/)	2	6	4.71	704
Substituted (/bis.**p**ɔ**g**/)	2	6	4.48	706

Note: [1] The phoneme position at which the nonword is no longer compatible with any word. [2] In milliseconds.

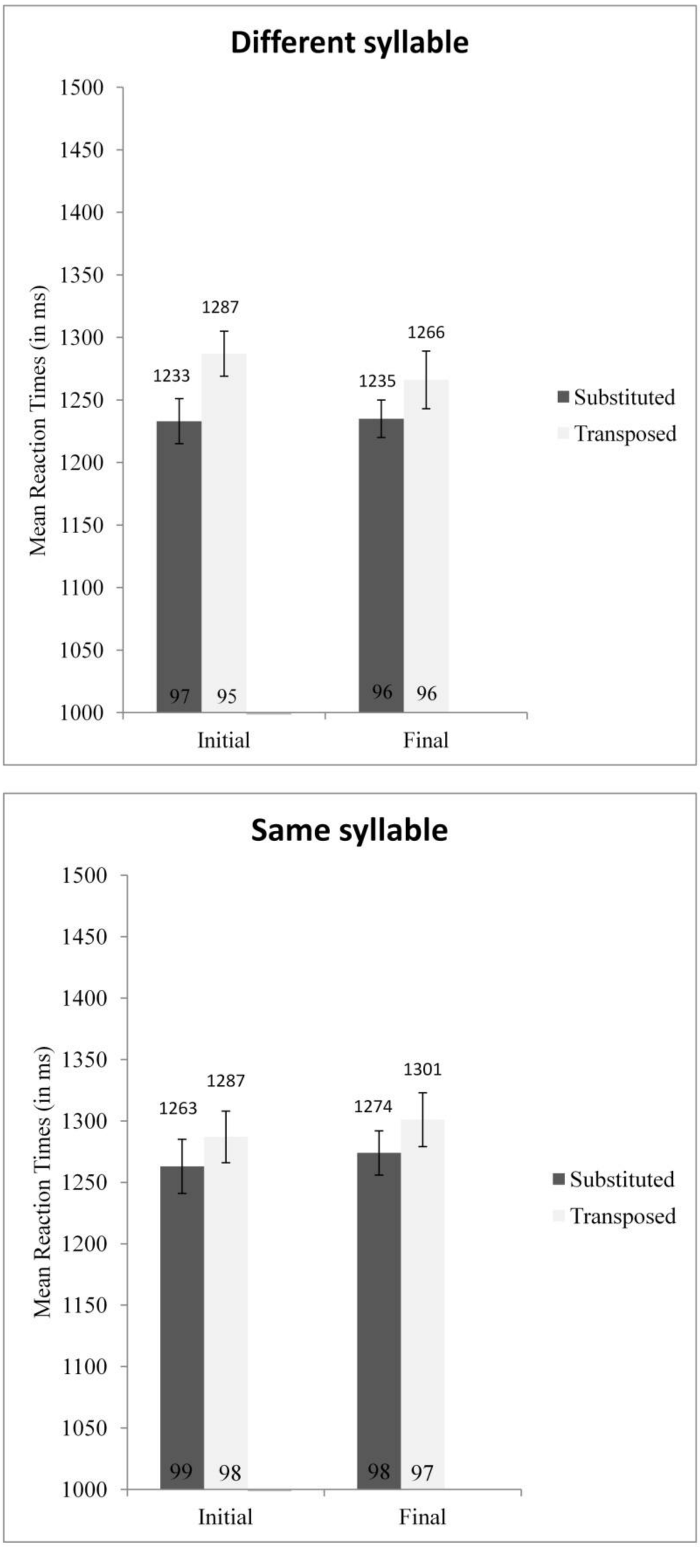

Figure 1: Mean Reaction Times (in ms) for the substituted and transposed nonwords in each of the position (initial vs. final) and syllable (same vs. different) conditions of Experiment 1. Percentages of correct responses are shown at the bottom of the graph. Error bars represent 95% confidence intervals.

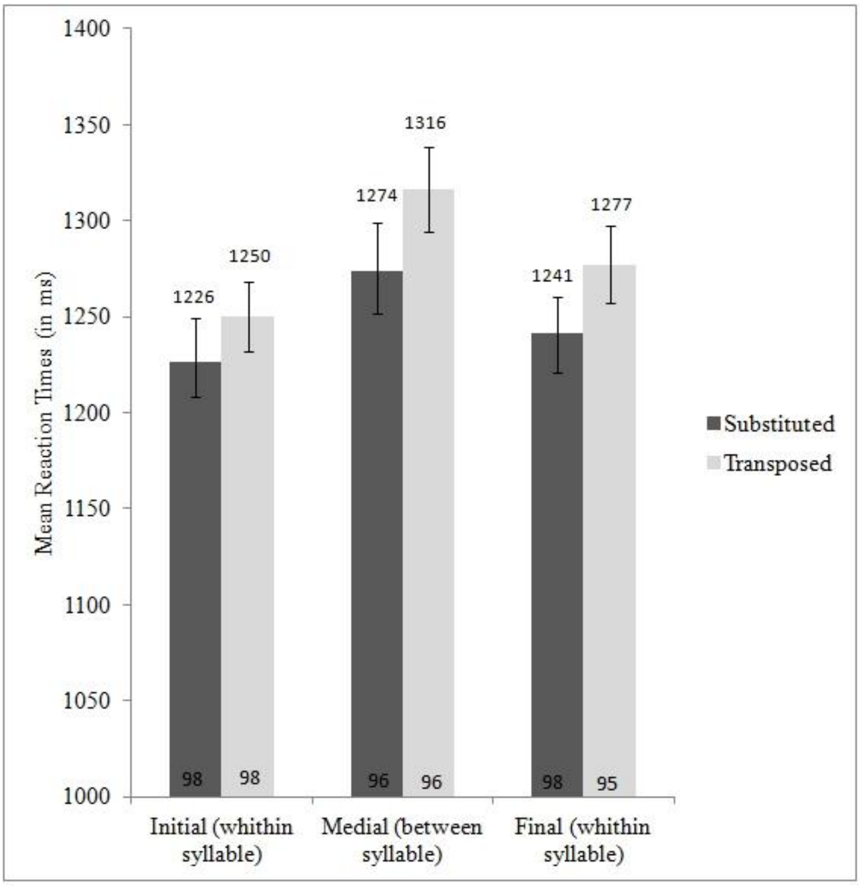

Figure 2: Mean Reaction Times (in ms) for the substituted and transposed nonwords in each position (initial, medial, final) of Experiment 2. Percentages of correct responses are shown at the bottom of the graph. Error bars represent 95% confidence intervals.

Appendix A: Results of the post-hoc analysis of Experiment 1.

A correlation analysis was conducted to examine whether the magnitude of the /ʃ*o***lok***a*/ effect, that is, the difference between the transposed and substituted nonwords in the different syllable final position condition, correlates with the average RT of participants in that condition. For each participant, the size of the transposed-phoneme effect was calculated by subtracting the mean RT (averaged across items) obtained for transposed-phoneme nonwords with that obtained for the substituted-phoneme nonwords. Pearson correlation coefficients and the associated *p*-value were calculated using statistica software. As illustrated in Figure A1, the correlation was positive (r(86)=.32) and significant (p<.01) showing that the size of the transposed-phoneme effect increased as RTs increased.

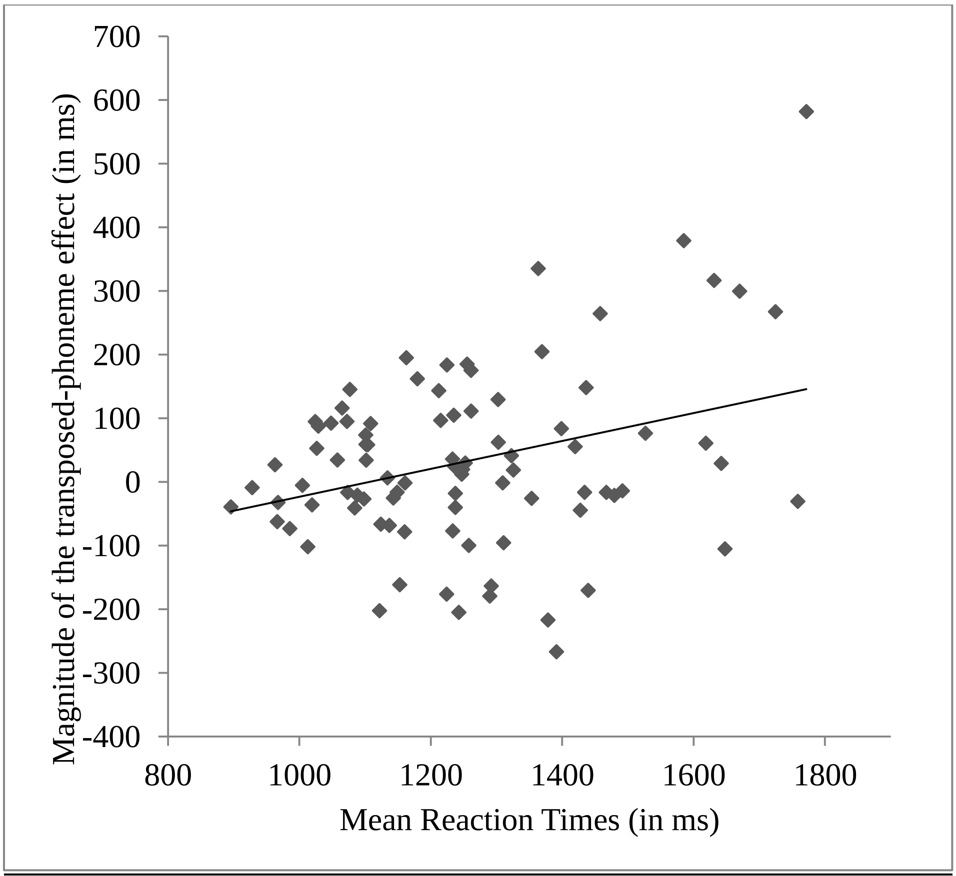

Figure A1: Scatterplot of the magnitude of the /ʃo**lok**a/ effect as a function of participants' mean RT in Experiment 1.

Appendix B: Replication of Experiment 1 with CV.CV.CV type base-words.

A total of 185 participants participated in the experiment. Twenty participants were excluded from the analyses. Among them 17 participants had an error rate superior to 50% and the 3 others had excessively long RTs (greater than 2000 ms on average and greater than 3 standard deviations above the general mean). The mean RT and percentage of correct responses to nonwords in each condition are presented in Figure B1.

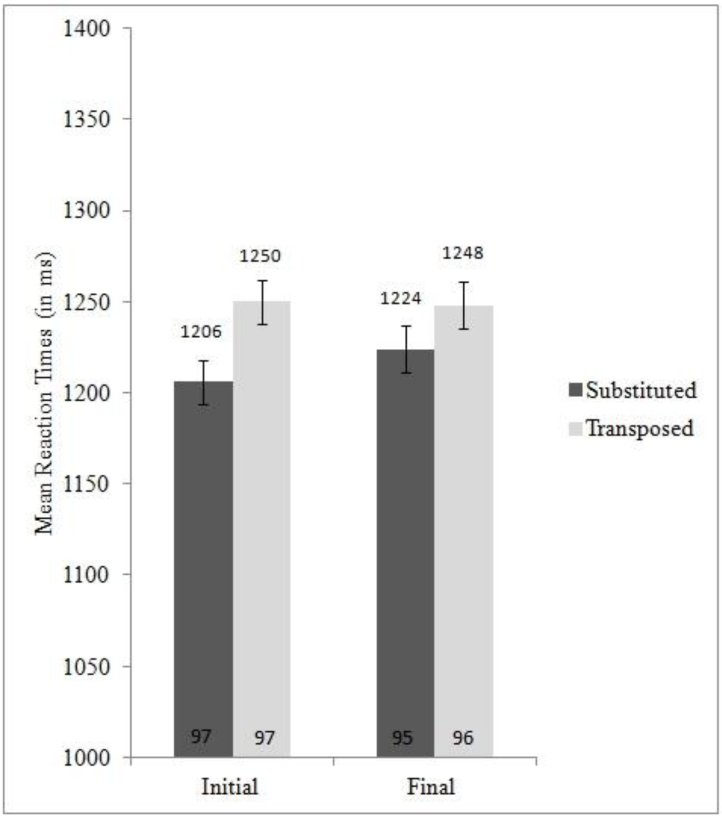

Figure B1: Mean Reaction Times (in ms) for the substituted and transposed nonwords in each position condition (initial, final) of the replication of Experiment 1. Percentages of correct responses are shown at the bottom of the graph. Error bars represent 95% confidence intervals.

RTs to nonwords (available at https://osf.io/gwk2j/) were analyzed using linear mixed effects models following the same procedure as in Experiment 1. The RT analysis was performed on correct responses, thus removing 305 data points out of 8580 (3.55%). Six RTs <600 ms and 34 RTs >5000 ms were considered as outliers (less than 1%) and were also excluded from the analysis. Additionally, within each condition and for each participant RTs lying more than 3 standard deviations from the participant's mean were excluded (29 data points; less than 1%). For the model to meet the assumptions of normally-distributed residuals and homogeneity of variance, a log transformation was applied to the RTs (Baayen & Milin, 2010) prior to running the model. The model was run on 8206 data points and included the variable Nonword Type (transposed, substituted), Position (initial, final) and their interaction as fixed effects. The nonword deviation point was also entered as factor. The model also included participants and items as random intercepts.

The effect of Position was not significant (*b* =0.0051, *SE* = .0077, *t* = 0.66). Crucially, the model revealed a significant effect of Nonword Type with RTs being slower for the transposed-phoneme nonwords than for the substituted-phoneme nonwords (*b* =0.0194, *SE* = 0.0078, *t* = 2.50). Again, this effect was of similar magnitude when the transposition occurred at the initial or final of the words, as revealed by the lack of interaction between Nonword Type and Position (*b* =-0.0189, *SE* = .0155, *t* = -1.22). As found in the post-hoc analysis of Experiment 1 (Appendix A), the correlation illustrated in Figure B2 was positive (r(163)=.30) and significant (p<.001), showing again that the size of the transposed-phoneme effect increased as RTs increased.

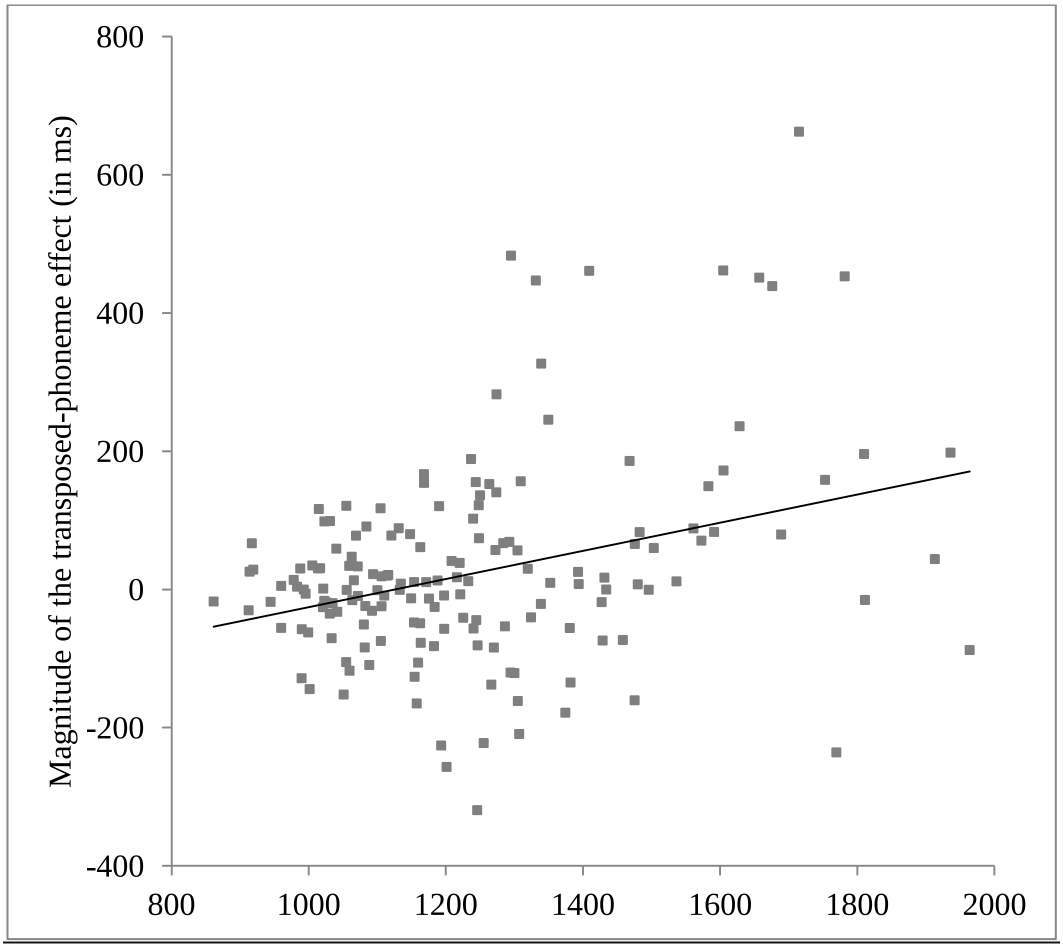

Figure B2: Scatterplot of the magnitude of the /ʃo.**l**o.**k**a/ effect as a function of participants' mean RT in the replication of Experiment 1.