When you hear /baks t/ do you think /bask t/? Evidence for transposed-phoneme effect with multisyllabic words.

Sophie Dufour, Jonathan Grainger

To cite this version:

Sophie Dufour, Jonathan Grainger. When you hear /baks t/ do you think /bask t/? Evidence for transposed-phoneme effect with multisyllabic words.. Journal of Experimental Psychology: Learning, Memory, and Cognition, American Psychological Association, In press, 10.1037/xlm0000978 . hal-03141336

HAL Id: hal-03141336
https://hal.archives-ouvertes.fr/hal-03141336

Submitted on 15 Apr 2021

When you hear /baksɛt/ do you think /*baskɛt*/?

Evidence for transposed-phoneme effects with multi-syllabic words

Sophie Dufour[1,3] 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

Author note: The data of the two experiments are available at https://osf.io/9pmb2/files/

(Open Science Framework; Foster & Deardorff, 2017)

Abstract: In this study we asked whether nonwords created by transposing two phonemes (/bi**ks**ɔt/) are perceived as being more similar to their base words (/*bi**sk**ɔt*/) than nonwords created by substituting two phonemes (/bi**pf**ɔt/). Using the short-term phonological priming and a lexical decision task, Experiment 1 showed that transposed-phoneme nonword primes lead to shorter RTs on the target base words than substituted-phoneme nonword primes. Using a single-presentation lexical decision task, Experiment 2 showed that transposed-phoneme nonwords lead to longer "no" decision responses than substituted-phoneme nonwords. In both Experiments 1 and 2, the transposed-phoneme effect was observed when the transposed phonemes were adjacent (/bi**ks**ɔt/-/*bi**sk**ɔt*/) but not when they were distant (/ʃo**l**o**k**a/-/*ʃo**k**o**l**a*/). Our findings suggest that nonwords created by transposing adjacent phonemes in real words generate more activation of the lexical representations associated with the base words than do matched control nonwords. More generally, our findings present a challenge for models of spoken word recognition that code for the precise order of speech segments.

One key characteristic of the speech signal is that it unfolds over time. Several decades of research have consequently led to the strong and widely held assumption that the sounds that make up spoken words are immediately assigned to their correct positions in words. However, four recent studies (Dufour & Grainger, 2019, 2020; Gregg et al., 2019; Toscano et al., 2013) report results that challenge this assumption and suggest, on the contrary, that the position coding of phonemes is more flexible than previously assumed, and also that position-independent phonemes may play a role in spoken word recognition.

The first demonstration was provided by Toscano et al. in 2013. Using the visual world paradigm, these authors examined the eye movements of participants who followed spoken instructions to manipulate objects pictured on a computer screen. They found more fixations on the picture representing a CAT than on a control picture (e.g., the picture of a MILL) when the spoken target was TACK. Such a finding thus suggests that CAT and TACK are confusable words even if the consonants that they shared are not in the same positions. Also, Toscano et al. (2013) showed that the probability of fixating transposed words was higher than the probability of fixating words sharing the same vowels at the same position plus one consonant in a different position (e.g., SUN-BUS). This finding suggests that the transposed-phoneme effect is due to more than just vowel position overlap in the transposed words, and that complete phonemic overlap is a necessary condition in order to obtain transposed-phoneme effects. The main finding of Toscano et al. (2013), that CAT and TACK are confusable words, was replicated by Gregg et al. (2019) with a larger set of items. At the same time, Gregg et al. (2019) showed that words without vowel position overlap (e.g., LEAF-FLEA) were not fixated more that unrelated words. Such a finding could argue for a special status for vowels, and in particular that positional vowel match is critical in the observation of transposed-phoneme effects (Gregg et al., 2019; see also Dufour & Grainger, 2019). It is also possible that the distance separating the transposed phonemes could be a factor determining the size of transposed-phoneme effects in line with findings showing that the size of transposed-letter effects diminish as the distance between the two letters increases (e.g., Perea et al., 2008).

Using the phonological priming paradigm, two other studies (Dufour & Grainger, 2019; 2020) conducted in French have also shown that speech input like [byt] facilitates not only the subsequent processing of an identical target word /byt/ BUT “goal” but also that of a target word /tyb/ TUBE “tube” that contains the same phonemes in a different order. This transposed-phoneme priming effect was found when unrelated words (*MOULE* /mul/ “mussel” – *TUBE* /tyb/ “tube”), vowel overlap words (*PUCE* /pys/ “flea” - *TUBE* /tyb/ “tube”) and vowel plus one consonant in a different position overlap words (*BULLE* /byl/ bubble – *TUBE* /tyb/ “tube”) were used as control conditions, thus providing further support to prior observations of transposed-phoneme effects. We also observed that the transposed-phoneme priming effect (/byt/-/tyb/) differed from repetition priming (/tyb/-/tyb/) by both its magnitude and its time course. The transposed-phoneme priming effect was significantly smaller than the repetition priming effect, and was only obtained using a short-term priming procedure with targets immediately following primes, while the repetition priming effect also occurred in a long-term priming paradigm with primes and targets presented in separated blocks of stimuli. In a follow-up study (Dufour & Grainger, 2020), we reported that the transposed-phoneme priming effect occurs when targets have a higher frequency than primes, but not when they have a lower frequency.

Altogether, these findings suggest that position-independent phonemes play a role in spoken word recognition. They are thus challenging for some of the most influential 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) that code for the precise order of segments, and assume that the phonological form of words consists of an ordered sequence of sounds. As we discussed in our preceding papers (Dufour & Grainger, 2019; 2020), the TISK model (Hannagan et al., 2013; see You & Magnuson, 2018, for a more recent implementation) is at present the sole model of spoken word recognition that can account for transposed-phoneme effects.[1] TISK is an interactive-activation model similar to the TRACE model (McClelland & Elman, 1986), but it replaces the position-dependent units in TRACE by both a set of position-independent phoneme units (see Bowers et al., 2016, for evidence for a role for position-invariant phonemes in spoken word recognition) and a set of open-diphone units that represent ordered sequences of contiguous and non-contiguous phonemes (cf. the open-bigram representations proposed by Grainger & van Heuven, 2004, for visual word recognition). Within such a framework, both the position-independent phoneme units and the open-dipone representations can contribute to transposed-phoneme effects. Thus, a nonword like /baksɛt/ will activate a set of open-diphone representations such as b-a, b-k, b-s, a-k, a-s, k-s, k-t, - many of which are compatible with the word */baskɛt/*. A nonword such as /bapfɛt/, on the other hand, activates many open-diphones that are incompatible with /basket/. The set of open-diphone representations that are generated by a given speech input is governed by the distance parameter that determines the degree of separation (in number of phonemes) that can be tolerated between the constituent phonemes and/or a weighting assigned as a function of the distance (see Hannagan &

---

[1] We note nevertheless, that transposed-phoneme effects are a relatively new phenomenon, and at the time that TRACE was developed, research on spoken word recognition was primarily concerned with the nature and the direction of information flow.

Grainger, 2012, for a discussion of the distance parameter within the more general framework of string kernels). Amount of overlap between a given phoneme sequence and a real word, both in terms of position-independent phonemes and open-diphones, determines how well that phoneme sequence can activate the lexical representation of the word.

In a priming context, therefore, a transposed-phoneme prime will partially activate the lexical representation corresponding to its base word, and more so than a control prime. This partial activation explains why transposed-phonemes effects are restricted to a short-term priming paradigm, and were not observed, contrary to repetition priming, in a long-term priming procedure (Dufour & Grainger, 2019). The partial activation of the transposed target words during prime processing quickly dissipates over time, and only fully activated lexical representations resist the longer delay and impact of intervening items in long-term priming. Although word frequency is not yet implemented in TISK, an implementation of frequency via, for example, variation in the connection strengths between sub-lexical phone and biphone representations and lexical representations (see Dahan et al., 2001, for simulations with a connectionist model) could also account for our observation that transposed-phoneme priming effects occur when the targets were of higher frequency than the primes (Dufour & Grainger, 2020). These bottom-up connections would allow the position-independent phoneme units and open-diphone representations to generate more activation in compatible lexical representations when the words increase in frequency.

Because, to this date, transposed-phoneme effects have only been observed with short-words, in this study, we aimed to examine the scope of transposed-phoneme effects, and whether they can be found with multisyllabic words. This is important because a complete account of the processes involved in spoken-word recognition must include multisyllabic as well as monosyllabic words. Because, from a strictly methodological point of view, it was not possible to find a sufficient number of pairs of long words created by transposing two phonemes, nonwords were used as primes and were created by transposing two medial consonants of long words (e.g. /bi**ks**ɔt/ created from /*bi**sk**ɔt*/ BISCOTTE "toasted bread").[2] The use of non-words constitutes perhaps the most important point of this study, because they enabled the testing of a key prediction of the TISK model. In TISK, a transposed nonword /bi**ks**ɔt/ sharing all of their phonemes, but in different positions, with a base word /*bi**sk**ɔt*/ should activate more strongly the lexical representation corresponding to the base word than a phonological control non-word created by substituting two phonemes /bi**pf**ɔt/ of the base word. As a result, in TISK, transposed-phoneme nonwords should be perceived as being more similar to their base words than substituted-phoneme nonwords. This prediction is not made by models that code for the precise order of segments (Gaskell & Marslen-Wilson, 1997; Marslen-Wilson & Warren, 1994; Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1990; McClelland & Elman, 1986; Norris, 1994). Since the number of shared phonemes with the base word at the same positions is identical in the two types of nonwords, models coding for the precise order of segments predict that transposed- and substituted- nonwords would similarly activate the lexical representation corresponding to the base word. As a result, in this kind of model, transposed-phoneme nonwords would not be perceived as being more similar to their base words than substituted-phoneme nonwords. Furthermore, because the study of Gregg et al. (2019) suggests that the distance separating the transposed phonemes could be an important factor in determining the size of transposed-phoneme effects, we also compared

---

[2] In order to distinguish examples of word and nonword stimuli in the present study, all words are printed in italics.

nonwords created by transposing adjacent phonemes (e.g. /bi**ks**ɔt/ from /*bi**sk**ɔt*/) and nonwords created by transposing nonadjacent phonemes (e.g. /ʃo**l**o**k**a/ from /*ʃo**k**o**l**a*/ CHOCOLAT “chocolate”). From a theoretical perspective, given that the distance parameter is known to impact on the ability of open-bigram coding to account for transposed-letter effects, we expected the same impact of distance for open-diphone coding with a smaller transposed-phoneme effect in the nonadjacent phoneme condition.

Experiment 1

Experiment 1 used the short-term priming paradigm with prime and target words separated by a 20 ms ISI. Two conditions of distance separating the transposed phonemes were tested: Adjacent (/bi**ks**ɔt/-/*bi**sk**ɔt*/) vs. nonadjacent (/ʃo**l**o**k**a/-/*ʃo**k**o**l**a*/) conditions. Within each condition of distance, primes were of three types: Transposed-phoneme nonwords (/bi**ks**ɔt/-/*bi**sk**ɔt*/; /ʃo**l**o**k**a/-/*ʃo**k**o**l**a*/), substituted-phoneme nonwords (/bi**pf**ɔt/-/*bi**sk**ɔt*/; /ʃo**r**o**p**a/-/*ʃo**k**o**l**a*/), and repeated prime words sharing with the target all of their phonemes in the same order (/*biskɔt*/-/*biskɔt*/; /*ʃokola*/- /*ʃokola*/). The predictions were straightforward. If as predicted by the TISK model, transposed-phoneme nonwords generate more activation in the lexical representations corresponding to the base words than substituted-phoneme nonwords, then faster RTs on the subsequent target words should be observed in the transposed non-word priming condition due to stronger residual activation associated with the target word. Moreover, if as suggested by the results of Gregg et al. (2019), the distance separating the transposed phonemes is an important factor in determining the size of transposed-phoneme effects, then greater priming effect could be observed in the adjacent transposed-phoneme condition in comparison with the non-adjacent transposed-phoneme condition.

Method

Participants. Forty-eight French speakers (8 men, 18-25 years, mean age=20.12) from Aix-Marseille University participated in the experiment. All participants reported having no hearing or speech disorders.

Materials. Fifty-four target words, five to six phonemes in length, with a CVCCV(C) syllabic structure were selected and were used in the adjacent condition. 54 other target words, six to seven phonemes in length, with a (C)CVCVCV syllabic structure were also selected and were used in the non-adjacent condition. For each target word, two nonword primes were created. One was created by transposing the two medial consonants of the target word (/bi**ks**ɔt/ for /*bi**sk**ɔt*/ BISCOTTE "toasted bread") in the adjacent condition and by transposing the two consonants adjoining the medial vowel (/ʃo**l**o**k**a/ for /*ʃo**k**o**l**a*/ CHOCOLAT "chocolate") in the nonadjacent condition. The other was created by replacing the two medial consonants of the target word (/bi**pf**ɔt/ for /*bi**sk**ɔt*/) in the adjacent condition and by replacing the two consonants adjoining the medial vowel (/ʃo**r**o**p**a/ for /*ʃo**k**o**l**a*/ CHOCOLAT "chocolate") in the nonadjacent condition. In each of the distance conditions, 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). For example the phonemes /p/ and /k/ of the nonwords

/bi**pf**ɔt/ and /bi**ks**ɔt/ are both voiceless plosives differing on place of articulation only, and /f/ and /s/ are both voiceless fricatives also differing on place of articulation only. Similarly, the phonemes /r/ and /l/ of the non-words /ʃo**r**o**p**a/ and /ʃo**l**o**k**a/ are both liquids, and /p/ and /k/ are again voiceless plosives. Position-specific phonetic similarity between the substituted and transposed nonwords with the original words was also evaluated. For the substituted nonwords of the adjacent condition, the average number of shared phonetic features was 1.70 and 1.65 out of four for the first and second consonant respectively (e.g., the /p/ of the nonword /bi**pf**ɔt/ with the /s/ of the word /*bi**sk**ɔt*/ and the /f/ of the nonword /bi**pf**ɔt/ with the /k/ of the word /*bi**sk**ɔt*/). For the transposed nonwords of the adjacent condition, it was on average 1.70 for the two consonants (e.g., the /k/ of the nonword /bi**ks**ɔt/ with the /s/ of the word /bi***sk***ɔt/). For the substituted nonwords of the nonadjacent condition, the average number of shared phonetic features was 1.61 and 1.57 out of four for the first and second consonant respectively (e.g., the /r/ of the nonword /ʃo**r**o**p**a/ with the /k/ of the word /*ʃo**k**o**l**a*/ and the /p/ of the nonword /ʃo**r**o**p**a/ with the /l/ of the word /*ʃo**k**o**l**a*/). For the transposed nonwords of the nonadjacent condition, it was on average 1.48 for the two consonants (e.g., the /l/ of the nonword /ʃo**l**o**k**a/ with the /k/ of the word /*ʃo**k**o**l**a*/). Note that the small differences in phonetic similarity for the two types of nonwords were due to only three stimuli among the 54 in the adjacent condition, and to 9 stimuli among the 54 in the nonadjacent conditions. Removing these stimuli did not change the pattern of results. The main characteristics of the prime and the target words are given in Table 1. The complete set of prime and target words are given in Appendix 1.

<Insert Table 1 about here>

Three experimental lists were created using a Latin-square design so that each of the 108 target words were preceded by the three types of prime (repeated, transposed, substituted) across different participants, and participants were presented with each target word only once. For the purpose of the lexical decision task, 108 target nonwords were added to each list. The nonwords were created by changing the last phoneme of words not used in the experiment (e.g. the nonword /garav/ derived from the word /*garaʒ*/ GARAGE "garage"). This allowed us to have wordlike nonwords, and to encourage participants to listen to the stimuli up to the end prior to giving their response. So that the target nonwords followed the same criteria as the target words, 36 of them were paired with a repeated prime sharing all phonemes in the same order (e.g. /ʃarpãd/-/ʃarpãd/), 36 other with a transposed-phoneme nonword prime (e.g. /pa**ʃ**a**r**yd/-/pa**raʃ**yd/), and the remaining 36 nonwords were paired with substituted-phoneme nonword prime (e.g. /ka**ʃt**yg**/-**/ka**ps**yg/). Finally, to avoid strategic anticipation from the primes, 438 fillers consisting in prime and target pairs without any relation were added to each list. Again, for the purpose of the lexical decision task, half of the filler targets were words and the other half were non-words. So that the filler target words mimicked the experimental target words, a third of the 219 filler target words were paired with a word prime and the remaining with a nonword prime. To avoid that word primes be paired only with words as targets, 109 filler target nonwords were preceded by a prime word. The remaining 110 filler target nonwords were preceded by a nonword prime. Thus, an equal number of target words and nonwords were preceded by a word (one third) or a nonword (two thirds). 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. Note that in order to minimize the influence of coarticulation effects the transposed nonwords were produced as such, and they were not created by inverting the critical phonemes directly in the speech signal.

Procedure. Participants were tested in a sound-attenuated booth. Stimulus presentation and recording of the data were controlled by a PC running E-Prime software. Primes and targets were presented over headphones at a comfortable sound level, and an interval of 20 ms (ISI) separated the offset of the prime and the onset of the target. Participants were asked to make a lexical decision as quickly and accurately as possible on the target stimuli, with "word" responses being made using their dominant hand on an E-Prime response box that was placed in front of them. RTs were recorded from the onset of target stimuli. The prime-targets pairs were presented randomly and an inter-trial interval of 2000 ms elapsed between the participant's response and the presentation of the next pair. Participants were tested on only one experimental list and began the experiment with 10 practice trials.

Results & Discussion

One participant and two target words in the adjacent condition that gave rise to an error rate of more than 70% were removed from the analyses. The mean RT and percentage of correct responses on target words in each priming condition and for each condition of distance are presented in Figure 1.

<Insert Figure 1 about here>

RTs on target words (available at https://osf.io/9pmb2/files/) were analyzed using linear mixed effects models with participants and target words 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 153 (3.07%) data points out of 4982. RTs greater than 1500 ms (2.87%) 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 4686 data points. We reported the results of a model with the variables prime type (repeated, transposed, substituted), distance (adjacent, nonadjacent) and their interaction entered as fixed effects. Model comparison using the log-likelihood ratio test revealed that this model fit the data significantly better than a model without the interaction term ($\chi^2 = 9.15$, $p<.05$). The model also included participants and items as random intercepts, plus random participant slopes for the within-participant factors prime type and distance, and item slopes for the within-item factor prime type (see Barr et al., 2013). Note that the model failed to converge when random slopes were included for the interaction term in addition to the main effects.

The intercept was the performance on the target words preceded by substituted-phoneme nonword primes in the adjacent condition. The full results are displayed in Appendix 2. The model revealed a significant repetition priming effect with RTs on target words in the adjacent condition being 78 ms shorter when preceded by repeated word primes in comparison to substituted-phoneme nonword primes ($\beta = -.10$, $SE = .01$, $t = -7.50$; $p<.001$). Crucially here, the model also revealed a significant transposed-phoneme priming effect with RTs on target words in the adjacent condition being 19 ms shorter when preceded by transposed-phoneme nonword primes in comparison to substituted-phoneme nonword primes ($\beta = -.02$, $SE = .01$, $t = -2.43$; $p<.05$). The model also revealed that the transposed-phoneme priming effect significantly interacted with the factor distance ($\beta = .03$, $SE = .01$, $t = 2.49$; $p<.05$). To understand the nature of this interaction, the model was releveled such that the performance on the target words preceded by substituted-phoneme nonword primes in the non-adjacent condition was the intercept. No significant transposed-phoneme priming effect was observed in the non-adjacent condition (β =.01, *SE* = .01, *t* = 0.77; *p*>.20).

The percentage of correct responses was analyzed using a mixed-effects logit model (Jaeger, 2008) following the same procedure as for RTs. The model revealed more correct responses on target words of the adjacent condition when preceded by substituted-phoneme nonword primes in comparison to repeated word primes (β =.65, *SE* = .29, *z* = 2.24; *p*<.05). This difference significantly interacted with the factor distance (β =-.88, *SE* = .43, *z* = -2.08; *p*<.05) and was observed only in the adjacent condition. No other differences were significant.

To sum-up, the results of Experiment 1 showed that nonwords created by transposing two phonemes of a target word prime more the processing of that target than nonwords created by substituting two phonemes, at least when the transposed phonemes are adjacent. Such a finding thus suggests that under some condition transposed-phoneme nonwords activate more strongly the lexical representations corresponding to the base words than do substituted-phoneme control nonwords. Before discussing the implications of these findings, another demonstration in favor of position-independent phonemes would be to show that transposed-phoneme nonwords are harder to classify as target nonwords in a lexical decision task than substituted-phoneme nonwords. This was tested in Experiment 2.

Experiment 2

In this experiment, we used an unprimed lexical decision task in which both the nonword and word primes used in Experiment 1 were used as targets.

Method

Participants. Forty-eight French speakers (8 men, 18-25 years, mean age=20.06) from Aix-Marseille University participated in the experiment. All participants reported having no hearing or speech disorders. None had participated in Experiment 1.

Materials and Procedure. The three lists of prime stimuli of Experiment 1, composed of 36 repeated primes, 36 transposed nonword primes and 36 substituted nonword primes (18 for each of the distance conditions - adjacent vs. non-adjacent) were used as targets in this experiment. The repeated primes became the word targets, and the two types of nonword primes became the two types of nonword targets. 72 nonword prime fillers from Experiment 1, created by changing the last phoneme of words, were included as targets in the present experiment, giving a total of 144 nonwords. In order to have an equal number of words and nonwords 108 word targets were added. The same procedure as in Experiment 1 was used except that the prime stimuli in that experiment became the targets in Experiment 2, and participants made lexical decisions to these nonwords and words. Note that the repeated word primes of Experiment 1 were maintained here only for the purpose of the lexical decision task, and thus they were not further analyzed. As in Experiment 1, these triplets of stimuli (base word, transposed-phoneme nonword, substituted-phoneme nonword) were rotated across three lists, and participants only saw one item from each triplet.

Results & Discussion

Data concerning one transposed-phoneme nonword in the adjacent condition that gave rise to an error rate of more than 70% was removed from the analyses. Data concerning the corresponding substituted-phoneme nonword was also discarded. The mean RT and percentage of correct responses to nonwords in each condition of distance are presented in Figure 2.

<Insert Figure 2 about here>

The RT analysis was performed on correct responses, thus removing 101 (2.95%) data points out of 3424. RTs greater than 2500 ms (less than 1%) were also excluded from the analysis. The model was run on 3298 data points (available at https://osf.io/9pmb2/files/). The model included nonword type (transposed, substituted), distance (adjacent, nonadjacent) and their interaction as fixed effects. Model comparison using the log-likelihood ratio test revealed that this model fit the data significantly better than a model without the interaction term ($\chi^2 = 4.01$, $p<.05$). The model also included participants and items as random intercepts, plus random participant slopes for the within-participant factors nonword type and distance.

The intercept was the performance on the substituted-phoneme nonwords in the adjacent condition. The full results are displayed in Appendix 2. The model revealed a significant effect of type of nonword with RTs in the adjacent condition being 46 ms slower for the transposed-phoneme nonwords than for the substituted-phoneme nonwords (*β* =.04, *SE* = .01, *t* = 2.95; *p*<.01). The effect of the distance was also significant (*β* =.04, *SE* = .01, *t* = 2.66; *p*<.01) with RTs for the substituted-phoneme nonwords being slower in the nonadjacent condition than for the adjacent condition. As in Experiment 1, the model revealed that the effect of nonword type significantly interacted with the factor distance (*β* =-.04, *SE* = .02, *t* = -1.998; *p*<.05). To understand the nature of this interaction, the model was releveled such that the performance on substituted-phoneme nonwords in the nonadjacent condition was the intercept. No significant effect of the type of nonword was observed in the nonadjacent condition (*β* =.002, *SE* = .01, *t* = 0.14; *p*>.20).

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, and so we will not discuss them further.

To sum-up, the results of Experiment 2 parallel those of Experiment 1 and indicate that adjacent transposed-phoneme nonwords are harder, in terms of longer RTs, to classify as nonwords than substituted-phoneme nonwords. Again, no evidence of a transposed-phoneme effect was found when the transposed phonemes were not adjacent.

General Discussion

One key prediction of the TISK model (Hannagan et al., 2013) is that nonwords (/bi**ks**ɔt/) created by transposing two phonemes of a real word (*/bi**sk**ɔt/*) should be perceived as being more similar to the base words (*/bi**sk**ɔt/*) than nonwords created by substituting two phonemes of the same words (/bi**pf**ɔt/). In accordance with this prediction, Experiment 1 revealed that when the critical phonemes are adjacent, 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, Experiment 2 tested nonwords presented in isolation as targets in a lexical decision task and showed that transposed-phoneme nonwords (/bi**ks**ɔt/) took longer to classify as nonwords compared with substituted-phoneme nonwords (/bi**pf**ɔt/). Similarly to Experiment 1, this transposed-phoneme effect in nonword decision latencies was only observed when the transposed phonemes were adjacent. Altogether, these findings indicate that nonwords created by transposing two adjacent phonemes activate to a greater degree the lexical representations of their base words than do nonwords created by substituting two phonemes. Consequently, both greater priming effects and longer lexical decisions for "no" responses were found with transposed-phoneme nonwords.

In both experiments there was no evidence for a transposed-phoneme effect when the transposition involved non-adjacent phonemes (e.g., /ʃo**l**o**k**a/ derived from */ʃo**k**o**l**a/*). As we discussed earlier, TISK has a distance parameter that governs open-diphone coding, and this leads the model to predict that phoneme transposition effects will be smaller with more distant transpositions. However, the null effect found in the non-adjacent condition in the present study is not in accordance with the robust transposed-phoneme effects found with CVC monosyllabic words in previous studies (Dufour & Grainger, 2019; 2020; Gregg et al., 2019; Toscano et al., 2013) and in which the transposed consonants were nonadjacent (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). A closer look at our materials revealed that the non-adjacent transposed phonemes in our multisyllabic items belonged to different syllables (/ʃo.**lo**.**k**a/ for */ʃo.**ko**.**la**/*), which was not the case with prior work investigating monosyllabic words, for which transposed phonemes inevitably belonged to the same syllable. This discrepancy between the present study and our prior work points to a possible role for syllable boundaries in transposed-phoneme effects. One possibility is that consonants could migrate across their respective positions within a syllable, but not across syllables[3]. Evidence for constraints imposed by syllable boundaries has already been observed in studies examining activation of embedded words. For example, Bowers et al. (2009) reported evidence for final embedded word activation that differed from the carriers words by at least three phonemes, when the embedded words were aligned with a syllable boundary (e.g., *bat* in acrobat) but not when they were misaligned with a syllable boundary (e.g. *ram* in diagram). This points to the possible need to integrate syllabic constraints in the TISK model, in terms, for example, of what position-independent phonemes can and cannot activate.

An alternative interpretation, however, is that there is a trade-off between the amount of bottom-up information that is compatible and that is incompatible with a given target word, and given a potential difference in the timing of bottom-up facilitation and inhibition this will depend on the location of this information in the target word. Thus, upon hearing /ʃo**lo…**/ the

---

[3] As in the present study, the transposed monosyllabic words in Dufour and Grainger's (2019) study were produced as such and were not created by inverting the critical phonemes directly in the speech signal. Hence, if we consider the word /ʀɔb/, the /ɔ/ vowel, due to coarticulation, contains acoustic traces corresponding to the /b/ of /ʀɔb/ and not to the /ʀ/ of the transposed word /bɔʀ/). We are thus confident that the transposed phoneme effect observed with nonwords like /bi**ks**ɔt/ is due to more than simple coarticulatory effects which would extend in this case to the nonadjacent /s/ phoneme.

amount of information that is already incompatible with the base word */ʃo**k**o**la**/* could outweigh the bottom-up input from location-invariant phonemes and compatible open-diphones, and especially when this information arrives too late. In this way, a prime such as /tyb/ can activate the target word /byt/ because the bottom-up support from location-invariant phonemes arrives rapidly enough to outweigh the negative evidence from incompatible open-diphones. This interpretation predicts that it is the location of the transposed phonemes in the non-adjacent condition that is critical. We would therefore expect to observe non-adjacent transposition effects when the transposition occurs early in the base word, such as with /**k**o**ʃ**ola/ derived from */**ʃ**o**k**ola/*. This interpretation in terms of temporal dynamics is to some extent in accordance with the observations made in other modalities. Indeed, although clear transposition effects were found in the visual modality with non-adjacent transposed letters in *chocolat* /ʃokola/ type words (e.g. Perea & Lupker, 2004), Perea et al. (2012) failed to find clear evidence for such an effect in the tactile modality with Braille reading, which like spoken word recognition involves serial processing. Also, in a more recent study, Marcet et al. (2019) reported a significant but reduced transposition effect in the visual modality with non-adjacent transposed letters when the stimuli were presented serially, letter by letter, thus again simulating processes involved in spoken word recognition. Hence, whether or not this temporal interpretation could be distinguished from the syllabic interpretation proposed above remains to be seen. Future simulation studies could examine the timing of bottom-up facilitation and inhibition and how this timing is modulated as a function of factors such as position and word length.

Our findings pose problems for models of spoken word recognition that code for the precise order of segments (Gaskell & Marslen-Wilson, 1997; Marslen-Wilson & Warren, 1994; Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1990; McClelland & Elman, 1986; Norris, 1994). In these models, transposed-phoneme nonwords and substituted-phoneme nonwords would produce similar levels of activation in the lexical representation associated with the base word, and transposed-phoneme nonwords should therefore not be perceived as more similar to their base words than the substituted-phoneme nonwords. Our results show that this was clearly not the case, since when the transposed phonemes were adjacent, transposed-phoneme nonwords led to more priming on the one hand, and to longer "no" lexical decisions on the other hand, than substituted-phoneme nonwords. One possible way to reconcile transposed-phoneme effects with these models 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), and their account of transposed-letter effects. For example, in Gomez et al. (2008)'s model, the representation of one letter is not strictly tied to a single letter position, but each letter in a letter string creates a distribution of activation over positions so that the representation of one letter extends into nearby letter positions. Incorporating such a mechanism in models like TRACE (McClelland & Elman, 1986) would allow the representation of one phoneme at a given position to be activated by the presence of that phoneme in adjacent positions. As a result, the word */bi**sk**ɔt/* would receive more activation from the transposed-phoneme nonwords /bi**ks**ɔt/ than from the phonological control nonword /bi**pf**ɔt/, thus accounting for the transposed-phoneme effect.

To sum-up, in the present work we have shown that transposed-phoneme effects are not restricted to short, monosyllabic, words but are also observed in longer, bisyllabic, words. More importantly, we have provided evidence that transposed-phonemes nonwords are perceived as being more similar to their base words than substituted-phoneme nonwords.

These results were predicted by the TISK model (Hannagan et al., 2013) that incorporates flexibility in the way in which phoneme order information is encoded. In addition, our results point to a possible role for syllable boundaries in driving transposed-phoneme effects, with current evidence suggesting that these effects are more likely to emerge when the transposed phonemes belong to the same syllable. More generally, the present study contributes to the old, but still current, debate concerning the nature of the units on which pre-lexical processing of spoken words is based. Although the psychological reality of phonemes has been repeatedly challenged (see Kazanina et al., 2018 for a review), and more recently the existence of linguistic units in general (Samuel, 2020), the present study provides a clear demonstration that pre-lexical linguistically defined units are extracted from the speech signal, and in particular that position-independent phonemes (see also Bowers et al., 2016) play a role in spoken word recognition. Here, we would argue that there is not a single such pre-lexical unit mediating spoken word recognition, but rather a combination of position-independent phonemes and a mechanism for flexibly encoding the order of phonemes that enables access to higher-order syllabic and word-level representations.

Acknowledgements

This research was supported by ERC grant 742141. The authors thank Julie Pynte for her assistance in running participants.

References

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/.

Bowers, J. S., Davis, C. J., Mattys, S. L., Damian, M. F., & Hanley, D. (2009). The activation of embedded words in spoken word identification is robust but constrained: Evidence from the picture–word interference paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1585–1597.

Bowers, J. S., Kazanina, N., & Andermane, N. (2016). Spoken word identification involves accessing position invariant phoneme representations. *Journal of Memory and Language*, *87*, 71–83.

Dahan, D., Magnuson, J.S., & Tanenhaus, M.K. (2001). Time course of frequency effects in spoken word recognition: evidence from eye movements. *Cognitive Psychology*, *42*, 317-367.

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

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

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association, 105*(2), 203–206.https://doi.org/10.5195/JMLA.2017.88.

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.

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.

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

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., & Grainger, J. (2012). Protein analysis meets visual word recognition: A case for String Kernels in the brain. *Cognitive Science, 36,* 575–606.

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.

Kazanina, N., Bowers, J.S., & Idsardi, W. (2018). Phonemes: lexical access and beyond. *Psychonomic Bulletin & Review*, 25, 560–585.

Marcet, A., Perea, M., Baciero, A., & Gomez, P. (2019). Can letter position encoding be modified by visual perceptual elements? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 72,* 1344–1353.

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.

Perea, M., & Duñabeitia, J. A., & Carreiras, M. (2008). Transposed-letter priming effects for close versus distant transpositions. *Experimental Psychology*, *55*,397– 406.

Perea, M., García-Chamorro, C., Martín-Suesta, M., & Gomez,P. (2012). Letter position coding across modalities: The case of Braille readers. *PLoS ONE*, *7*(10), e45636.

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.

Samuel, A. G. (2020). Psycholinguists should resist the allure of linguistic units as perceptual units. *Journal of Memory and Language, 111*, 104070.

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.

You, H. & Magnuson, J. (2018). TISK 1.0: An easy-to-use Python implementation of the time-invariant string kernel model of spoken word recognition. *Behavior Research Methods, 50*, 871-889.

Table 1: Characteristics of the stimulus sets (mean values)

	Frequency[1]	Syllable number	Phoneme number	Uniqueness point[2]	Duration[3]
Adjacent					
Target/repeated words primes (/*bi**sk**ɔt*/)	15	2	5.48	5.70	630
Transposed nonword primes (/bi**ks**ɔt/)	-	2	5.48	-	631
Substituted nonword primes (/bi**pf**ɔt/)	-	2	5.48	-	629
Nonadjacent					
Target/repeated word primes (/*ʃo**k**o**l**a*/)	14	3	6.06	6.07	616
Transposed nonword primes (/ʃo**l**o**k**a/)	-	3	6.06	-	615
Substituted nonword primes (/ʃo**r**o**p**a/)	-	3	6.06	-	615

Note: [1] In number of occurrences per million. [2] The phonemic position at which the auditory target words can be reliably identified. [3] In milliseconds.

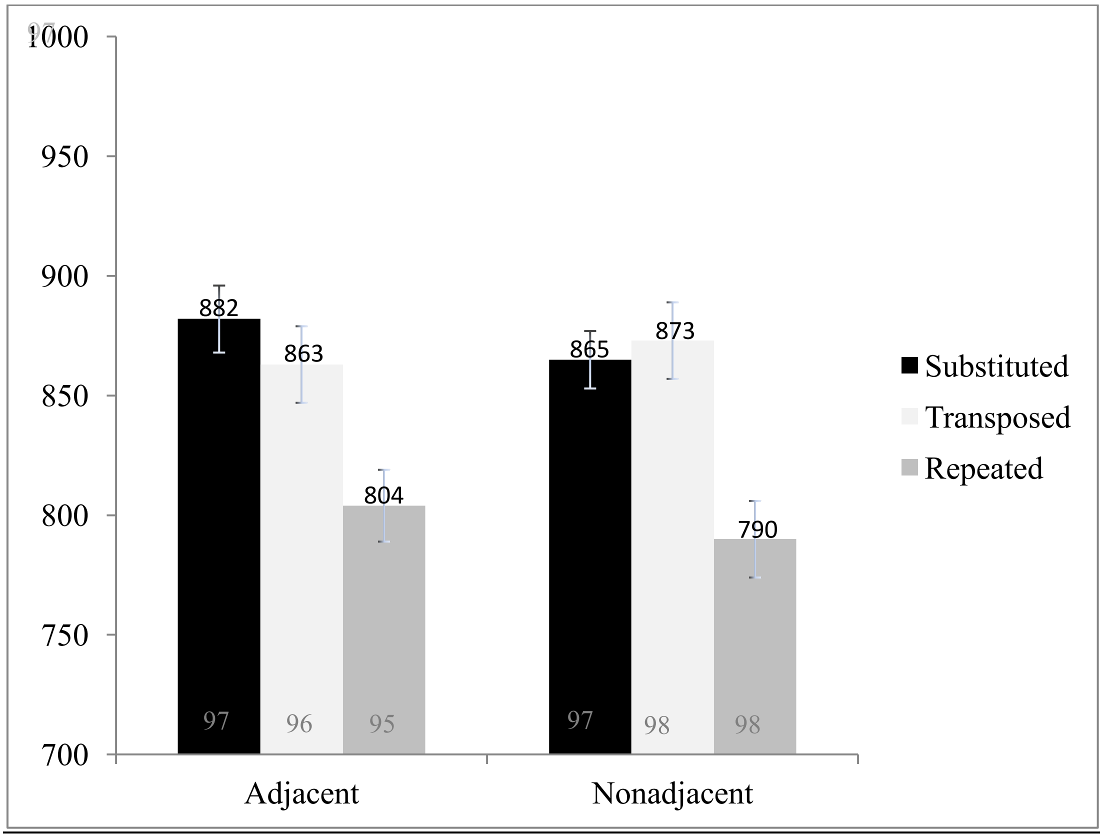

Figure 1: Mean Reaction Times (in ms) in each condition of Experiment 1. Percentages of correct responses are shown at the bottom of the graph. Error bars are standard errors.

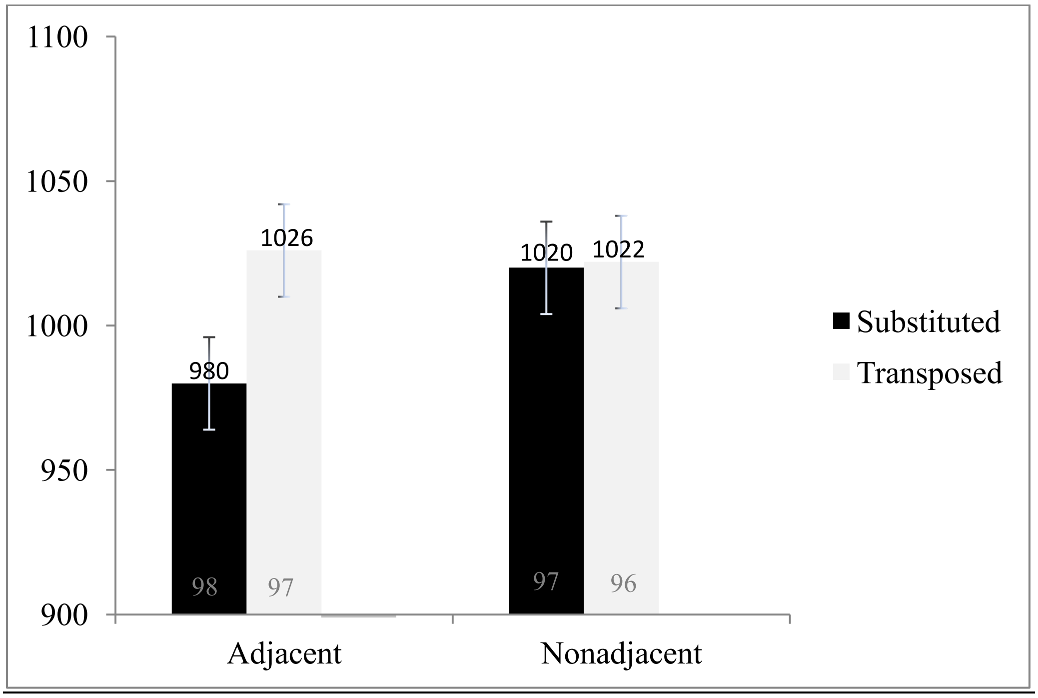

Figure 2: Mean Reaction Times (in ms) in each condition of Experiment 2. Percentages of correct responses are shown at the bottom of the graph. Error bars are standard errors.

Appendix 1: Nonwords and words used in Experiments 1 and 2.

Adjacent phonemes			**Nondjacent phonemes**		
Substitued nonwords	**Transposed nonwords**	**Repeated primes/ Target words**	**Substitued nonwords**	**Transposed nonwords**	**Repeated primes/ Target words**
bipfotte	bicsotte	biscotte	chakilé	chatiré	charité
bourpette	boulkette	bouclette	bouvanrer	boujanler	boulanger
bochteur	boskeur	boxeur	camadon	canabon	cabanon
capraire	caclaire	calcaire	syntigat	synkidat	syndicat
canrant	camlant	calmant	catamé	capané	canapé
caltice	carpice	caprice	calami	carani	canari
caglone	cabrone	carbone	camijo	caniso	casino
cablo	cagro	cargo	ceinlupon	ceinruton	ceinturon
caklette	caprette	carpette	chanureau	chamuleau	chalumeau
caplon	catron	carton	chakibeau	chatipeau	chapiteau
caplouche	catrouche	cartouche	choropat	cholocat	chocolat
capfade	cacsade	cascade	cibapin	cidatin	citadin
catchette	caksette	casquette	mikirant	mitilant	militant
capchor	catsor	castor	compadant	comtabant	combattant
chalbin	charguin	chagrin	cobénie	codémie	comédie
chaglon	chabron	charbon	coquiné	cotimé	comité
cilpon	cirton	citron	couvaleux	coujareux	courageux
conrsit	conlfit	conflit	dépudant	détubant	débutant
conlpat	conrtat	contrat	dériché	délifé	défilé
couvlette	coujrette	courgette	détinrant	déquinlant	délinquant
depchin	detsin	destin	défapant	déchatant	détachant
dipchours	dicsours	discours	famileux	fanireux	farineux
dikfute	dipsute	dispute	falozi	farovi	favori
dopgeur	dotkeur	docteur	galadit	garabit	gabarit
dourgure	doulbure	doublure	gapanlie	gatanrie	garantie
fapgeur	fatkeur	facteur	grarumé	gralуné	granulé
fagleau	fadreau	fardeau	javourie	jasoulie	jalousie
fekchin	fetsin	festin	kanloubou	kanrougou	kangourou
fipchon	fitson	fiston	ladazo	labavo	lavabo
foublon	fougron	fourgon	malaton	maracon	macaron
founli	foumri	fourmi	majabin	masaguin	magasin
goulbon	gourdon	goudron	mabarie	madalie	maladie
gounland	goumrand	gourmand	mapalon	matharon	marathon
jablin	jadrin	jardin	méborie	médolie	mélodie
jadlon	jagron	jargon	mochapin	mossaquin	mocassin
laglon	ladron	lardon	nuléno	nurémo	numéro
makleau	matreau	marteau	panrakon	panlaton	pantalon
mapchotte	macsotte	mascotte	paguolie	padorie	parodie

milpo	mirco	micro	pétiran	pékilan	pélican
miltobe	mircobe	microbe	povaker	pojater	potager
mildaine	mirgaine	migraine	pramiré	pranilé	praliné
nomldil	nomrbil	nombril	pélumie	pérunie	pénurie
paglon	padron	pardon	pynava	pymaja	pyjama
paslun	pafrum	parfum	ripovo	ritoso	risotto
pikfon	pitson	piston	salachi	sarafi	safari
salchan	sarfan	safran	sabari	samali	salami
sanrdot	sanlgot	sanglot	stirunant	stilumant	stimulant
sagline	sadrine	sardine	tamloudin	tamroubin	tambourin
tardette	talbette	tablette	tomrogua	tomloba	tombola
tapline	tatrine	tartine	vadakond	vabaguond	vagabond
toslon	tochron	torchon	vepouré	vetoulé	velouté
tuglan	tubran	turban	vébanla	védanra	véranda
vipgime	vitkime	victime	vépilé	vétiré	vérité
viblule	vigrule	virgule	rénuvé	rémusé	résumé

Appendix 2: Summary of the mixed effects models for Experiments 1 & 2

Table A1: Summary of the mixed effects model for Experiment 1. The intercept represents the target words preceded by substituted-phoneme nonword primes in the adjacent condition.

Effect	***β***	**SE**	**t**	**p**
(Intercept)	6.77	0.02	389.93	<.001
Repeated prime	-0.10	0.01	-7.50	<.001
Transposed prime	-0.02	0.01	-2.43	<.05
Nonadjacent	-0.01	0.02	-0.99	>.20
Repeated prime : Nonadjacent	-0.01	0.01	-0.53	>.20
Transposed prime : Nonadjacent	0.03	0.01	2.49	<.05

Table A2: Summary of the mixed effects model for Experiment 1 after releveling. The intercept represents the target words preceded by substituted-phoneme nonword primes in the nonadjacent condition.

Effect	***β***	**SE**	**t**	**p**
(Intercept)	6.75	0.02	386.56	<.001
Repeated prime	-0.10	0.01	-8.17	<.001
Transposed prime	0.01	0.01	0.77	>.20
Adjacent	0.01	0.02	0.99	>.20
Repeated prime : Adjacent	0.01	0.01	0.53	>.20
Transposed prime : Adjacent	-0.03	0.01	-2.49	<.05

Appendix 2 continuation

Table B1: Summary of the mixed effects model for Experiment 2. The intercept represents the substituted-phoneme nonwords in the adjacent condition.

Effect	***β***	**SE**	**t**	**p**
(Intercept)	6.87	0.02	389.33	<.001
Transposed nonword	0.04	0.01	2.95	<.01
Nonadjacent	0.04	0.01	2.66	<.01
Transposed nonword : Nonadjacent	-0.04	0.02	-1.998	<.05

Table B2: Summary of the mixed effects model for Experiment 2 after releveling. The intercept represents the substituted-phoneme nonwords in the nonadjacent condition.

Effect	***β***	**SE**	**t**	**p**
(Intercept)	6.91	0.02	406.78	<.001
Transposed nonword	0.002	0.01	0.14	>.20
Adjacent	-0.04	0.01	-2.66	<.01
Transposed nonword : Adjacent	0.04	0.02	1.998	<.05