Another cup of TEE? The processing of second language near-cognates in first language reading

Kristin Lemhöfer, Lynn Huestegge & Kimberley Mulder

View supplementary material

Published online: 19 Feb 2018.

Submit your article to this journal

Article views: 619

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

REGULAR ARTICLE

OPEN ACCESS | Check for updates

Another cup of TEE? The processing of second language near-cognates in first language reading*

Kristin Lemhöfer[a], Lynn Huestegge[b] and Kimberley Mulder[c]

[a]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands; [b]Institute of Psychology, University of Würzburg, Würzburg, Germany; [c]Centre for Language Studies, Radboud University, Nijmegen, The Netherlands

ABSTRACT
A still unresolved issue is in how far native language (L1) processing in bilinguals is influenced by the second language (L2). We investigated this in two word recognition experiments in L1, using homophonic near-cognates that are spelled in L2. In a German lexical decision task (Experiment 1), German-Dutch bilinguals had more difficulties to reject these Dutch-spelled near-cognates than other misspellings, while this was not the case for non-Dutch speaking Germans. In Experiment 2, the same materials were embedded in German sentences. Analyses of eye movements during reading showed that only non-Dutch speaking Germans, but not Dutch-speaking participants were slowed down by the Dutch cognate misspellings. Additionally, in both experiments, bilinguals with larger vocabulary sizes in Dutch tended to show larger near-cognate effects. Thus, Dutch word knowledge influenced word recognition in L1 German in both task contexts, suggesting that L1 word recognition in bilinguals is non-selective with respect to L2.

Introduction

Since the early beginnings of psycholinguistic research on the structure of the bilingual lexicon, the predominant question has been whether and to what extent the two language systems can function independently of each other (e.g. Beauvillain & Grainger, 1987; de Groot, Delmaar, & Lupker, 2000; Kolers, 1963; Macnamara & Kushnir, 1971; Starreveld, de Groot, Rossmark, & Van Hell, 2014). In this context, there is one class of words of which the processing has been studied particularly intensively, namely, cognates. Cognates are words that share form and meaning between two languages, such as the word *rose* in French and English. Bilinguals process cognates faster, more accurately and with less difficulty than non-cognates in almost all language tasks we know of, both in language production (Costa, Santesteban, & Caño, 2005; Hoshino & Kroll, 2008; Pureza, Soares, & Comesaña, 2016; Strijkers, Costa, & Thierry, 2010) and in comprehension (Bultena, Dijkstra, & van Hell, 2013; Lemhöfer & Dijkstra, 2004; Lemhöfer, Dijkstra, et al., 2008; Midgley, Holcomb, & Grainger, 2011). The fact that even in a monolingual word processing task, words that also exist in the currently inactive language are processed differently from words that do not, has been taken as evidence for the so-called *non-selective account* of bilingual language processing, that is, the view that one language processing system does not normally function completely independently of the other. In consequence, all present models of bilingual language processing include cross-language interactions at the lexical level (de Bot, 2004; Dijkstra & van Heuven, 1998, 2002; French, 1998; Green, 1998; Zhao & Li, 2010). Cognate effects can then be explained by word activation levels for cognates that are higher than those for non-cognate words because they accumulate across languages, either on the feature level, the word form level, and/or on the semantic level (recall that cognates have the same meaning in both languages; see the General Discussion for more details).

However, there also appear to be limits to the extent of the cross-talk between words from both languages. For instance, cognate and other cross-language effects are not symmetrical across a bilingual's two languages: They are most reliably found for word recognition tasks carried out in the second language (L2), while effects in the reverse direction, i.e. from L2 on first language (L1) processing, are usually smaller and/or less robust (Caramazza & Brones, 1979; Davis et al., 2010; de Groot,

Borgwaldt, Bos, & van den Eijnden, 2002; Gollan, Forster, & Frost, 1997; Poarch & van Hell, 2012). Incorporating this asymmetry into the non-selective view of the function of the bilingual lexicon is not trivial and has been achieved in various ways, e.g. by assuming lower resting levels and therefore temporally delayed activation of L2 lexical representations (Dijkstra & van Heuven, 2002; van Heuven, Dijkstra, & Grainger, 1998), or by assigning the L1, but not the L2 the status of a "base language" that can never fully be deactivated (Grosjean, 1998).

A second limitation of non-selectivity of bilingual lexical access seems to arise as soon as words are embedded in sentences rather than presented in isolation. When bilinguals read sentences in their second language, cognate effects can disappear or become substantially smaller, for example when the cognate words are predictable from the sentence context (Libben & Titone, 2009; Schwartz & Kroll, 2006; van Hell & de Groot, 2008), when the cognates are in mid-sentence (rather than final) position (Macizo & Bajo, 2006), for verb rather than noun cognates (Bultena, Dijkstra, & van Hell, 2014), or when the purpose of reading is mere repetition (Macizo & Bajo, 2006). Another relevant factor seems to be L2 proficiency: less skilled L2 speakers are more susceptible to cognate effects in L2 than very proficient ones (Bultena et al., 2014; Libben & Titone, 2009). Finally, the degree of overlap between the two cognate readings matters with respect to how they are processed: orthographically non-identical cognates (so-called *near-cognates*) generally give rise to (sometimes disproportionally) smaller cognate effects than identical ones do (Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010; Duyck, van Assche, Drieghe, & Hartsuiker, 2007).

Taken together, these studies suggest that the cognate effect as the most robust indicator of non-selectivity of the bilingual lexicon is not as universal as one might think. Rather than being language-selective or non-selective on principle, the bilingual word processing system appears to be flexible in order to allow for higher or lower degrees of between-language cross-talk, depending on the factors mentioned above. In particular, the extent of between-language interactions seems to decrease when the target language is L1 rather than L2, and when words are not processed in isolation, but in sentence context. Thus, investigating cognate / near-cognate effects in bilinguals while they are *reading sentences in their L1* seems to be a good testing case to explore the lower limits of language-nonselectivity, i.e. whether the degree of cross-language co-activation can be reduced to zero.

To our knowledge, there are three previous studies that have examined cognate effects during L1 sentence processing in bilinguals. These are all eye movement studies of sentence reading in bilinguals, where reading times for cognates are compared to those of matched control words. Even though all three of these studies succeed in demonstrating (facilitatory) cognate effects, the evidence remains somewhat feeble due to the extremely small observed effect sizes. van Assche, Duyck, Hartsuiker, and Diependaele (2009) found that (identical and non-identical) cognates embedded in Dutch sentences read by Dutch-English bilinguals yielded first fixation durations that were 5 ms shorter than those for control words; the effects for gaze and regression path durations were 8 and 10 ms, respectively, and became significant only with (despite multiple comparisons on interdependent measures) non-corrected p-levels of .05 and extremely large degrees of freedom (due to analyses on non-aggregated data). Titone, Libben, Mercier, Whitford, and Pivneva (2011) found cognate facilitation effects of similar sizes in English-French bilinguals reading English sentences, but these occurred only in low-constraint sentences (i.e. when the target was not predictable) and, in the case of early measures, only for bilinguals who acquired their L2 during childhood. Finally, in a corpus reading study on the data of participants reading an entire book (Cop, Dirix, van Assche, Drieghe, & Duyck, 2017), cognate or cross-language orthographic similarity effects in L1 were extremely small as well and found only in few of the measures and in complex interactions with other variables (e.g. a 2 ms advantage in first fixation durations only for words longer than nine letters).

With the extremely small effect sizes reported in these studies and the rising awareness that such fragile effects are especially endangered to be affected by publication biases (Open Science Collaboration, 2015), there appears to be sufficient reason to examine the question of non-selectivity and cognate effects during L1 reading again, ideally from a different angle than from the one adopted in all three mentioned studies. This is what we set out to do in the present study. The new angle we chose was to look at the processing of *incorrect,* namely second-language spellings of cognates in the context of a first-language task (analogous to, e.g. the English spelling of *coffee* instead of the correct Dutch spelling *koffie* in a Dutch context). Additionally, given the cited previous reports of strong task-set dependency of (near-)cognate effects (e.g. purpose of reading, Macizo & Bajo, 2006), another new angle of our present study was to compare effects on sentence reading for comprehension with those in a classic single-word processing task, namely lexical decision, by using the same set of target items.

Looking at how "misleading foils" are processed follows an established research tradition in psycholinguistics. For instance, in Van Orden's (Van Orden, 1987) famous study, participants were found to often erroneously classify *rows* as a flower, misled by the word's shared phonology with *rose* (for other word recognition research on "misleading" items, see, e.g. Besner & Davelaar, 1983; Forster & Hector, 2002; Pecher, 2001). In the field of bilingual word recognition, the same logic has been applied to monolingual lexical decision tasks in which words of the other language occurred among the foils requiring a "no"-response (Gerard & Scarborough, 1989; Nas, 1983; Rodriguez-Fornells, Rotte, Heinze, Nösselt, & Münte, 2002; Scarborough, Gerard, & Cortese, 1984). The idea here is that if lexical access was truly language-selective, responding "no" to words from the non-target language should be as easy as saying "no" to other nonwords.

Here, we apply a similar principle. To find out in how far native language word recognition is immune to influence from the weaker L2, we will look at how German speakers of Dutch who are reading in their L1 (German) process homophonic, but orthographically non-identical cognate words (co-called *near-cognates*) that are spelled *in the Dutch way* (e.g. STOK "stick" instead of the correct German spelling STOCK). We will use two highly divergent task sets that are both commonly used in bilingual word recognition research, namely lexical decision (Experiment 1) as a starting point, and sentence reading for comprehension (Experiment 2) using the same set of target items.

Thus, what happens when German speakers of Dutch who read in their L1 German encounter Dutch spellings such as STOK instead of the correct German variant STOCK? One possibility is that they, when reading in their dominant L1, are hardly influenced by their L2 Dutch at all, especially when they did not acquire this L2 before adulthood. In that case, the absence of cross-language activation would contradict theories which claim that the bilingual language system is principally and always non-selective, i.e. that there is always an influence of the non-target language. Consequently, Dutch-speaking German participants might not be any more inclined to confuse the Dutch spelling STOK with the correct German one (STOCK) than they would be for other types of "normal", non-Dutch misspellings (like TRIK instead of TRICK "trick"). Given the weaker evidence for cognate effects in sentence reading as compared to single word recognition, this might especially be the case in Experiment 2. This would be in line with the extremely small size of the cognate effects observed in L1 sentence reading before, as reviewed above. It would also be in line with modelling accounts that assume

comparatively low activation levels of L2 when it is not selected (like, for instance, when placed in a "monolingual mode"; Grosjean, 1998) or when other language cues, such as language-specific orthography, reduce language co-activation (Casaponsa & Duñabeitia, 2016).

In contrast, if knowledge of the fairly recently acquired language Dutch does already affect word recognition in the participants' native language, the Dutch spelling STOK might not easily be identified as wrong in German. STOK and STOCK share the same phonology, meaning, and a large part of their orthography, which might complicate the German / non-German distinction. This should have different consequences depending on the task: In a task where the exact identification of the correct spelling (and its discrimination from the other, currently incorrect one) is required, such "Dutch misspellings" would be easily confused with the correct German ones, and in particular, more often than other misspellings (cognate inhibition effect). On the other hand, in tasks focusing on word *meaning* rather than on spelling, the processing of the incorrect spelling STOK (as a misspelling of STOCK) should be *less* difficult than that of other misspellings, because an exact discrimination of the two forms is unnecessary for understanding the meaning of the sentence. Rather, the fact that STOK is already represented in the bilingual lexicon with the same meaning as STOCK would be advantageous because it would already activate the correct meaning (cognate facilitation effect). In any case, no matter whether facilitatory or inhibitory, the existence of near-cognate effects in a situation that does not favour such effects to arise (late bilinguals in an L1 task) would be an argument for the bilingual word recognition system to be fundamentally non-selective with respect to language.

We investigated the language-selective vs. non-selective nature of near-cognate processing in two experiments. Because it was our aim to explore the lower limits of between-language cross-talk (non-selectivity), both experiments were conducted with native speakers of German who acquired Dutch as a foreign language relatively recently and as adults, namely, before the start of their university programme. The first experiment was a German lexical decision task in which Dutch spellings of homophonic near-cognates such as STOK were included in the list, and required a "no" response. As mentioned before, the inclusion of non-target language words as "nonwords" in the lexical decision task is a common method to investigate language selectivity in bilingual word recognition, but it has not yet been used in connection with cognate spellings. Such near-cognates, however, have been used as targets in a language decision task by Dijkstra et al. (2010), who

observed *inhibition* for near-cognates, i.e. longer RTs for words whose translation is very similar in the other language. Nevertheless, in that study, the task (language decision) was different from ours (lexical decision). It is thus unclear whether we will also observe a (near-)cognate inhibition effect in the present study. If our learners of Dutch were influenced by their L2 during L1 (German) word recognition, they should reject Dutch near-cognate spellings like STOK less accurately and/or more slowly than other misspellings, i.e. we should observe *cognate inhibition*. If, on the other hand, our participants' L1 lexicon was immune to L2 interference (due to some sort of "protected" nature of L1, the late and recent L2 acquisition, or similar reasons), they should not show such differences.

In the second experiment, we used the same "misspelled" items embedded in German sentences. German learners of Dutch read these sentences on the screen while their eye movements were measured. Previous monolingual research on the processing of typographical errors has shown that processing times for misspelled words are typically prolonged (White & Liversedge, 2004). Unlike lexical decision, reading for comprehension does *not* require a precise identification of the correct German spelling. Reading a misspelled word that is already known from Dutch and linked to the correct target meaning (STOK), should then be easier than reading a word that is misspelled in a "normal" way (TRIK). Thus, if word recognition during L1 sentence reading is influenced by L2, bilingual readers should exhibit cognate *facilitation* effects in this task.

As critical near-cognate items, we used German-Dutch translation pairs with identical or close-to-identical phonology in the two languages and a highly similar, but non-identical orthography, like *stok-Stock*. Reading performance on these items was, first, to be compared to that of "ordinary", non-Dutch misspellings of existing German words like TRIK instead of TRICK (so-called *pseudohomophones*). However, it has to be noted that such pseudohomophones differ from our Dutch near-cognate spellings (STOK) in *two* respects: first, they are non-cognates, i.e. they do not represent the Dutch translation of the respective German target word (in this case, TRUC); and second, they do not represent *any* existing Dutch word at all, therefore, they do not have a lexical representation and are thus orthographically unfamiliar even to someone who knows Dutch. In consequence, any potential difference between how German speakers of Dutch process items like STOK and those like TRIK might arise out of any, or both, of these two differences. In particular, orthographic familiarity has been shown to influence word recognition both in single word

processing (Zeelenberg, Wagenmakers, & Shiffrin, 2004) and in sentence reading (White & Liversedge, 2004). Therefore, if interested in the genuine cognate effect, i.e. the effect of replacing words with their highly form-similar translations, it seems desirable to have an additional control condition of German pseudohomophones that *do* represent Dutch words (and are thus orthographically familiar to L2 speakers of Dutch), but that are non-cognates (henceforth called *Dutch non-cognate homophones*). Even though such Dutch words are relatively rare, they do exist, like the Dutch word HEK: it sounds like German HECK ("rear" of a car, ship, etc.), but does not have the same meaning (it means "fence", in German: "Zaun"). Any difference between how German speakers of Dutch process these two nonword types (HEK vs. STOK) should be due to the special cognate status of STOK with highly overlapping (form and meaning) representations, and not to its higher orthographic familiarity or its lexical status in Dutch.

Because the matching of the three critical item categories (STOK; TRIK; HEK) might not be possible in a perfect way, especially with this highly restricted item pool, we decided to adopt a more complex study design which does not rely on between-item comparisons only, but rather on a comparison between a group of Germans with Dutch as a second language (henceforth called "Dutch-speaking" participants, in short DS) and *an additional baseline group* of German speakers without knowledge of Dutch (henceforth called "non-Dutch speaking" participants, abbreviated NDS). In contrast to the DS participants, these NDS participants should not be influenced by the cognate status of Dutch near-cognates like STOK, and process them in the same way as other misspellings, like HEK or TRIK. As a result, our crucial analyses do not hinge on the comparison of performance between item types (Dutch near-cognates vs. Dutch non-cognate homophones vs. German pseudohomophones), but rather on the potential interaction between item type and group (DS vs. NDS individuals).

Experiment 1: lexical decision

Experiment 1 was a lexical decision task in German, the participants' native language, in which the critical conditions contained the nonword items requiring a "no"-response: Dutch spellings of near-cognates (e.g. STOK) were compared to Dutch spellings of Dutch-German non-cognate homophones (HEK) and (non-Dutch) German pseudohomophones (TRIK). Two groups of native speakers of German were tested, those with (DS) and without (NDS) knowledge of Dutch. Thus, we are

looking for an interaction of item category by speaker group.

To avoid that Dutch-speaking participants felt that all nonwords in the experiment were misspellings of existing German words, we also included "normal" German-like pseudowords (e.g. WULG) as fillers. Furthermore, we included word items with *correct* German spellings of near-cognates (e.g. BLICK, Dutch *blik*) and non-cognate control words to assess whether there is a "standard" cognate effect in the word data.

Method

Participants

Twenty-five native speakers of German (six men), all without (known) dyslexia, studying at Radboud University Nijmegen, the Netherlands, formed the DS group. They all had acquired Dutch in an intensive course preparing them for their Dutch exam required for their study program at Radboud University Nijmegen at the age of around 20, apart from one participant who had learned Dutch at high school from the age of 17 onwards. They were between 20 and 31 years old (mean = 23.6 years), with between one and five years of experience with Dutch (mean = 3.4 years). In a language background questionnaire, they rated how often they used Dutch and how much experience they had with it. The results of these ratings, given on a scale from one to seven, are summarised in Table 1. Participants were also tested for their Dutch vocabulary knowledge using the Dutch version of LexTALE, a short non-speeded lexical decision task on pseudowords and infrequent words (which is validated only for English so far; Lemhöfer & Broersma, 2012). The mean score (a percentage-correct measure adjusted for yes- or no-bias) was 72.6 (minimum 55.0, maximum 85.5, SD = 8.2), which, in English, would roughly correspond to the Common European Framework of Reference for Languages (CEF) proficiency level B2 (Upper intermediate).

Table 1. Self-ratings from the language background questionnaire given by German speakers of Dutch in Experiment 1.

	Mean	SD	Range
How often do you read Dutch literature?	5.1	1.6	2–7
How often do you speak Dutch?	5.8	1.3	2–7
How often do you listen to Dutch radio / watch Dutch TV?	4.4	1.7	1–7
Self-rated reading experience in Dutch	5.3	1.0	3–7
Self-rated writing experience in Dutch	5.0	1.0	3–7
Self-rated speaking experience in Dutch	5.4	1.1	4–7
Mean Dutch experience (mean of previous 3)	5.3	0.8	3.7–7.0

Note. SD = Standard deviation.
[a]Self-ratings were given on a scale from 1 (low / rarely) to 7 (high / very often).

The Dutch-speaking participants also reported to know other foreign languages, particularly English (all participants), which, on average, they reported to use with a frequency of 5.0 (on a 7-point scale; SD = 1.3, range 3–7). Twenty-one of the participants stated however that they used Dutch more often than English, two stated the opposite, and for one participant, the frequency of use of these two languages was equal. Besides English, 22 participants indicated to speak yet another foreign language, like French (12), Spanish (4), or Polish (2), but none of them reported to speak this language more often than Dutch.

Because there are hardly any native speakers of German at Nijmegen University who do not speak (at least some) Dutch, the non-Dutch speaking control group was tested at RWTH University, Aachen, Germany. They were 23 (11 male) non-dyslexic participants with German as their only mother tongue, mostly students, and aged between 19 and 28 years (mean: 23.4). They all stated to speak English as a foreign language; 17 reported to speak one or even two additional foreign languages, most often French (13) or Spanish (7). The most frequently used foreign language (English in 19 cases) was spoken with an average rated frequency (7-point scale) of 5.0 (SD = 1.7, range 2–7). None of the participants in this group reported to know any Dutch.

Materials

Nonwords. The three critical conditions in the experiment were all formed by items requiring a "no"-response, namely, Dutch spellings of near-cognates (STOK), Dutch spellings of non-cognate cross-language homophones (HEK), and German (non-Dutch) pseudohomophones (TRIK). We also included other German-like pseudowords (e.g. WULG) as fillers.

The stimuli from the mentioned categories (20 items per category) were selected such that they did not contain any letters or letter combinations that can serve as clear orthographic cues for either German or Dutch language membership (e.g. German Umlauts, or the frequent Dutch bigram "ij"). For the categories of Dutch-German near-cognates and non-cognate homophones, we selected only Dutch words that are pronounced in the same (or highly similar) way in German, and for which the grapheme-to-phoneme conversion rules of the two languages overlap. Thus, for instance, the homophonic near-cognate BLOED(D)-BLUT(G) ("blood") was not selected because the bigram "oe" is mapped to a different sound in German (/ø:/) than in Dutch (/u:/). We also constructed twenty German pseudohomophones that were not identical to Dutch words. This was done by misspelling existing German words in a way that

resembled the orthographic changes between Dutch and German in the other two categories (e.g. TRIK, changed from German TRICK "trick").

These nonword categories were matched item-by-item for length in letters, (logarithmic and absolute) frequency of both the Dutch (if existing) and the German reading according to the SUBTLEX databases (Brysbaert et al., 2011; Keuleers, Brysbaert, & New, 2010), as well as for Levenshtein distance (or edit distance) as a measure of orthographic similarity of the two readings (e.g. Schepens, Dijkstra, & Grootjen, 2012). Matching was confirmed by one-way ANOVA's (when all three categories were compared) or independent-sample t-tests (for Dutch frequencies that were only available for two of the three categories) with p-values above .11. The number and summed (type) frequency of neighbours in German, both calculated using the tool WordGen (Duyck, Desmet, Verbeke, & Brysbaert, 2004), are also reported. Even though they were not explicitly matched, there were no significant differences in terms of these two variables (both $p > .19$). Because of the close similarity of German and Dutch to English, it could not be avoided that two near-cognates and two non-cognate homophones were also existing English words. Similarly, 22 of the items (roughly equally distributed across the conditions) were near-homophones to existing English words (e.g. KOOR which sounds very similar to the English word "core"). The characteristics of these stimulus conditions are summarised in Table 2. The list of critical stimuli can be found in the Appendix.

We also included twenty German non-pseudohomophone pseudowords (e.g. WULG) that were also matched with the other nonword types, to provide a "baseline" for nonword rejection performance. We treat these pseudowords as fillers, but will report their response times and error rates to provide a full picture of nonword processing in this task. Additionally, to reduce the proportion of "Dutch" nonwords in the stimulus list further, we included 40 additional German-like pseudowords that were not explicitly matched to the other four nonword types. All these sixty non-

pseudohomophone pseudowords were taken from several sources. They did not exist in German, Dutch, or English, and were monosyllabic as well as orthographically legal and pronounceable in German. The total number of nonwords used in the experiment was 120 (20 near-cognates, non-cognate homophones, pseudo-homophones, and matched pseudowords, respectively; 40 filler pseudowords).

Words. The 120 words requiring a "yes"-response in the experiment were 20 Dutch-German near-cognates, different from those presented in their Dutch spelling above, in their *German* spelling (e.g. BLICK "gaze", Dutch translation BLIK), 20 non-cognate control words (e.g. WALD "forest", Dutch translation BOS), and 80 additional filler words (both cognates and non-cognates). The cognate and control words were included to assess whether there would be a cognate effect for near-cognates in the responses to words. They were matched to each other in terms of German (absolute and logarithmic) frequency according to SUBTLEX-DE (Brysbaert et al., 2011) and length. Their mean length was 4.0 letters for both cognates and control words; their mean frequency (per million occurrences) was 42.3 for near-cognates (logarithmic 1.22) and 28.4 for control words (logarithmic 1.20).

Procedure

Participants performed a German visual lexical decision task, i.e. they decided whether or not the visually presented stimulus was an existing German word by pressing a button corresponding to either the answer "yes" or "no". The experiment was controlled by the software package *Presentation* version 13.0 (Neurobehavioural Systems, www.nbs.com). The visual stimuli were presented in white, 24 point Arial capital letters on a dark grey background in the middle of the screen. German nouns are always written with a capital as first letter, while Dutch ones are not. To avoid this kind of language cue, we presented all nouns in capitals. Participants were tested individually in a soundproof room. They first read the German

Table 2. Characteristics of the four matched nonword conditions.

	Near-cognates (STOK)	Non-cognate homophones (HEK)	Pseudo-homophones (TRIK)	Pseudowords (WULG)
Mean no. of letters	3.95 (0.76)	3.95 (0.76)	4.35 (0.49)	4.15 (0.75)
Mean log. freq. of Dutch spelling	1.23 (0.49)	1.16 (0.60)	–	–
Mean absolute freq. of Dutch spelling (o.p.m.)	30.9 (46.6)	32.0 (52.4)	–	–
Mean log. freq. of correct German spelling	1.20 (0.48)	1.22 (0.66)	0.97 (0.49)	–
Mean absolute freq. of correct German spelling (o.p.m.)	25.9 (29.4)	62.5 (145.0)	15.6 (19.3)	–
Mean Levenshtein distance between actual and correct German spelling	1.15 (0.37)	1.15 (0.37)	1.20 (0.41)	–
Mean summed German bigram frequency	5939 (3284)	7483 (6169)	9139 (6554)	7587 (6100)
Mean no. of German orthographic neighbours	2.30 (2.32)	3.60 (3.44)	2.50 (1.76)	2.60 (2.14)

Note. Standard deviations are presented in parentheses.
o.p.m. = occurrences per 1 million words, freq. = frequency, log. = logarithmic.

instructions informing them about their task. They were asked to react as accurately and quickly as possible.

Each trial started with the presentation of a black fixation cross, which was displayed in the middle of the screen for 700 ms. After the presentation of an empty black screen for 300 ms, the target stimulus was presented. It remained on the screen until the participant responded or until a timeout of 1500 ms was reached. After another black screen had been shown for 500 ms, a new trial was started.

The experiment was divided into two blocks of equal length (120 trials). The first block was preceded by 20 practice trials with different stimuli than the experimental ones, but with a similar distribution of conditions as in the main experiment. After the practice trials, the participant could ask questions before continuing with the main experiment. Each participant received a different item order that was random with the only restriction that no more than three words or nonwords followed each other. Additionally, each block started with three dummy trials that were not analyzed. The complete experiment took approximately 10 min.

Results

Nonwords

We analyzed mean error rates and RTs in the three experimental nonword conditions Dutch near-cognates (STOK), Dutch non-cognate homophones (HEK), and German pseudohomophones (TRIK). One Dutch near-cognate (POOL, the Dutch spelling of German Pol "pole") had to be excluded because it is also a (correctly spelled) English loanword ("swimming pool"). For the analysis of RTs, incorrect trials were excluded. Furthermore, outliers that were more than two standard deviations away from the item and participants mean in a given condition

were removed from RT analyses (1.9% of correct trials in the DS group, and 2.4% in the NDS group).

The analyses were repeated-measures ANOVAs across participants (F_1) and items (F_2), with nonword category as within-participants, but between-items factor, and participant group (DS vs. NDS) as between-participants, but within-items factor. Degrees of freedom were Greenhouse-Geisser corrected where sphericity was violated.

The mean error rates and RTs for the three critical nonword categories are shown in Figures 1 and 2, and are numerically summarised in Table 3.[1]

The statistical results of the ANOVA's on error rates and RTs are summarised in Table 4. In the *error rates*, there was no effect of Nonword Type or of Group, but there was a highly significant interaction of the two, indicating that the three nonword categories were processed differently by the two groups of participants. Planned contrasts showed that this interaction was a consequence of different processing of the Dutch near-cognates, both relative to German pseudohomophones and to Dutch non-cognate homophones. In particular, subsequent t-tests showed that when comparing near-cognates to German pseudohomophones, NDS speakers showed an effect in the opposite direction than what was expected, with more errors for pseudohomophones than for near-cognates (t_1 (22) = −2.58, p = .017; t_2 (37) = −1.71, p = .096; marginally significant across items), while DS speakers displayed the reverse, but only marginally significant pattern (t_1 (24) = 2.03, p = .054; t_2 (37) = 1.85, p = .072). When applying the stricter control condition (Dutch homophones that also form Dutch words), there was no difference in error rates between near-cognates and non-cognate homophones in the NDS group (t_1 (22) = −1.58, p = .128; t_2 (37) = −1.00, p = .32), while DS participants made more errors on near-cognates than on homophones (t_1 (24) = 2.46, p = .021; t_1 (37) =
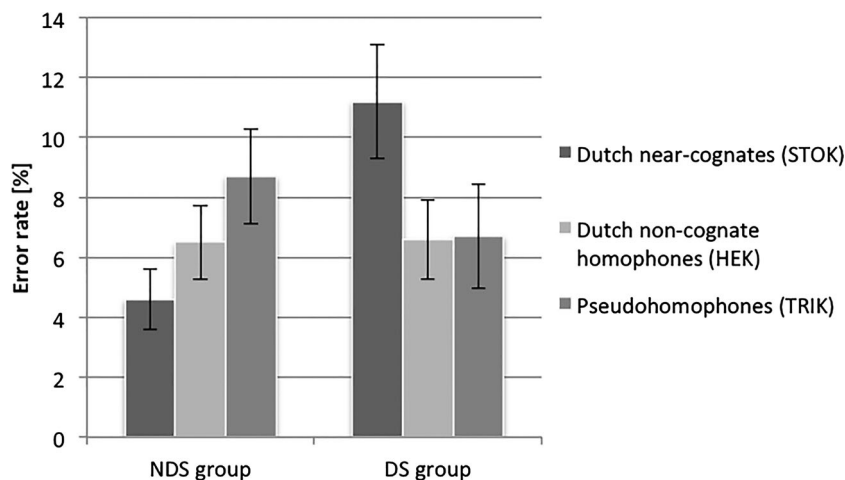

Figure 1. Mean error rates for the three nonword categories in the two groups. Error bars represent standard errors.
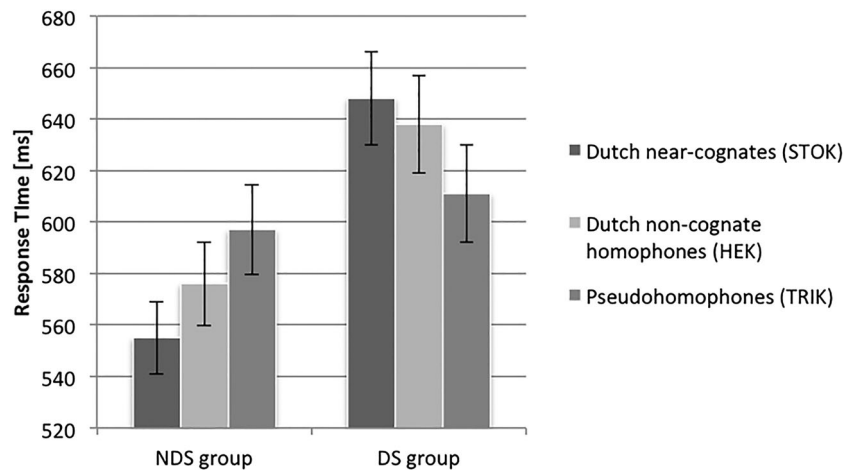
Figure 2. Mean RTs for the three nonword categories in the two groups. Error bars represent standard errors.

Table 3. Mean RTs and error rates (standard deviations in parentheses) for the three experimental nonword conditions, as well as the matched filler condition, in Experiment 1.

	NDS group		DS group	
Nonword Type	RT	Error rate	RT	Error rate
1 Dutch near-cognates (STOK)	555 (67)	4.6 (4.8)	648 (91)	11.2 (9.5)
2 Dutch non-cognate homophones (HEK)	576 (78)	6.5 (5.9)	638 (95)	6.6 (6.6)
3 Pseudohomophones (TRIK)	597 (83)	8.7 (7.6)	611 (94)	6.7 (8.7)
Near-cognates vs. pseudohomophones (row 1–row 3)	−42	−4.1	37	4.5
Near-cognates vs. non-cognate homophones (row 1–row 2)	−21	−1.9	10	4.6
Filler condition: Pseudowords (WULG)	589 (59)	3.9 (4.0)	599 (67)	4.2 (5.9)

1.78, $p = .083$; marginally significant across items). In summary, the DS speakers had most difficulty rejecting the Dutch-spelled near-cognates relative to the other kinds of misspellings, while NDS speakers, on the contrary, were *more* accurate on these near-cognates than on pseudohomophones, and equally accurate as on non-cognate homophones.

The ANOVA on *RTs* further confirmed the overall pattern observed in the error rates. Again, there was no main effect of Nonword Type, but a main effect of Group, with longer RTs for DS (632 ms) than for NDS participants (576 ms). Critically, the reversal of response time patterns visible from Figure 2 again surfaced in an interaction of Nonword Type by Group. It was highly significant for the comparison of near-cognates and pseudohomophones; for the contrast between near-cognates and homophones, it was significant across participants and marginally significant across items. Following up on this interaction, pairwise t-tests showed that NDS participants were *faster* rejecting Dutch near-cognates than German pseudohomophones (t_1 (22) = −5.35, $p = .000$; t_2 (37) = −2.40, $p = .022$), while DS participants showed the reverse

Table 4. Results of the ANOVA's for Error Rates and Reaction Times in Experiment 1.

		across participants				across items			
Measure	Effect	df	F	p	η_p^2	df	F	p	η_p^2
Error rates	Nonword Type	2;92	.76	.47	.02	2;56	.25	.78	.01
	Group	1;46	.88	.35	.02	1;56	1.97	.17	.03
	Nonword Type x Group	2;92	7.41	.002**	.14	2;56	6.26	.004**	.18
	Planned contrasts:								
	• near-cognates vs. PHPs	1;46	9.68	.003**	.17	1;37	10.11	.003**	.22
	• near-cognates vs. HPs	1;46	8.26	.006**	.15	1;37	6.81	.013**	.16
Reaction Times	Nonword Type	2;92	.44	.61	.01	2;56	.08	.93	.003
	Group	1;46	5.56	.02*	.11	1;56	92.03	.000**	.62
	Nonword Type x Group	2;92	23.66	.000**	.34	2;56	13.92	.000**	.33
	Planned contrasts:								
	• near-cognates vs. PHPs	1;46	33.60	.000**	.42	1;37	30.11	.000**	.45
	• near-cognates vs. HPs	1;46	6.75	.01*	.13	1;37	3.35	.075(*)	.08

Note: PHP = German pseudohomophones, HP = Dutch homophones.
Asterisks indicate levels of significance: **$p < .01$; *$p < .05$, (*)$p < .10$.

effect (t_1 (24) = 3.39, p = .002; t_2 (37) = 2.01, p = .051). When comparing Dutch near-cognates to homophones, these two conditions differed significantly for *NDS individuals* (but only over participants) in the opposite direction than what was expected for DS individuals, i.e. with faster RTs for near-cognates compared to homophones (t_1(22) = 2.97, p = .01; t_2(37) = 1.52, p = .14). This difference in the baseline is not entirely surprising, given the difficulties of matching highly constrained lexical items (see Introduction). In contrast, the difference between these two nonword types, though in the hypothesised direction descriptively, was not significant for DS participants (t_1(24) = 1.06, p = .30; t_2(37) = .53, p = .60). To summarise, as we will elaborate on in the Discussion, the important point in the RT data is that the hypothesised interaction was obtained in the expected direction (i.e. when regarding the NDS RTs as a baseline, relatively longer RTs for near-cognates than for other misspellings in the DS group).

Additional role of Dutch proficiency in DS participants

Having obtained Dutch vocabulary size measures for the DS speakers by using the Dutch version of LexTALE (Lemhöfer & Broersma, 2012), we carried out an additional analysis investigating whether the size of the near-cognate effect depended on vocabulary knowledge in Dutch. For this, we conducted ANCOVAs on the DS

participants' RTs and error rates with Nonword Type as a two-level factor only, including Dutch near-cognates and non-cognate homophones, and each participant's LexTALE score as a covariate. Because the LexTALE score is a participant characteristic, this analysis could only be conducted across participants. For RTs, there was no effect of Nonword Type, LexTALE score, or an interaction (all $F < 1$). However, for error rates, there was a significant main effect of Nonword Type (F_1 (1, 23) = 5.24, p = .032, η_p^2 = .185), and, crucially, an interaction with LexTALE score (F_1 (1, 23) = 6.80, p = .016, η_p^2 = .228). The main effect of LexTALE score was not significant ($F < 1$). Figure 3 shows a scatter plot illustrating the observed relation between vocabulary scores and the near-cognate effect in error rates (calculated by error rate for near-cognates minus that for homophones). The correlation of the two variables was r = .48. Thus, the larger a DS speaker's Dutch vocabulary was, the greater the near-cognate effect (i.e. the more errors she or he displayed on Dutch near-cognates compared to Dutch homophones).

Words. To assess whether a near-cognate effect was also present in the word data, we report also the means for (correctly spelled) near-cognates (like BLICK) and non-cognates. Outliers were excluded with the same criterion as for the nonwords (1.7% of correct responses in DS, and 1.6% in NDS participants). The mean RTs and error rates for these two word categories are shown in Table 5.
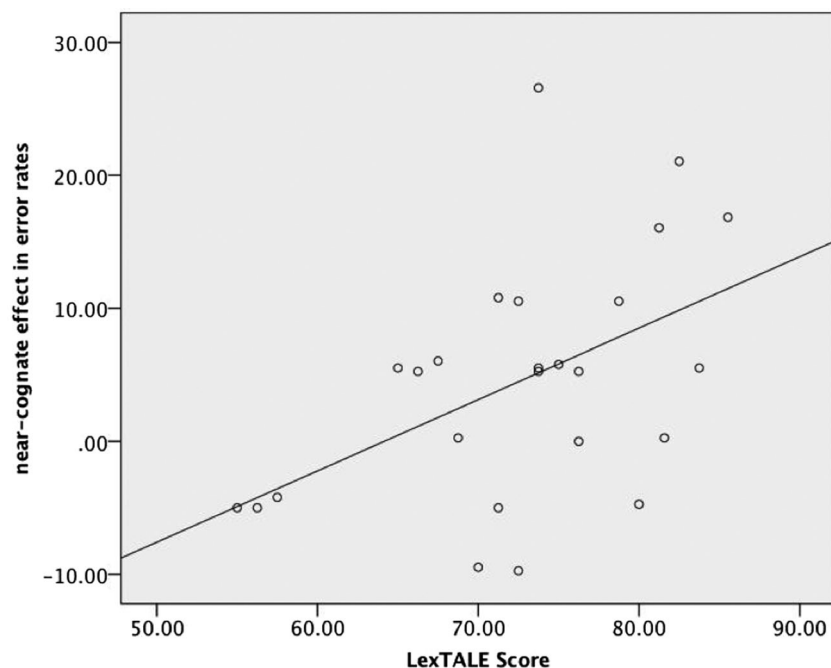

Figure 3. Scatter plot and regression line showing the correlation between the near-cognate effect (near-cognates minus homophones) in error rates and Dutch vocabulary size, as measured by LexTALE.

Table 5. Mean RTs and error rates (standard deviations in parentheses) for German near-cognates and non-cognates in Experiment 1.

	NDS group		DS group	
Word Type	RT	Error rate	RT	Error rate
German near-cognates (BLICK)	531 (65)	6.0 (5.8)	562 (71)	5.1 (5.4)
German controls (WALD)	529 (74)	7.6 (5.8)	563 (75)	6.4 (6.2)
Near-cognate effect (row 2– row 1)	−2	1.6	1	1.3

Repeated-measures ANOVAs with Word Type (near-cognates vs. non-cognates) as within-participants and between-items factor and Group as between-participants and within-items factor showed a trend (and a significant effect in the item analysis) towards faster overall RTs in the NDS group (530 vs. 562 ms; $F_1(1, 46) = 2.75$, $p = .10$, $\eta_p^2 = .056$; $F_2(1, 37) = 50.12$, $p < .001$, $\eta_p^2 = .575$), but no effect of Word Type on RTs (both $F < 1$), and no interaction between the two (both $F < 1$). For error rates, none of the main effects nor the interaction were significant (main effect Group: $F_1 < 1$; $F_2 (1, 37) = 1.33$, $p = .26$; main effect Word Type: $F_1(1, 46) = 2.43$, $p = .13$; $F_2 < 1$; Group x Word Type: both $F < 1$). Thus, there was no evidence whatsoever for a near-cognate effect in the word data.

Discussion

In the German (L1) lexical decision task in Experiment 1, we observed that native speakers of German with knowledge of Dutch had trouble rejecting near-cognates that were spelled in the Dutch way, like STOK (German spelling: STOCK), as compared to native German speakers who do not speak any Dutch. Our results were especially clear in the error rates, which is often the case in L2 research (e.g. Haigh & Jared, 2007; Ko, Wang, & Kim, 2011; Lemhöfer, Spalek, & Schriefers, 2008): DS participants made more errors on the near-cognates than on the other two types of misspellings, while NDS participants only showed a difference between near-cognates and pseudohomophones, but in the other direction (they were *less* accurate on pseudohomophones).

In the RT data, the critical Group by Nonword Type interaction was also significant in the hypothesised direction. Thus, relative to the non-cognate homophones and pseudohomophones, DS participants were slowed down more by nonwords that were near-cognates than NDS participants were. In fact, like in the error data, NDS participants were not slowed down by near-cognates at all, but were showing effects in the opposite direction (slower RTs for pseudohomophones than for near-cognates).

Altogether, the interaction of Group and Nonword Type in both measures represents a (near-)cognate effect in L1, and suggests the co-activation of Dutch lexical representations of near-cognates in a German task in DS speakers. This effect arose, first, with German pseudohomophones as a control condition in which nonwords did not represent Dutch words at all, suggesting that Dutch word status caused differences in how the two groups of participants processed the misspelled words. Importantly, though, they also arose with a very conservative control condition, i.e. Dutch homophones that did also represent Dutch words, but were co-activated probably to a lesser degree by the DS speakers because of lacking semantic overlap.

A closer inspection of the RT pattern within each group also reveals the anticipated difficulties associated with between-items comparisons despite careful matching (see Table 2). Specifically, we observed *faster* RTs (and slightly lower error rates) for near-cognates than for the other two nonword types in the NDS group, who, ideally, should have shown no differences between the three conditions. Apparently, to native speakers of German without knowledge of Dutch, the near-cognates were *easier* to distinguish from existing German words (i.e. to say "no" to) than were homophones and pseudohomophones. However, because these data obtained from the NDS group serve as a baseline for "pure" item difficulty unrelated to Dutch, we are still able to assess whether this a-priori difference is modulated by the knowledge of Dutch in the DS speakers. This modulation, i.e. an interaction between Group and Nonword Type, was indeed obtained in the hypothesised direction: rather than having *less* difficulty with Dutch near-cognates (as would be expected based on the NDS group), they had *more* difficulties with them (except for the comparison with homophones in RTs, where the difference was not significant). Thus, we can conclude that relative to the NDS baseline, DS speakers showed a greater inhibitory near-cognate effect.

In an additional analysis investigating the role of Dutch proficiency in the DS group, we found that the inhibitory near-cognate effect (measured as the difference between near-cognates and homophones) in the error rates was significantly greater in individuals with a larger vocabulary size in Dutch. This lends further support to the notion that the difference in near-cognate effects between groups was due to experience with Dutch (and increased with increasing amounts of experience), and not to some other between-group difference.

Finally, in contrast to our nonword data in which there was a near-cognate effect and thus evidence for co-activation of the Dutch lexicon, we did not observe a similar

near-cognate effect in the *word data*, neither in the facilitatory nor in the inhibitory direction. A facilitatory effect could have been expected on the basis of earlier observations of such cognate effects in L1 sentence reading (Cop et al., 2017; Titone et al., 2011; van Assche et al., 2009); however, it has to be kept in mind that these effects were very small and only observed with large data sets, and mostly obtained with identical cognates. In contrast, an inhibitory effect would also have been possible here as a result of confusion between German and Dutch spellings in this situation with Dutch words as nonword foils, similar to what has been shown for a language decision task (Dijkstra et al., 2010). However, what we observed was no near-cognate effect at all, which is in line with previous reports of null effects for cognates in L1 lexical decision (Caramazza & Brones, 1979; de Groot et al., 2002; but see van Assche et al., 2009, for a report of cognate effects in a pilot L1 lexical decision experiment). We will return to this issue in the General Discussion, after we have seen whether there are any word cognate effects in Experiment 2.

Taken together, even though our data did not show any difference between near-cognates and control words for "yes" responses, our main nonword manipulation did reveal that the DS group had difficulties rejecting the Dutch spellings of near-cognates as non-existing words in German. This shows that even individuals who were clearly dominant in their L1 German could not completely ignore their relatively newly acquired L2, Dutch, when recognising German words. This supports the claim of a fundamental language non-selective quality of the bilingual language system (e.g. Dijkstra & van Heuven, 2002). However, an even harder test of this notion would be to embed the same, Dutch-spelled near-cognates in German sentences. As already mentioned, cross-language co-activation tends to decrease in monolingual sentence contexts, presumably because the sentence context acts as a strong language membership cue. Experiment 2 served to explore these lower limits of language non-selectivity.

Experiment 2: sentence reading

In Experiment 2, we used a more naturalistic reading task than the lexical decision task in Experiment 1, namely, (German) sentence reading for comprehension. The sentences contained the incorrect, Dutch spellings of Dutch-German near-cognates (STOK). If the Dutch lexicon, and particularly, the Dutch readings of cognates, were indeed co-activated even in purely German (L1) sentences, it should hardly be a problem to integrate the meaning of such a word spelled in Dutch in the meaning of the whole sentence. In contrast, this integration should be harder for non-cognate homophones (HEK), where meanings differ and language boundaries should be clearer. Finally, integrating near-cognates should also be harder than integrating German pseudohomophones (TRIK) because the latter have no word representation in any language at all. Thus, we were expecting a result opposite from Experiment 1, i.e. a *facilitatory* effect of Dutch near-cognates in the DS group.

Again, we compared native speakers of German with and without knowledge of Dutch. In contrast to the DS participants, German NDS speakers should not be influenced by the cognate status of near-cognate spellings, i.e. they should show less facilitation by near-cognates. Again, the inclusion of a German NDS baseline group was necessary since unanticipated differences between items in the two critical categories might occur. The same words that were previously used as critical items in Experiment 1 were now included as target items in the sentences. Eye movement measures were employed to investigate how both groups of participants processed Dutch near-cognates and the two control conditions. By using not only simple, non-Dutch misspellings of German words (pseudohomophones), but also Dutch-German non-cognate homophones (spelled in Dutch, like HEK) as a control condition, we again intended to rule out that near-cognate effects in the DS group would be a result of a higher visual familiarity with the Dutch spelling. While misspellings have generally been shown to cause delays in eye movements during sentence reading (e.g. Pynte et al., 2004; White & Liversedge, 2004), we are thus expecting fewer problems when DS participants read Dutch-spelled near-cognates because they are co-activated anyway along with the correct German spelling.

Given that the analysis of eye movements in reading typically yields a great variety of potentially interesting dependent variables, we restricted our main hypothesis to two variables which, based on previous research (see Rayner, 1998, 2009), are considered most sensitive to the cognitive word processing effects envisioned here: re-fixation time and re-reading time (see below). Nevertheless, we will still report other standard parameters to provide a more complete picture of oculomotor control in our task.

Method

Participants

Sixteen native German speakers with and 16 without experience with Dutch participated. Because the use of the same eye tracker and experimental procedure

across groups was considered important, and no non-Dutch speaking native speakers of German are available in Nijmegen, all participants were recruited and tested in Aachen, Germany.

The DS speakers (five male) were non-dyslexic and had experience with Dutch due to their study, travelling, family, or other circumstances. They were between 20 and 69 years old (mean: 30.5) and had between 0.25 and 32 years of experience with Dutch (mean: 9.0 years). Their self-ratings in terms of their use of Dutch and other foreign languages are summarised in Table 6. These self-ratings are lower than those of the participants in Experiment 1, which reflects the fact that they were not immersed in a Dutch-speaking environment at the time of the experiment like participants in Experiment 1. Despite this difference, though, their test scores in the Dutch LexTALE vocabulary test were not significantly lower (p = .20 in a t-test), with a mean of 68.8 (minimum 51.25, maximum 91.25, SD = 10.55). Like in Experiment 1, this corresponds to a CEF level of B2 (upper intermediate).

All DS participants reported to speak English with a mean frequency of 4.8 on the 7-point scale (SD = 2.0, range 1–7), and mostly more frequently than Dutch (twelve out of the 16 participants). Fifteen participants stated to also speak another foreign language like French (9) or Spanish (4) with a mean reported frequency of 2.7 on the scale; six participants spoke this language more frequently than Dutch.

The NDS participants (five male) were native speakers of German who were between 21 and 54 years old (mean = 28.4); none of them reported to be dyslexic. All of them spoke English as a foreign language, and 13 of them another one or two languages, mostly French (6), Spanish (4), and Italian (2). The most frequently spoken foreign language (English in 15 cases) was used with a frequency of 4.1 (SD = 1.5, range 2–7). None of the participants in this group reported to speak Dutch.

Materials

The same critical words and nonwords were used as in Experiment 1, but they were now embedded in

Table 6. Self-ratings from the language background questionnaire for the German speakers of Dutch in Experiment 2.

	Mean	SD	Range
How often do you read Dutch literature?	3.3	2.1	1–7
How often do you speak Dutch?	3.3	1.6	1–6
How often do you listen to Dutch radio / watch Dutch TV?	2.6	2.1	1–7
Self-rated reading experience in Dutch	4.5	1.4	2–7
Self-rated writing experience in Dutch	3.2	0.9	2–5
Self-rated speaking experience in Dutch	4.3	1.2	2–6
Mean Dutch experience (mean of previous 3)	4.0	1.0	2.0–5.3

Note: SD = Standard deviation.
[a]Self-ratings were given on a scale from 1 (low/rarely) to 7 (high/very often).

German sentences, like the Dutch-spelled near-cognate STOK:.

DAS FOTO ZEIGT DEN AUF EINEN STOK GESTÜTZTEN PRÄSIDENTEN. (target word underlined; literal translation: "the photo shows the on a stick leaning president").

The three critical conditions and the items (20 per condition) were identical to Experiment 1. However, note that the filler condition of simple German nonwords that were no misspellings of German words could not be included here (because they would not have been possible to integrate in a sentence meaning at all). Besides the 60 sentences containing misspellings, there were 120 sentences that contained only correctly spelled words: 20 with German spellings of near-cognates as in Experiment 1 (BLICK), 20 with the matching non-cognate control words (WALD), and 80 with the filler words (both cognates and non-cognates) from Experiment 1.

Given the semantic sentence constraints for the target words, it was not possible to use exactly the same sentence for a set of the critical conditions. However, we took great care to create comparable sentence environments across target conditions. The target words were always located in the middle of the sentence (i.e. not at the first two or last two word positions). All sentences were relatively simple and without commas. Many of them were taken from (or inspired by) the sentence examples provided by the *Deutscher Wortschatz* from the University of Leipzig (http:// wortschatz.uni-leipzig. de); these examples stem from German newspapers.

Sentence length was, on average, 11.0 words (SD = 1.2, range = 8–14), with a mean target position of 6.4 (SD = 2.1, range = 3–11). Sentence length and target position were matched across the three nonword and two word conditions (Dutch near-cognates, Dutch non-cognate homophones, German pseudohomophones; German near-cognates, German control words).

Because the predictability of words is known to influence eye movements during reading (e.g. Inhoff, 1984), we measured it by means of an online cloze test with a new sample of native speakers of German. Twenty-six participants (14 females, age range: 19–66 years) who were raised with German as only mother tongue took part. In the cloze test, sentences were displayed up to the target word and participants were required to type in a possible continuation of the sentence. For each target sentence, we counted how many participants correctly predicted the upcoming (N + 1) target word. We either used a strict criterion (N + 1 identical with target word) or a more relaxed criterion (N + 1 or N + 2 identical with target word; or target word is part of either N + 1 or N + 2). Table 7 shows the mean probabilities for these two criteria for the three critical nonword conditions.

Table 7. Mean cloze probabilities (standard deviations in parentheses) for the targets in the three experimental nonword conditions in Experiment 2.

Nonword Type	Strict criterion	Relaxed criterion
1 Dutch near-cognates (STOK)	0.16 (0.25)	0.17 (0.25)
2 Dutch non-cognate homophones (HEK)	0.10 (0.22)	0.11 (0.22)
3 Pseudohomophones (TRIK)	0.23 (0.34)	0.25 (0.36)

Note: Strict criterion = target word was the first word of the answer, relaxed criterion = target word was (or was contained in) the first two words of the answer.

A between-item ANOVA revealed no significant difference between these nonword categories in the mean percentage of correctly predicted target words (strict criterion: $F(2,57) = 1.21$, $p = .31$; relaxed criterion: $F(2,57) = 1.19$, $p = .31$).

Sentences were presented in uppercase letters for the same reason as in Experiment 1, i.e. to avoid the use of German-specific noun capitals as a language cue. Thirty-two of the 180 sentences (18%) were followed by a comprehension question (e.g. "Is the president in good physical shape?") that had to be answered with "yes" or "no" (50% each) indicated by a button press. The occurrence of the comprehension questions was not predictable for the participants.

Apparatus, task & procedure

Sentences were presented on a 21″ CRT monitor (100 Hz, 1240 × 1068 pixels) at a viewing distance of 67 cm. We utilised a 500 Hz EyeLink II head-mounted eye tracker (SR Research, Canada) with a chin rest. Horizontal three-point calibration routines were conducted at the beginning of the experiment and prior to each sentence trial.

The experiment was run in a single session. Participants were instructed to read the sentences silently and for comprehension, and to answer the (irregularly) interspersed comprehension questions ("yes" vs. "no"). They were explicitly instructed not to attend to any potentially occurring spelling errors. The first two trials consisted of filler sentences to accommodate participants to the procedure.

Each sentence trial started with the presentation of a fixation cross (height & width 1/2°) on the left side of the screen at the first letter position of the following sentence. Participants were instructed to press the space bar of the keyboard when they fixated the middle of the fixation cross. Afterwards, the sentence was displayed. Participants ended sentence viewing time by pressing the space bar of the computer keyboard. This was followed by either the fixation cross of the next sentence trial or by a comprehension question, which was presented centrally on the screen.

Comprehension questions occurred at random positions in the lists, but equally often for each target sentence category. No feedback regarding responses was provided.

Design

The independent variables (Nonword or Word Type and Group) were the same as in Experiment 1. Due to the differences in sentence environments across target word conditions, eye movement analyses focused on the target words. Specifically, we computed first fixation durations (i.e. the duration of the first fixation on the word), gaze durations (i.e. the sum of fixation durations until the eyes leave the word for the first time), and total reading times (i.e. the sum of all fixation durations on the word including regressions). However, these standard parameters are strongly confounded (in that the latter include the former), which somewhat limits the interpretation of the temporal dynamics of the effects. Thus, to disentangle the underlying processing dynamics, we statistically analyzed the mutually independent measures *first fixation durations* (early measure), *re-fixation time* (equivalent to gaze duration minus first fixation duration), and *re-reading time* (late measure, equivalent to total reading time minus gaze duration), the latter being the two central variables referred to in our a priori hypotheses.[2] Specifically, we reasoned that the subtle orthographic differences and the processing of cognate status (near-cognate vs. non-cognate) should require advanced processing of target words and should thus be less likely to become evident in early (vs. late) measures of word processing (see Rayner, 2009). Additionally, for the sake of completion, we computed the *total number of fixations* on a word, *regression probability* (i.e. the probability of a target word to be re-fixated after the word has been fixated for the first time), and *word skipping probability* (i.e. the proportion of instances where a word was not fixated at all).

Results

The percentage of incorrectly answered question was, on average, 1.9% (range = 0–9.0%) in the NDS and 2.2% (range = 0–3.1%) in the DS group. Thus, participants from both groups showed good understanding of the sentences.

For the analysis of the eye movement parameters mentioned above, the following trials with extreme values (unlikely representing typical word processing) were discarded: those in which first fixation durations were below 70 ms or above 900 ms; those in which gaze durations were above 3000 ms; and those in

which total reading times were above 5000 ms. This way, 2.3% of the data (2.5% in NDS, 2.0% in DS participants) were excluded.

Nonwords (misspelled words)

The means of the two groups concerning the three measures of fixation duration (first fixation duration, re-fixation time, and re-reading time; see above) are shown in Figure 4, and listed in Table 8.

We conducted separate ANOVAs on these three different (mutually independent) duration measures across participants and items, involving Nonword Type as within-participant and between-item factor, and Group as between-subject, but within-item factor. The full results of these ANOVAs can be found in Table 9.[3]

Focusing on the crucial Nonword Type x Group interaction, it can be seen from Table 9 that the hypothesised interaction emerged for the two late reading measures referred to in our hypotheses, namely re-fixation time and (though less robust statistically) re-reading time. Planned contrasts (see also Table 9) showed that this interaction arose both with German pseudohomophones and with Dutch homophones as control conditions. Further pairwise t-tests showed that DS participants' re-fixation times were shorter for near-cognates than for both pseudohomophones ($t_1(15) = -3.03$,

$p = .009$; $t_2(38) = -2.91$, $p = .006$) and non-cognate homophones ($t_1(15) = 3.77$, $p = .002$; $t_2(38) = -2.64$, $p = .012$), while there was no difference between these nonword types in the NDS group (cognates vs. pseudohomophones: $t_1(15) = -1.61$, $p = .13$; $t_2(38) = -1.32$, $p = .19$; cognates vs. homophones: $t_1(15) = -.71$, $p = .49$; $t_2(38) = -.79$, $p = .44$). Similarly, the DS group had shorter re-reading times (a trend across items) for near-cognates than for pseudohomophones ($t_1(15) = -7.00$, $p < .001$; $t_2(38) = -1.16$, $p = .08$) and homophones, but here only across participants ($t_1(15) = 4.61$, $p < .001$; $t_2(38) = -1.16$, $p = .25$). In contrast, the NDS group showed a significant difference (though smaller than in the DS group, hence the interaction, and only across participants) only when comparing near-cognates to pseudohomophones ($t_1(15) = -3.12$, $p = .007$; $t_2(38) = -1.17$, $p = .25$), but not when comparing them to homophones ($t_1(15) = .08$, $p = .94$; $t_2(38) = -.18$, $p = .86$).

Thus, DS speakers re-fixated shorter (either before or after leaving the word) on near-cognates than on both German pseudohomophones and homophones spelled in Dutch, even though the results were statistically more robust for re-fixation time (before leaving the word) than for re-reading time (after leaving the word) in the item analysis. For NDS participants, there was generally no difference between the near-cognates and the
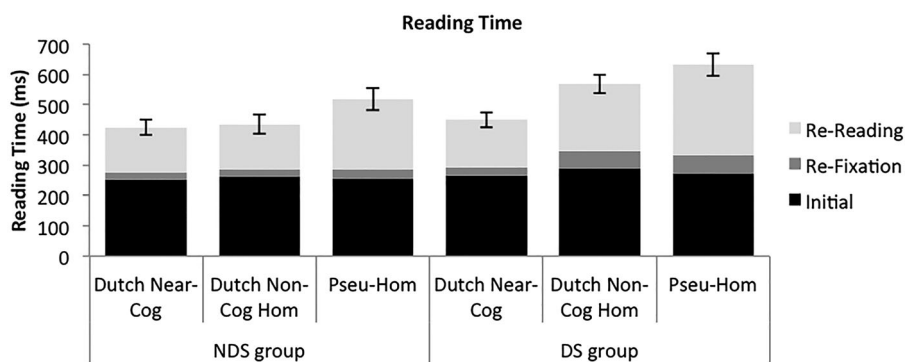
Figure 4. Bar graph of fixation measures (first fixation duration, re-fixation duration, re-reading time) for the three critical misspelling categories in the two groups. Cog = cognates, Hom = homophones, Pseu = pseudo.

Table 8. Mean temporal reading parameters (standard deviations in parentheses) for the three experimental nonword conditions in Experiment 2.

	NDS group			DS group		
Nonword Type	First fix. duration	Refix. time[a]	Reread. time[b]	First fix. duration	Refix. time[a]	Reread. time[b]
1 Dutch near-cognates (STOK)	253 (33)	21 (18)	150 (66)	265 (45)	27 (31)	158 (78)
2 Dutch non-cognate homoph. (HEK)	262 (36)	25 (21)	148 (69)	288 (56)	58 (40)	222 (101)
3 Pseudohomophones (TRIK)	256 (38)	29 (21)	233 (113)	272 (47)	59 (46)	301 (117)
Near-cognates vs. pseudohomophones (row 1–row 3)	−3	−8	−88	−7	−32	−143
Near-cognates vs. non-cognate homophones (row 1–row 2)	−9	−4	−2	−23	−31	−64

[a] re-fixation time = gaze duration - first fixation duration.
[b] re-reading time = total reading time - gaze duration (time spent on the word during regressions).

Table 9. Results of the ANOVA's for the three fixation duration measures in Experiment 2.

Measure	Effect	across participants				across items			
		df	F	p	η_p^2	df	F	p	η_p^2
First fixation duration	Nonword Type	2, 60	4.46	.02*	.13	2, 56	1.68	.20	.06
	Planned contrasts:								
	• near-cognates vs. PHPs	1, 30	.67	.42	.02	–	–	–	–
	• near-cognates vs. HPs	1, 30	12.48	.001**	.29	–	–	–	–
	Group	1, 30	1.79	.19	.06	1, 56	10.40	.002**	.16
	Nonword Type x Group	2, 60	.87	.43	.03	2, 56	.75	.48	.03
Re-fixation time	Nonword Type	2, 60	7.19	.002**	.19	2, 56	3.59	.03*	.11
	Planned contrasts:								
	• near-cognates vs. PHPs	1, 30	11.71	.002**	.28	1, 37	7.06	.01*	.16
	• near-cognates vs. HPs	1, 30	12.38	.001**	.29	1, 37	5.34	.03*	.13
	Group	1, 30	6.72	.02*	.18	1, 56	26.97	.000**	.33
	Nonword Type x Group	2, 60	3.32	.04*	.10	2, 56	3.51	.04*	.11
	Planned contrasts:								
	• near-cognates vs. PHPs	1, 30	4.27	.05*	.13	1, 37	6.26	.02*	.15
	• near-cognates vs. HPs	11, 30	7.38	.01*	.20	1, 37	5.67	.02*	.13
Re-reading time	Nonword Type	2, 60	27.60	.000**	.48	2, 56	1.44	.25	.05
	Planned contrasts:								
	• near-cognates vs. PHPs	1, 30	45.60	.000**	.61	–	–	–	–
	• near-cognates vs. HPs	1, 30	6.35	.02*	.18	–	–	–	–
	Group	1, 30	3.82	.08(*)	.10	1, 56	23.89	.000**	.30
	Nonword Type x Group	2, 60	2.71	.07(*)	.08	2, 56	2.48	.09(*)	.08
	Planned contrasts:								
	• near-cognates vs. PHPs	1, 30	3.31	.08(*)	.10	1, 37	5.25	.03*	.12
	• near-cognates vs. HPs	1, 30	7.05	.01*	.19	1, 37	3.18	.08(*)	.08

Note: HP = Dutch homophones, PHP = pseudohomophones.
Asterisks indicate levels of significance: **$p < .01$; *$p < .05$, (*)$p < .10$.

two types of misspellings except for longer re-reading times for pseudohomophones than for near-cognates (significant only across participants), but this effect was smaller than that in DS participants, giving rise to the above mentioned interaction of Group and Nonword Type.

In addition to fixation duration measures, we also computed three additional spatial word processing measures (number of fixations, regression probability, and word skipping probability), which will be reported here for the sake of a more complete picture of oculomotor processing in the reading task (i.e. without any explicit a priori hypotheses regarding these variables, see Table 10).

We conducted ANOVAs on these three dependent variables analogous to those on fixation durations reported above. The results are reported in Table 11.

The only dependent variable that showed a (trend towards a) Nonword Type x Group interaction was the total number of fixations (marginally significant across participants). This is not surprising, since this parameter is statistically associated with total reading times (see above). The planned contrasts showed that this interaction was more strongly carried by the comparison of near-cognates and pseudohomophones than by that of near-cognates and Dutch homophones, though the latter showed a trend in the analysis across participants. Pairwise t-tests revealed, similar to the results of the fixation duration measures, that NDS participants displayed no differences in number of fixations except for more fixations on pseudohomophones than on near-cognates, which was significant across participants only ($t_1(15) = -2.16$, $p = .047$; $t_2(38) = -1.19$, $p = .24$; near-cognates vs. homophones: $t_1(15) = -.53$, $p = .61$; $t_2(38) = -.44$, $p = .66$). However, the fact that an interaction of Group and Nonword Type arose for this comparison shows that this effect was smaller than in DS participants. In the t-tests, this effect (near-cognates vs. pseudohomophones) was indeed highly significant ($t_1(15) = -7.82$, $p < .001$; $t_2(38) = -2.26$, $p = .03$). In contrast, the comparison of near-cognates and homophones showed a significant difference only across participants ($t_1(15) = -5.89$, $p < .001$; $t_2(38) = -1.35$, $p = .19$; but recall that for this comparison, the interaction had also not become significant in the item analysis). Overall, this analysis on the total number of fixations – though less robust in the analysis across items – reflects the data pattern present in re-fixation and re-reading times.

Additional role of Dutch proficiency

Since we observed a significant positive correlation between Dutch vocabulary size and the (inhibitory) near-cognate effect in the error rates in Experiment 1, we also explored the data in this experiment to look for a similar relationship, even though the fairly low number of participants (N = 16 in the DS group) might make such a relation hard to observe. Still, we

Table 10. Mean spatial fixation parameters (standard deviations in parentheses) for the three experimental nonword conditions in Experiment 2.

	NDS group			DS group		
Nonword Type	Total no. fix.	Regr. prob.	Skipping prob.	Total no. fix.	Regr. prob.	Skipping prob.
1 Dutch near-cognates (STOK)	1.58(.41)	.26(.12)	.34(.19)	1.71(.45)	.30(.14)	.38(.17)
2 Dutch non-cognate homoph. (HEK)	1.64(.29)	.36(.15)	.34(.17)	2.00(.42)	.35(.15)	.39(.14)
3 Pseudohomophones (TRIK)	1.89(.50)	.35(.18)	.27(.17)	2.33(.49)	.40(.13)	.32(.14)
Near-cognates vs. pseudohomophones (row 1–row 3)	−.31	−.09	.07	−.62	−.10	.06
Near-cognates vs. non-cognate homophones (row 1–row 2)	−.06	−.10	0	−.29	−.05	−.01

Note: Fix. = fixations, Regr. = regression, prob. = probability.

carried out ANCOVAs on the DS participants' data with the Dutch LexTALE score as covariate, and considering those three variables that did show (at least marginally) significant Nonword Type x Group interactions in the analyses reported above: re-fixation times, re-reading times, and the total number of fixations. Recall that the relation between proficiency and the size of the near-cognate effect in Experiment 1 was only present in error rates, not in RTs, even though the latter did show the critical Nonword Type x Group interaction. It was thus possible that only some of these three variables (if any) would show a correlation.

Again, this analysis was conducted across (DS) participants only, and with only two nonword types (Dutch near-cognates and non-cognate homophones). Results showed that there was one measure among the three for which the Nonword Type x LexTALE score interaction was significant, which was the total number of fixations ($F_1(1,14) = 4.52$, $p = .05$). For the other two measures, re-fixation and re-reading time,

this interaction was absent (both $F < 1$). Figure 5 shows the scatterplot of the relationship between the (facilitatory) near-cognate effect (the difference of near-cognates and non-cognate homophones) in terms of the total number of fixations and LexTALE score in the DS group. The size of the correlation was $r = −.49$.

Thus, this analysis shows that the more Dutch words a DS speaker knew, the fewer fixations he or she had on Dutch near-cognates compared to homophones.

Words

Table 12 shows the mean first fixation durations, re-fixation times, and re-reading times for cognate vs. non-cognate words.

The two ANOVAs with Word Type and Group as factors showed no effect of Word Type on any of the three parameters (all $F < 1$). There was a main effect of Group for first fixation duration ($F_1(1, 30) = 4.28$, $p = .05$, $\eta_p^2 = .13$; $F_2(1, 38) = 11.67$, $p = .002$, $\eta_p^2 = .24$) and re-reading time ($F_1(1, 30) = 5.42$, $p = .03$, $\eta_p^2 = .15$; $F_2(1, 38) = 15.44$, p

Table 11. Results of the ANOVA's for the three spatial fixation parameters in Experiment 2.

		across participants				across items			
Measure	Effect	df	F	p	η_p^2	df	F	p	η_p^2
Total no. fix.	Nonword Type	2, 60	20.86	.000**	.41	2, 56	1.89	.16	.06
	Planned contrasts:								
	• near-cognates vs. PHPs	1, 30	32.64	.000**	.52	–	–	–	–
	• near-cognates vs. HPs	1, 30	8.29	.01**	.22				
	Group	1, 30	5.73	.02*	.16	1, 56	46.02	.000**	.45
	Nonword Type x Group	2, 60	2.56	.09(*)	.08	2, 56	3.29	.04*	.11
	Planned contrasts:								
	• near-cognates vs. PHPs	1, 30	3.76	.06(*)	.11	1, 37	7.18	.01*	.16
	• near-cognates vs. HPs	1, 30	3.67	.07(*)	.11	1, 37	2.18	.15	.06
Regr. prob.	Nonword Type	2, 60	5.84	.01**	.16	2, 56	1.46	.24	.05
	Planned contrasts:								
	• near-cognates vs. PHPs	1, 30	8.72	.01**	.23	–	–	–	–
	• near-cognates vs. HPs	1, 30	8.75	.01**	.23	–	–	–	–
	Group	1, 30	.47	.50	.02	1, 56	1.24	.27	.02
	Nonword Type x Group	2, 60	.62	.54	.02	2, 56	1.32	.28	.05
Skipping prob.	Nonword Type	2, 60	5.09	.01**	.15	2, 56	.99	.38	.03
	Planned contrasts:								
	• near-cognates vs. PHPs	1, 30	6.18	.02*	.17	–	–	–	–
	• near-cognates vs. HPs	1, 30	.17	.68	.01	–	–	–	–
	Group	1, 30	1.02	.32	.03	1, 56	6.33	.02*	.10
	Nonword Type x Group	2, 60	.05	.95	.00	2, 56	.20	.82	.01

Note: Fix. = fixations, Regr. = regression, prob. = probability, HP = Dutch homophones, PHP = pseudohomophones.
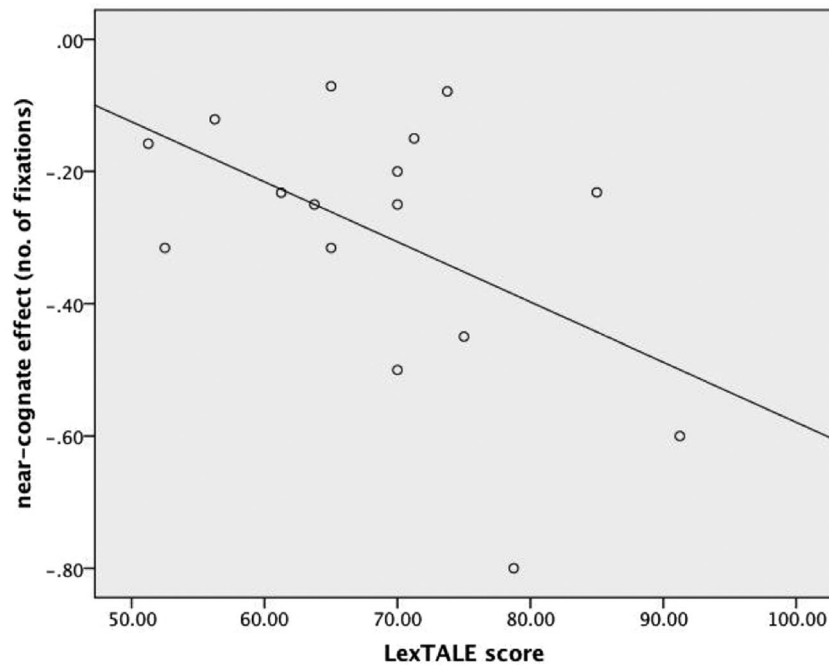Asterisks indicate levels of significance: **$p < .01$; *$p < .05$, (*)$p < .10$.

Figure 5. Scatter plot and regression line showing the correlation between the near-cognate effect (near-cognates – homophones) concerning the total number of fixations and Dutch vocabulary size, as measured by LexTALE.

< .001, $\eta_p^2 = .29$), and across items also for re-fixation times (F_1 (1, 30) = 1.36, $p = .25$, F_2 (1, 38) = 4.57, $p = .04$, $\eta_p^2 = .11$), all showing longer viewing times for DS compared to NDS individuals. Crucially, though, the interaction of Word Type with Group was not significant for any of the parameters (all $F < 1$). Thus, there was no evidence of a near-cognate effect in the word data.

Table 13 shows the spatial eye movement parameters (number of fixations, regressions, and skipping probabilities) across (correctly spelled) cognates and non-cognates.

Similar to the viewing times reported above, these eye movement parameters showed no differences between the two word types (F_1 (1, 30) = 2.92, $p = .10$, $\eta_p^2 = .09$ for skipping probability, all other $F < 1$). However, DS individuals displayed generally more fixations than NDS individuals (F_1 (1, 30) = 4.68, $p = .04$, $\eta_p^2 = .14$; F_2 (1, 38) = 25.44, $p < .001$, $\eta_p^2 = .40$) and had a slightly higher regression probability which was significant only across items (F_1 (1, 30) = 2.30, $p = .14$; F_2 (1, 38) = 11.23, $p = .002$, $\eta_p^2 = .23$), while skipping probability was unaffected by Group (both $F < 1$). Again, importantly, there

was no evidence whatsoever for a differential near-cognate effect in the two groups, i.e. a Word Type by Group interaction (all $F < 1$).

Discussion

In the German sentence reading experiment (Experiment 2), we observed facilitatory effects for near-cognates spelled in the incorrect (Dutch) way for German speakers of Dutch, but not or to a weaker extent for Germans without knowledge of Dutch. DS speakers spent less time fixating these misspelled near-cognates (like STOK) than both "simple" misspellings (pseudohomophones, like TRIK) and Dutch homophone misspellings (like HEK), as apparent in re-fixation and re-reading times (and, as a trend, also in the number of fixations), while NDS participants showed either smaller or no significant differences between these kinds of misspellings. In other words, considering that any kind of misspelling normally induces longer reading times (e.g. Pynte et al., 2004; White & Liversedge, 2004), the Dutch spelling of near-cognates represented a

Table 12. Mean temporal reading parameters (standard deviations in parentheses) for the German near-cognates and control words in Experiment 2.

Word Type	NDS group			DS group		
	First fix. duration	Refix. time[a]	Reread. time[b]	First fix. duration	Refix. time[a]	Reread. time[b]
1 German near-cognates (BLICK)	239(31)	14(12)	99(48)	255(26)	23(27)	150(95)
2 German controls (WALD)	232(35)	14(23)	104(54)	258(38)	18(18)	143(63)
Near-cognate effect (row 2–1)	−7	0	5	3	−5	−7

[a]re-fixation time = gaze duration – first fixation duration.
[b]re-reading time = total reading time – gaze duration (time spent on the word during regressions).

comparably small difficulty to those German speakers who knew Dutch. In fact, when comparing the incorrect Dutch spellings with the *correctly* spelled near-cognate word items, additional analyses (pairwise t-tests, across participants only) showed that DS participants displayed no significant difference between the German and the Dutch spellings of cognates (BLICK vs. STOK) for any of the parameters mentioned in the Results section (all $p > .30$)[4], while NDS participants did show longer first fixation durations, re-reading times (both $p < .02$), and a trend ($p = .067$) towards more fixations for Dutch compared to German near-cognates. This implies that DS participants, in contrast to the NDS group, were not even slowed down significantly when a near-cognate was spelled in Dutch rather than in German.

These results are in line with the earlier findings of a co-activation of the L2 readings of cognates during L1 reading as observed in the word data of eye movement studies (Cop et al., 2017; Titone et al., 2011; van Assche et al., 2009). First, the fact that DS speakers read near-cognates spelled in the Dutch way faster than "non-Dutch" misspellings (pseudohomophones) indicates that Dutch word knowledge played a role in this reading task, even though only German (L1) sentences were being read. Second, the boundary between the languages seems to be less clear for near-cognates than for other form-similar pairs (HEK – HECK), causing Dutch spellings to be relatively easily integrated into German sentences by individuals who have knowledge of Dutch. For sentence reading, current models of oculomotor control in reading typically assume an early lexicality or familiarity check (e.g. in the E-Z reader model of eye-movement control during reading, see Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Tokowicz, Liu, & Perfetti, 2011). Our results suggest that the Dutch near-cognates passed this check, as well as the further word processing stages, without any apparent problems in DS speakers. Because the presented spellings did not form existing German words, this means that this check has to be non-selective with respect to language.

Note that in Experiment 2, unlike in Experiment 1, there were no differences between the two Dutch-spelled nonword types (near-cognates and homophones) in the NDS group; however, the difference between pseudohomophones and near-cognates in that group persisted (but was larger in the DS group). Apparently, the dimensions along which the two types differed in the previous experiment were not equally relevant to lexical decision and sentence reading, suggesting that successful item matching may also depend on the task context. Why the pseudohomophones caused longer processing times in both tasks and also for the NDS group, is not clear to us. Apparently, actually existing (Dutch) spellings seem more "natural" even to speakers without knowledge of Dutch than fully artificial misspellings. Again, this observation highlights the importance of including a (non-L2 speaking) baseline group for interpreting any differences in (non)word type processing.

The effects obtained in Experiment 2 complement the inhibitory effects observed for the same targets in the German lexical decision task in Experiment 1, where the Dutch spellings had to be rejected. In analogy to the error rates in Experiment 1, for which the near-cognate effect correlated with Dutch vocabulary size, we again found that a discrete measure – the number of fixations – showed an (albeit small) near-cognate effect in the DS group that correlated with how well participants knew Dutch: the more words they knew, the fewer fixations they exhibited on the Dutch spelling of near-cognates, relative to the homophone misspellings. Thus, having a larger vocabulary in Dutch meant having problems identifying a Dutch near-cognate spelling as non-German in Experiment 1, and at the same time having very little trouble integrating the meaning of that near-cognate spelled in Dutch in a (German) sentence context in Experiment 2. The discrepancy of the effects in the two tasks across experiments (i.e. inhibition vs. facilitation) also points at certain constraints for using lexical decision procedures as a proxy for assessing processing mechanisms in normal reading for comprehension (Kuperman, Drieghe, Keuleers, & Brysbaert, 2013; see also Huestegge, Radach, Corbic, & Huestegge, 2009, for similar constraints with respect to single word naming tasks).

Similar to Experiment 1, we did not observe any evidence for a near-cognate effect in the word data, i.e. the correct German spellings of near-cognates. This again is

Table 13. Mean spatial fixation parameters (standard deviations in parentheses) for the German near-cognates and control words in Experiment 2.

Word Type	NDS group			DS group		
	Total no. fix.	Regr. prob.	Skipping prob.	Total no. fix.	Regr. prob.	Skipping prob.
1 German near-cognates (BLICK)	1.41(.29)	.20(.10)	.34(.15)	1.64(.40)	.27(.19)	.31(.11)
2 German controls (WALD)	1.39(.40)	.18(.09)	.27(.19)	1.64(.31)	.25(.13)	.29(.19)
Near-cognate effect (row 2–1)	−.02	−.02	−.07	0	−.02	−.02

Note: Fix. = fixations, Regr. = regression, prob. = probability.

in line with other findings showing smaller or non-existent (identical) cognate effects in L1 compared to L2 (Caramazza & Brones, 1979; de Groot et al., 2002). At the same time, it is at odds with the studies that did succeed in finding cognate effects in L1 reading, though with usually many more data points than we had (Cop et al., 2017; Titone et al., 2011; van Assche et al., 2009). We will return to this issue in the General Discussion.

General discussion

In two written language processing experiments carried out in the participants' native language German, we assessed the processing of homophonic near-cognates that were spelled in the non-target second language. The aim was to examine whether the functioning of the bilingual lexical system is principally non-selective with respect to language (i.e. both languages interact at all times), or whether circumstances can be so much in favour of the activation of only one language that an influence of the other is not observable any more. To test these lower limits of non-selectivity, we created a situation in which near-cognate effects seemed most unlikely to occur: participants were unbalanced L2 speakers who acquired their L2 Dutch relatively late in life, and who conducted the task in their clearly dominant language, German; the critical items were non-identical in the two languages; and in Experiment 2, they were embedded in sentences, which typically reduces cross-language activation.

Our results support the first hypothesis. In the lexical decision task (Experiment 1), German speakers of Dutch had larger difficulties rejecting Dutch near-cognates like STOK than both Dutch-German homophones like HEK and (non-Dutch) pseudohomophones like TRIK, compared to a baseline of German speakers without knowledge of Dutch ("NDS participants"). This is in line with the claim that the Dutch near-cognate representations could not be ignored in this task and affected rejection performance for the critical near-cognates.

In Experiment 2, the same target words were read in German sentence contexts. Here, participants were reading for comprehension, thus, the task did not explicitly require access to language membership. The inhibition previously observed for Dutch near-cognate spellings now turned into facilitation for DS participants, while NDS participants showed a significantly smaller effect (with pseudohomophones as control condition) or no effect (with homophones as control condition). It was thus easier for DS speakers to integrate Dutch-spelled words into the sentence context – in fact, almost as easy as when they were spelled correctly – when they were Dutch near-cognates than when they

were either purely "incidental" misspellings, or Dutch words that were homophonic to the German base word that was the correct target.

In several additional analyses requested by a reviewer, we made sure that our effects were not carried by those words in our stimulus set that overlapped with existing English words, either in terms of orthography (homographs) or phonology (homophones; see Footnotes 1 and 3). This kind of cross-language overlap between three West-Germanic languages had been inevitable.

These results can be explained theoretically in terms of co-activation of Dutch and German lexical representations, as in the BIA and BIA+ models, where lexical access is assumed to be language-nonselective. In these models, the two spellings of near-cognates like STOK and STOCK are listed separately in the bilingual orthographic lexicon. When the Dutch variant of a near-cognate like STOK is encountered, it will firstly activate its (Dutch) orthographic, phonological, and semantic representation, unlike non-existing misspellings without a lexical representation like TRIK, but just as other Dutch words (e.g. homophones like HEK) will. Due to the high form overlap, the input will also activate the orthographic and phonological representations of their German counterparts STOCK or HECK to a certain degree. However, the crucial difference between the two Dutch word categories STOK and HEK is that STOK also overlaps with STOCK in terms of meaning, while HEK does not share its meaning with HECK. Thus, STOK also already activates the meaning "stick", which in turn feeds activation back to its German orthographic representation STOCK, adding even more activation to this already co-activated word entry. Such a feedback from meaning to orthographic representations is missing for HEK / HECK, resulting in a lower co-activation – and thus less influence on processing during our experiments – of the Dutch homophone.

The lower co-activation of HECK as compared to STOCK has different consequences in the two tasks. In lexical decision, when STOCK is strongly co-activated upon the presentation of STOK, it will also activate the German language node (i.e. the representation of language membership), which should, in this task, be mapped to the "yes"-response. The activation of the German language node will occur simultaneously to the already present activation of the Dutch language node, activated by the actual stimulus STOK, which must be mapped to the "no"-response. The relative difficulty of DS individuals to reject Dutch near-cognates is likely to be a result of competition between these two responses, as there is both evidence for a "no"-response (Dutch language node activation) and for a "yes"-response (German language node activation).

In contrast to these cases, pseudohomophones like TRIK will not activate any entries in the Dutch part of the lexicon. Instead, they will only (but not as strongly as a correctly spelled German word) activate the (German) representation of the existing German word (TRICK) to a similar degree as STOK activates STOCK and HEK activates HECK, but because of the lack of an activation of the Dutch language node, this particular language (and response) conflict is not at play in this case. Therefore, pseudohomophones are relatively easy to reject by DS individuals, compared to Dutch near-cognates.

Note that our account of the data in Experiment 1 does not necessarily hinge on the assumption of language nodes. Instead, it would also hold when the lexical decision response involved a different sort of language check ("accept German words only"), for example, one that is operating on *language tags* that are part of the lexical representations (Green, 1998).

Accordingly, in the sentence reading experiment (Experiment 2), the same strong co-activation of the German word STOCK (when STOK is presented) leads to its easy integration into the sentence context. In contrast to the lexical decision task, this task likely involves a fundamentally different *task set* (e.g. Monsell, 2003) which is not primarily based on language nodes, but rather on nodes that code for the word's meaning. As both STOK and STOCK – that are both already active due to their form overlap – converge on the same meaning, this meaning representation will quickly pass the activation threshold – i.e. it promises to be "recognized" – and trigger the oculomotor command to leave the word. In contrast, the co-activation of HEK and HECK will lead to the activation of two competing meaning representations, and this competition takes time to be resolved. Similarly, pseudohomophones like TRIK are harder for DS speakers to integrate than near-cognates because there is no co-activated Dutch word representation that shares the same meaning as the target word, feeding activation back to the correct German orthographic representation TRICK. Furthermore, special problems for this nonword type might already arise at the familiarity check preceding lexical access in sentence reading (e.g. Reichle et al., 1998; see above) because mere misspellings are visually unfamiliar.

In summary, currently prevailing models of bilingual visual word recognition (Dijkstra & van Heuven, 2002; van Heuven et al., 1998) can account for our results, if it is assumed (as they do) that even during word recognition in the DS speakers' dominant and native language, words from the much weaker second language are also activated and compete for recognition.

Having said this, one cautionary note is in place. Of course, our stimulus materials contained Dutch words, which may have boosted the co-activation of Dutch relative to a reading situation without Dutch stimuli. It would indeed be more elegant to demonstrate the influence of a non-target language during word recognition entirely without any presence of that language. However, we do not know of any robust indicator of cross-language activation that satisfies this requirement; effects of, for instance, cross-language neighbourhoods or homophones that do not form non-target language words have been notoriously unstable (see, e.g. Lemhöfer & Dijkstra, 2004). We wanted to make use of the cognate effect as it is the most reliably observed indicator of the influence of the non-target language, even though it by definition involves the presence of non-target language words. Moreover, in our case it involved even the presence of Dutch, non-German spellings, which admittedly may have activated the Dutch lexicon to an even larger degree than the presence of identical cognates. However, the proportion of Dutch words in the experiment was rather small (20% in Experiment 1, and, when taking all words in the sentences into account, only 2% in Experiment 2 that were moreover accompanied by other, non-Dutch misspellings). Especially for the sentence reading experiment, we think it is plausible to argue that a predominantly language-selective L1 processing system should easily be able to ignore or suppress these two percent of non-target language words. The fact that we did nevertheless not obtain evidence for such a suppression is, in our view, still a strong argument against principally separated language systems in bilinguals.

One question that remains, however, is why, given this evidence for cross-language activation of cognate representations, we did not find a (near-)cognate effect in the *word* data. First of all, it has to be kept in mind that there have been failures to find such L2 cognate effects in L1 word recognition before (Caramazza & Brones, 1979; de Groot et al., 2002), and that those previous studies that did find such (descriptively very small) effects most often used identical cognates and large numbers of data points (Cop et al., 2017; Titone et al., 2011; van Assche et al., 2009). Here, because the cognate effect in the words was not our primary research question, we had a restricted number of orthographically non-identical cognates and a restricted number of participants, which might not have been sufficient to let the cognate effect surface.

However, the descriptive data on all various dependent variables do not even point at a trend for such an effect, different to what one would expect if low

statistical power was the only reason. Also, the near-cognate effect in the nonword data did reach significance (in Experiment 1, and in some measures in Experiment 2), even though statistical power was the same there. Thus, an alternative explanation for the null effect of cognate status in the word data is that L2 cognate representations (like BLIK) are co-activated also when correct German spellings are presented (BLICK), but that they do not have enough time to influence reaction times when these are fast (Dijkstra & van Heuven, 2002); particularly in Experiment 1, RTs were much (84 ms) faster for existing German words like BLICK than for the near-cognate nonword foils. A third possibility is that representations from the weaker language are taken into consideration mainly when they are directly activated by the input and when the search in the target language lexicon has been unsuccessful so far, as in the case of incorrect Dutch spellings like STOK. In contrast, when there is already a "hit" for the presented input string like BLICK, additional active representations like BLIK might not be considered, especially when from the "wrong" language. We do not currently have the means to distinguish between these three possible explanations.

Summary and conclusion

Returning to our original research question, we asked whether L2 cognate representations can be found to influence L1 processing even when circumstances are not conducive for such effects to occur, i.e. in full sentences, in unbalanced L2 speakers reading in their dominant language L1, and for cognates that have clearly distinct orthographic representations in the two languages. By doing this, we hoped to investigate an old, but unresolved issue – whether bilingual processing principally non-selective or not – from a new angle. In particular, we intended to look at L2 cognate effects in L1 reading in an alternative way to the so far only other available sort of evidence, namely, comparing bilinguals' (DS speakers') recognition times for (mostly identical) cognates to those for frequency-matched control words (with equivocal results). In both lexical decision and sentence reading, we did find clear effects of Dutch (L2) near-cognate status on nonword processing, suggesting that visual word recognition in L1 is indeed non-selective in terms of language even under circumstances that disfavour L2.

Notes

1. To rule out that the four words (two near-cognates, two non-cognate homophones) that were also existing English words disproportionally influenced the results, we also ran these analyses with these stimuli excluded. However, the overall pattern of results remained essentially unchanged. Similarly, on request of a reviewer, we looked at the data that resulted when excluding the fairly large set of items (22) that are near-homophonic to existing English words. Despite the loss of statistical power and of giving up item-by-item matching (some members of matched item sets were excluded in this analysis while other were not), the effects (and in particular the critical Nonword Type x Group interaction) remained descriptively stable (no statistics were computed in this analysis). Thus, the observed effects are highly unlikely to be affected by lexical overlap of the selected items with English words.

2. Note that in our computation of mean re-fixation time we deliberately included cases in which a re-fixation did not occur in terms of "0" values, so that our resulting measure also reflects those instances in which no re-fixations occurred. The same logic holds for the computation of re-reading times, which includes cases without any regressions in terms of "0" values, so that the overall measure can be interpreted as (roughly) reflecting the amount of late (i.e., second-pass) word processing.

3. Again, we re-calculated the analyses without the four stimuli that were existing English words, and obtained essentially unchanged results, suggesting that our main findings were not driven by the inclusion of these particular four items. Similarly to Experiment 1, we also looked at the descriptive data for all items excluding the 22 that had an English near-homophone. For first fixation duration (no substantial difference between DNS and DS participants) and re-fixation time (effects in DS, but not in NDS participants), the effects remained remarkably stable. In re-reading times, the critical Nonword Type x Group interaction patterned persisted in the predicted direction, but shifted a bit on the facilitation / inhibition continuum (what was small facilitation or no effect for near-cognates compared to the other two conditions in the NDS speakers previously now became inhibition, while the facilitation in DS speakers partly became weaker). We think that these changes in re-reading times reflect the fact that this measure is likely especially fragile statistically (and thus more strongly affected by the loss of statistical power associated with item pool reduction), since it depends on the few trials in which the participants actually go back to a previously fixated word. However, even with this complication, the overall pattern of results continued to show more facilitation for near-cognates vs. the two control conditions for DS relative to NDS speakers.

4. An exception is skipping rate, for which the effect was in the wrong direction (more frequent skipping for incorrect Dutch than for correct German near-cognates), $p = .029$.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Beauvillain, C., & Grainger, J. (1987). Accessing interlexical homographs: Some limitations of a language-selective access. *Journal of Memory and Language*, 26(6), 658–672.

Besner, D., & Davelaar, E. (1983). Suedohomofoan effects in visual word recognition: Evidence for phonological processing. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 37(2), 300–305. doi:10.1037/h0080719

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58(5), 412–424. doi:10.1027/1618-3169/a000123

Bultena, S., Dijkstra, T., & van Hell, J. G. (2013). Cognate and word class ambiguity effects in noun and verb processing. *Language and Cognitive Processes*, 28(9), 1350–1377. doi:10.1080/01690965.2012.718353

Bultena, S., Dijkstra, T., & van Hell, J. G. (2014). Cognate effects in sentence context depend on word class, L2 proficiency, and task. *Quarterly Journal of Experimental Psychology*, 67(6), 1214–1241. doi:10.1080/17470218.2013.853090

Caramazza, A., & Brones, I. (1979). Lexical access in bilinguals. *Bulletin of the Psychonomic Society*, 13(4), 212–214.

Casaponsa, A., & Duñabeitia, J. A. (2016). Lexical organization of language-ambiguous and language-specific words in bilinguals. *Quarterly Journal of Experimental Psychology*, 69(3), 589–604. doi:10.1080/17470218.2015.1064977

Cop, U., Dirix, N., van Assche, E., Drieghe, D., & Duyck, W. (2017). Reading a book in one or two languages? An eye movement study of cognate facilitation in L1 and L2 reading. *Bilingualism: Language and Cognition*, 20(4), 747–769. doi:10.1017/S1366728916000213

Costa, A., Santesteban, M., & Caño, A. (2005). On the facilitatory effects of cognate words in bilingual speech production. *Brain and Language*, 94(1), 94–103. doi:10.1016/j.bandl.2004.12.002

Davis, C., Sánchez-Casas, R., García-Albea, J. E., Guasch, M., Molero, M., & Ferré, P. (2010). Masked translation priming: Varying language experience and word type with Spanish-English bilinguals. *Bilingualism: Language and Cognition*, 13(2), 137–155. doi:10.1017/s1366728909990393

de Bot, K. (2004). The multilingual lexicon: Modelling selection and control. *International Journal of Multilingualism*, 1(1), 17–32. doi:10.1080/14790710408668176

de Groot, A. M. B., Borgwaldt, S., Bos, M., & van den Eijnden, E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language*, 47(1), 91–124. doi:10.1006/jmla.2001.2840

de Groot, A. M. B., Delmaar, P., & Lupker, S. J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: Support for nonselective access to bilingual memory. *The Quarterly Journal of Experimental Psychology Section A*, 53(2), 397–428. doi:10.1080/713755891

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, 62(3), 284–301. doi:10.1016/j.jml.2009.12.003

Dijkstra, T., & van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 189–225). Mahwah, NJ: Lawrence Erlbaum.

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5(3), 175–197. doi:10.1017/S1366728902003012

Duyck, W., Desmet, T., Verbeke, L. P. C., & Brysbaert, M. (2004). Wordgen: A tool for word selection and nonword generation in Dutch, English, German, and French. *Behavior Research Methods, Instruments, and Computers*, 36(3), 488–499. doi:10.3758/BF03195595

Duyck, W., van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for non-selective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 663–679. doi:10.1037/0278-7393.33.4.663

Forster, K., & Hector, J. (2002). Cascaded versus noncascaded models of lexical and semantic processing: The turple effect. *Memory & Cognition*, 30(7), 1106–1117. doi:10.3758/BF03194328

French, R. M. (1998). A simple recurrent network model of bilingual memory. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the twentieth annual cognitive science society conference* (pp. 368–373). Hillsdale, NJ: Lawrence Erlbaum.

Gerard, L. D., & Scarborough, D. L. (1989). Language-specific lexical access of homographs by bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(2), 305–315.

Gollan, T. H., Forster, K. I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(5), 1122–1139.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1(2), 67–81. doi:10.1017/S1366728998000133

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, 1(2), 131–149. doi:10.1017/S136672899800025X

Haigh, C. A., & Jared, D. (2007). The activation of phonological representations by bilinguals while reading silently: Evidence from interlingual homophones. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 623–644. doi:10.1037/0278-7393.33.4.623

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition*, 106(1), 501–511. doi:10.1016/j.cognition.2007.02.001

Huestegge, L., Radach, R., Corbic, D., & Huestegge, S. M. (2009). Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision Research*, 49(24), 2948–2959. doi:10.1016/j.visres.2009.09.012

Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning and Verbal Behavior*, 23, 612–624.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, 42(3), 643–650. doi:10.3758/brm.42.3.643

Ko, I. Y., Wang, M., & Kim, S. Y. (2011). Bilingual reading of compound words. *Journal of Psycholinguistic Research*, 40(1), 49–73. doi:10.1007/s10936-010-9155-x

Kolers, P. A. (1963). Interlingual word associations. *Journal of Verbal Learning and Verbal Behavior*, 2(4), 291–300.

Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *Quarterly Journal of Experimental Psychology*, 66(3), 563–580. doi:10.1080/17470218.2012.658820

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, 44(2), 325–343. doi:10.3758/s13428-011-0146-0

Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlexical homographs: Effects of code similarity in language specific and generalized lexical decision. *Memory and Cognition*, 32(4), 533–550. doi:10.3758/BF03195845

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A mega-study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(1), 12–31. doi:10.1037/0278-7393.34.1.12

Lemhöfer, K., Spalek, K., & Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *Journal of Memory and Language*, 59(3), 312–330. doi:10.1016/j.jml.2008.06.005

Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(2), 381–390. doi:10.1037/a0014875

Macizo, P., & Bajo, M. T. (2006). Reading for repetition and reading for translation: Do they involve the same processes? *Cognition*, 99(1), 1–34. doi:10.1016/j.cognition.2004.09.012

Macnamara, J., & Kushnir, S. L. (1971). Linguistic independence of bilinguals: The input switch. *Journal of Verbal Learning and Verbal Behavior*, 10(5), 480–487.

Midgley, K. J., Holcomb, P. J., & Grainger, J. (2011). Effects of cognate status on word comprehension in second language learners: An ERP investigation. *Journal of Cognitive Neuroscience*, 23(7), 1634–1647. doi:10.1162/jocn.2010.21463

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7(3), 134–140. doi:10.1016/S1364-6613(03)00028-7

Nas, G. (1983). Visual word recognition in bilinguals: Evidence for a cooperation between visual and sound based codes during access to a common lexical store. *Journal of Verbal Learning and Verbal Behavior*, 22(5), 526–534.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). doi:10.1126/science.aac4716

Pecher, D. (2001). Perception is a two-way junction: Feedback semantics in word recognition. *Psychonomic Bulletin and Review*, 8(3), 545–551. doi:10.3758/BF03196190

Poarch, G. J., & van Hell, J. G. (2012). Cross-language activation in children's speech production: Evidence from second language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, 111(3), 419–438. doi:10.1016/j.jecp.2011.09.008

Pureza, R., Soares, A. P., & Comesaña, M. (2016). Cognate status, syllable position and word length on bilingual Tip-Of-the-tongue states induction and resolution. *Bilingualism: Language and Cognition*, 19(3), 533–549. doi:10.1017/S1366728915000206

Pynte, J., Kennedy, A., & Ducrot, S. (2004). The influence of parafoveal typographical errors on eye movements in reading. *European Journal of Cognitive Psychology*, 16(1–2), 178–202. doi:10.1080/09541440340000169

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

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62(8), 1457–1506. doi:10.1080/17470210902816461

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105(1), 125–157. doi:10.1037/0033-295X.105.1.125

Reichle, E. D., Tokowicz, N., Liu, Y., & Perfetti, C. A. (2011). Testing an assumption of the E-Z Reader model of eye-movement control during reading: Using event-related potentials to examine the familiarity check. *Psychophysiology*, 48(7), 993–1003. doi:10.1111/j.1469-8986.2011.01169.x

Rodriguez-Fornells, A., Rotte, M., Heinze, H.-J., Nösselt, T., & Münte, T. F. (2002). Brain potential and functional MRI evidence for how to handle two languages within one brain. *Nature*, 415(6875), 1026–1029. doi:10.1038/4151026a

Scarborough, D. L., Gerard, L., & Cortese, C. (1984). Independence of lexical access in bilingual word recognition. *Journal of Verbal Learning and Verbal Behavior*, 23(1), 84–99.

Schepens, J., Dijkstra, T., & Grootjen, F. (2012). Distributions of cognates in Europe as based on levenshtein distance. *Bilingualism: Language and Cognition*, 15(1), 157–166. doi:10.1017/S1366728910000623

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, 55(2), 197–212. doi:10.1016/j.jml.2006.03.004

Starreveld, P. A., de Groot, A. M. B., Rossmark, B. M. M., & Van Hell, J. G. (2014). Parallel language activation during word processing in bilinguals: Evidence from word production in sentence context. *Bilingualism: Language and Cognition*, 17(2), 258–276. doi:10.1017/S1366728913000308

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: Electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex*, 20(4), 912–928. doi:10.1093/cercor/bhp153

Titone, D., Libben, M., Mercier, J., Whitford, V., & Pivneva, I. (2011). Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1-L2 intermixing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(6), 1412–1431. doi:10.1037/a0024492

van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. *Psychological Science*, 20(8), 923–927. doi:10.1111/j.1467-9280.2009.02389.x

van Hell, J. G., & de Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica*, 128(3), 431–451. doi: 10.1016/j.actpsy.2008.03.010

van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, 39(3), 458–483.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory and Cognition*, 15(3), 181–198.

White, S. J., & Liversedge, S. P. (2004). Orthographic familiarity influences initial eye fixation positions in reading. *European Journal of Cognitive Psychology*, *16*(1–2), 52–78. doi:10.1080/09541440340000204

Zeelenberg, R., Wagenmakers, E.-J., & Shiffrin, R. M. (2004). Nonword repetition priming in lexical decision reverses as a function of study task and speed stress. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 270–277. doi: 10.1037/0278-7393.30.1.270

Zhao, X., & Li, P. (2010). Bilingual lexical interactions in an unsupervised neural network model. *International Journal of Bilingual Education and Bilingualism*, *13*(5), 505–524. doi:10.1080/13670050.2010.488284