Journal of Experimental Psychology: Learning, Memory, and Cognition

Cross-Lingual Neighborhood Effects in Generalized Lexical Decision and Natural Reading

Nicolas Dirix, Uschi Cop, Denis Drieghe, and Wouter Duyck

Online First Publication, January 16, 2017. http://dx.doi.org/10.1037/xlm0000352

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/17/$12.00 http://dx.doi.org/10.1037/xlm0000352

Cross-Lingual Neighborhood Effects in Generalized Lexical Decision and Natural Reading

Nicolas Dirix and Uschi Cop
Ghent University

Denis Drieghe
University of Southampton

Wouter Duyck
Ghent University

The present study assessed intra- and cross-lingual neighborhood effects, using both a generalized lexical decision task and an analysis of a large-scale bilingual eye-tracking corpus (Cop, Dirix, Drieghe, & Duyck, 2016). Using new neighborhood density and frequency measures, the general lexical decision task yielded an inhibitory cross-lingual neighborhood density effect on reading times of second language words, replicating van Heuven, Dijkstra, and Grainger (1998). Reaction times for native language words were not influenced by neighborhood density or frequency but error rates showed cross-lingual neighborhood effects depending on target word frequency. The large-scale eye movement corpus confirmed effects of cross-lingual neighborhood on natural reading, even though participants were reading a novel in a unilingual context. Especially second language reading and to a lesser extent native language reading were influenced by lexical candidates from the nontarget language, although these effects in natural reading were largely facilitatory. These results offer strong and direct support for bilingual word recognition models that assume language-independent lexical access.

Keywords: bilingualism, eye tracking, reading, cross-lingual interactions, neighborhood effects

Supplemental materials: http://dx.doi.org/10.1037/xlm0000352.supp

During written word recognition, we are faced with the complex task of activating and identifying the correct lexical representation among a large group of orthographically similar, but not identical, representations. The term orthographic *neighbor,* coined by Coltheart, Davelaar, Jonasson, and Besner (1977), is used to denote such a similar word. Coltheart et al.'s (1977) definition of such a neighbor is any word that can be created by changing one letter of the target word while preserving letter positions (example: *house* is a neighbor of the word *horse*; see also Landauer & Streeter, 1973). We refer to this kind of neighbor as a s*ubstitution* neighbor from now on. Most studies examining neighbor effects used this definition. The number of neighbors of a particular target word is called the *neighborhood density* (*N* density).

In the word recognition literature, most models of (monolingual) word recognition hypothesize that a written word activates a set of possible lexical candidates. This means that at some point the correct target word has to be selected out of a number of neighbors. The search model (Forster, 1976) and the activation verification model (Paap, Newsome, McDonald, & Schvaneveldt, 1982) both predict that the neighborhood density will affect language performance because the actual decision is established by a frequency-ordered lexical search within those candidates. The longer the list of neighbors, the longer it would take to select the correct representation. Another influential model of word recognition, the interactive activation model (IA model; McClelland & Rumelhart, 1981), also makes the prediction that the number of activated candidates should affect lexical access, but proposes that the reason for this is lateral inhibition. In the IA model, word identification starts with letter identification. These letters feed forward activation to lexical candidates. Each of these representations has a resting level of activation, which is determined by the frequency of the word. The activated representations feed activation backward to the letter level. Word recognition is the end result of a competitive process between the activated lexical candidates, each inhibiting the others activation. The representation whose activation level first rises significantly above the identification threshold, is selected.

The most intuitive hypothesis formed by the IA model is that words with more orthographically similar lexical items would receive more lateral inhibition from these neighbors and this would slow lexical access to the target word (e.g., Grainger & Jacobs, 1993). On the other hand, a facilitative effect of a larger neighborhood is also not impossible within the IA model. More neighbors could cause greater overall excitation in the lexicon, which could help in specific tasks like the lexical decision task (e.g., Andrews, 1997; Grainger & Jacobs, 1996). Also, the feedback activation of multiple lexical candidates to partic-

Nicolas Dirix and Uschi Cop, Department of Experimental Psychology, Ghent University; Denis Drieghe, School of Psychology, University of Southampton; Wouter Duyck, Department of Experimental Psychology, Ghent University.

This research was funded by an aspirant scholarship from the Fonds voor Wetenschappelijk Onderzoek Vlaanderen (FWO) and by a concerted research action, Grant BOF13/GOA/032, from Ghent University.

Correspondence concerning this article should be addressed to Nicolas Dirix, Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium. E-mail: nicolas.dirix@ugent.be

ular letters, again activating the target representation, could facilitate activation of the correct lexical representation, so that large neighborhoods could again speed up word recognition in some instances. In the IA model, word frequency determines the resting level activation of representations, and lateral inhibition between the activated lexical candidates belonging to the orthographic neighborhood is also a function of their frequency. For this reason, it could be expected that recognition of low-frequent words would show larger effects of neighborhood density and frequency. This is because a low-frequent representation will need more time to accumulate enough activation to significantly rise above the activation levels of the higher frequency neighbors, thus delaying lexical access to the target word.

Within the IA architecture, precise predictions about the time course of neighborhood effects and whether the combination of these counteracting effects would result in facilitation or inhibition of recognition are difficult to make. With its complex interactions between parallel activation of letters and words and lateral inhibition among words, the IA model can account for a lot of different effects. Indeed, model simulations of the IA model have shown that effects can be both inhibitory (Jacobs & Grainger, 1992) and facilitatory (Coltheart & Rastle, 1994; Pollatsek, Perea, & Binder, 1999) depending on stimulus materials and small adjustment to the parameters of the model. For a more detailed discussion on this matter, we refer to Grainger and Jacobs (1996). Their Multiple Read-Out model allowed simulations of both inhibitory and facilitatory effects, based on multiple response criteria. As we discuss below, empirical investigations of neighborhood effects have also yielded a complicated mix of findings, with multiple moderating variables. This mimics the complicated pattern of neighborhood effect simulations that the computational models may exhibit.

Monolingual Neighborhood Effects

Isolated Word Studies

In the empirical search for neighborhood effects, mainly two variables have been manipulated. The first one is the neighborhood density. Coltheart et al. (1977) were the first to show neighborhood density effects for isolated word recognition. In a lexical decision task, they found inhibitory effects for nonwords with increasing neighborhood density, but no effects for words (see also Holcomb, Grainger, & O'Rourke, 2002). After this, multiple authors investigated the effects of neighborhood density on lexical decision word performance. As Andrews (1997) argued in a review paper, large neighborhoods are almost always associated with better performance in standard lexical decision tasks. Indeed, most of these experiments pointed toward a facilitatory effect of increasing neighborhood size, for both the speed and accuracy of lexical decision (Andrews, 1989, 1992; Carreiras, Perea, & Grainger, 1997; Forster & Shen, 1996; Huntsman & Lima, 2002; Johnson & Pugh, 1994; Laxon, Coltheart, & Keating, 1988; Perea & Rosa, 2000; Pollatsek et al., 1999; Sears, Campbell, & Lupker, 2006; Sears, Hino, & Lupker, 1995; for an additional review, see Mathey, 2001). Similar results were found for naming (Peereman & Content, 1995; Sears et al., 1995) and semantic categorization tasks (Carreiras et al., 1997; Forster & Shen, 1996). Perceptual identification tasks have shown mixed results. Carreiras et al. (1997) reported slower reaction times (RTs) for words with a large neighborhood density, whereas Snodgrass and Mintzer (1993) found a null effect in their Experiment 1, facilitation in Experiment 2 and inhibition in Experiment 3, 4 and 5. Andrews (1997) concluded that inhibitory effects of large neighborhoods observed for perceptual identification tasks are the result of unusual stimulus environments or elaborate guessing strategies.

Another neighborhood measure that is used regularly is whether the target word has a more frequent neighbor or not. We refer to this factor as *neighborhood frequency* (*N* frequency). In lexical decision tasks it is usually found that RTs are longer and accuracy is lower when a more frequent neighbor is present (Carreiras et al., 1997; Davis & Taft, 2005; Grainger, 1990; Grainger & Jacobs, 1996; Grainger, O'Regan, Jacobs, & Segui, 1989; Grainger, O'Regan, Jacobs, & Segui, 1992; Grainger & Segui, 1990; Huntsman & Lima, 1996; Perea & Pollatsek, 1998). This effect is also present for perceptual identification tasks (Carreiras et al., 1997; Grainger & Jacobs, 1996; Grainger & Segui, 1990).

Although research on neighborhood effects has predominantly used isolated word tasks, such as lexical decision tasks and naming tasks, there is some debate as to whether these tasks capture the cognitive processes underlying lexical access (e.g., Balota & Chumbley, 1984; Rayner & Pollatsek, 1989). The most important argument is that isolated word tasks entail a decision component or behavioral response, decreasing the validity of the measure (e.g., Paap & Johansen, 1994; Rayner & Liversedge, 2011; Snodgrass & Mintzer, 1993). Because of this decision component, and specifically in the case of neighborhood effects, the lexical decision task is for instance insensitive to the cases where the participant makes a response to the more frequent neighbor of the target word instead of the target itself and still responds with a correct "Yes" answer.

Kuperman, Drieghe, Keuleers, and Brysbaert (2013) indeed showed that the lexical decision task and a more natural reading method, that is, sentence reading in context are distinguishable and measure, to a large extent at least, different language processes. They found that lexical decision RTs only explained 5–17% of the variance in gaze durations on target words embedded in sentences after partialing out the effects of word frequency and word length. This dropped to 0.2% of the variance in fixation durations in natural reading when not only the target words, but all words in the sentences are analyzed.

Eye-Tracking Studies

It becomes clear that a more natural reading task, like sentence reading monitored by an eye tracker, could produce measures that are a closer approximation of natural language processes. Eye tracking can be used to assess the time that the eyes remain fixated on a word and thus provide more direct evidence for the existence of neighborhood influence on lexical access. In the case of neighborhood effects, eye tracking can be especially useful because it has a very high temporal resolution. This allows a specific investigation of the time course of potential *N* effects. Indeed, some eye movement measures (such as single fixation durations) reflect early stages in visual word recognition, whereas others like total reading time reflect higher-order language processes such as semantic integration. Eye tracking thus allows the study of language processing through multiple dependent variables reflecting several stages of word recognition, whereas the lexical decision task only allows investigation of RTs and accuracy scores. Eye tracking during natural reading should therefore contribute to the study of cross-lingual neighborhood effects above and beyond lexical decision results.

So far, only a handful of studies investigated neighborhood effects in sentence reading using eye tracking. Only one of those investigated the effect of neighborhood density (Pollatsek, Perea, & Binder, 1999). In Experiment 2 of Pollatsek et al. (1999), English monolinguals read target embedded sentences for comprehension. Half of the target words had a lot of neighbors (average = 8.5), the other half few (average = 2.2). All of these targets had at least one more frequent neighbor. Their first analysis showed an inhibitory effect of neighborhood density for gaze duration and total reading time. Because in this analysis the number of neighbors was confounded with the number of more frequent neighbors, Pollatsek et al. conducted another analysis, in which they held the number of more frequent neighbors constant. Under these conditions, they found that words with more low-frequent neighbors were skipped more often, but these words were also regressed to more often. The authors noted that the facilitatory effect on skipping rates might be due to initial misidentification of the target word. However they did find a facilitatory effect in gaze durations that could not be due to such misidentification because it was stronger in the sentences where the highest frequent neighbor was implausible in the sentence context.

Perea and Pollatsek (1998) conducted another reading study, this time investigating the effect of neighborhood frequency. In their Experiment 2 they instructed English monolingual participants to read sentences for comprehension. The embedded target words in these sentences were matched on number of neighbors. Half of the target words had an orthographic neighbor with a higher word frequency and the other half did not. The results showed more regressions toward the target word when it had a higher frequency neighbor than when it did not. Also, spillover effects were larger when the target word had a more frequent neighbor. These effects were larger for low-frequent target words. Davis, Perea, and Acha (2009) and Slattery (2009) conducted similar reading studies and confirmed that inhibitory effects of neighborhood frequency might occur late in the reading process. Davis et al. (2009) found an inhibitory effect of neighbor frequency for gaze durations and total reading time. Although there were also more regressions toward words with a more frequent neighbor, this effect was not significant. Slattery (2009) found an inhibitory effect of the presence of a more frequent neighbor in a sentence-reading task. More regressions were made and the total reading time was longer when the target word had a more frequent neighbor. He pinpointed this effect on the initial misidentification of the target word, by showing that these effects are no longer present when the more frequent word is not compatible with the prior sentence context. However, Sears et al., (2006) failed to find similar neighborhood frequency effects in an extensive set of reading experiments. They concluded that, at least in English, neighborhood frequency has no direct effect on reading times and has little to no effect on postidentification processes.

It becomes clear that all previous experiments examining neighborhood effects, either in isolated word studies or eye-tracking studies, have focused on one of the two neighborhood variables, density or frequency, while holding the other one constant. It is not clear what the net result would be of either variable in natural reading when both vary simultaneously.

Bilingual Neighborhood Effects

In the field of bilingualism, one of the most important questions has been whether word recognition involves activation of lexical candidates from the nontarget language. This question is tied in with the architecture of the bilingual lexical models, which may have one integrated, or two separate lexicons. Some have argued that lexical access for bilinguals is language-selective, meaning that when reading one language, only representations of that language are activated (Caramazza & Brones, 1979; Scarborough, Gerard, & Cortese, 1984). More recently however, a consensus has evolved in the literature that word recognition involves cross-lingual activation for bilinguals (for an overview see Dijkstra, 2007). The evidence for this mechanism comes mostly from studies using words that share features across two languages, such as interlingual homographs (words sharing orthography but not meaning across languages) and cognates. The latter are translation equivalent words that not only overlap in meaning but also in orthography (example of an identical cognate is the word "piano" in English and in Dutch). Cognates are recognized faster and more accurately than control words in behavioral studies that present words in isolation, such as lexical decision tasks (Bultena, Dijkstra, & van Hell, 2013; Dijkstra, Grainger, & van Heuven, 1999; Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010; Duyck, Van Assche, Drieghe, & Hartsuiker, 2007; Peeters, Dijkstra, & Grainger, 2013; Van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011). Similarly, cognate effects have been observed when bilinguals read text (Duyck et al., 2007; Van Assche et al., 2011; Van Assche, Duyck, & Brysbaert, 2013), even in the native language (Van Assche, Duyck, Hartsuiker, & Diependaele, 2009). This is remarkable because the language of a running text might serve as a useful cue in restricting access to the target language and therefore could speed up word recognition in this way (Altarriba, Kroll, Sholl, & Rayner, 1996; Van Assche, Duyck, & Hartsuiker, 2012).

Generally, these cognate effects are attributed to spreading activation between representations of both languages. Alternatively, because cognates share the exact same orthography and almost exact phonology and semantics, it has been argued that identical cognates could have a single representation across languages (see Dijkstra et al., 2010). This is important, because there is only very indirect evidence that cognates would actually be represented separately for each language, which is necessary for an explanation in terms of cross-lingual lexical activation. So, a more conservative test of cross-lingual lexical activation would be one in which representations that are certainly language-specific, such as neighbors, interact with other language-specific representations of the other language.

This is why the most compelling evidence for cross-lingual lexical access would come from cross-lingual neighborhood effects in bilingual reading. However, there is only study so far providing such evidence (van Heuven et al., 1998). In this study, Dutch–English bilinguals performed a blocked and mixed-progressive demasking task, a generalized lexical decision task and an English lexical decision task. Four item conditions were constructed by orthogonally manipulating the number of English and Dutch substitution neighbors in the CELEX database of the target words. In the progressive demasking task, participants had to identify four-letter words that gradually appeared on a screen as fast as possible. In the blocked version of the task, the experiment consisted of two blocks, one containing only L1 words, the other containing only L2 words. Both in the English and Dutch block of the progressive demasking task, van Heuven et al. found an inhibitory effect of nontarget *N* density, but this effect only reached full significance in the L2 block. In the mixed progressive demasking task, L1 and L2 words were presented in a random order. Here the authors expected to find larger effects, because in a mixed language setting, both languages have to be active to perform the task. In this experiment, inhibition from the nontarget neighbors was found for English and Dutch items. In the generalized lexical decision task, participants had to decide as fast and accurately as possible whether the target stimulus was a word (Dutch or English) or not. For the generalized lexical decision task, van Heuven et al. again found inhibition of Dutch *N* and facilitation for English *N* for RTs to the English items. No neighborhood effects were found for the Dutch items. In the English lexical decision task, monolingual and bilingual participants had to decide whether the presented stimulus was an English word or not. Here, again an inhibitory effect of Dutch *N* was found, showing that cross-lingual activation is not limited to mixed language contexts. All of these results were taken as evidence that words automatically activate substitution neighbors both pertaining to the target and nontarget language. Although van Heuven et al.'s (1998) results (nor design) were never directly replicated, two ERP studies supported the existence of cross-lingual *N* density effects, by showing a more negative N400 ERP component for words with more cross-lingual neighbors (Grossi, Savill, Thomas, & Thierry, 2012; Midgley, Holcomb, van Heuven, & Grainger, 2008). It is interesting to note that van Heuven et al. (1998) did not find any effects of cross-lingual *N* density in a blocked or selective L1 setting. Because this is the only study reporting cross-lingual *N* effects, so far there has been no direct evidence of cross-lingual activation of neighbors in L1 reading in a purely unilingual context. The present study will assess such an effect in bilingual natural reading.

Also, the effect of cross-lingual *N* frequency has never been investigated. In the monolingual literature, it is clear that the presence of a more frequent neighbor influences RTs and error rates in lexical decision tasks (e.g., Carreiras et al., 1997; Davis & Taft, 2005; Grainger, 1990; Grainger & Jacobs, 1996; et al., 1989; Perea & Pollatsek, 1998). Also, several studies provided evidence for an important role of this factor in *N* density effects (Carreiras et al., 1997; Davis & Taft, 2005; Grainger & Jacobs, 1996; Perea & Pollatsek, 1998). We address this issue in a bilingual context.

BIA+ Model

The findings on cross-lingual activation in bilingual reading described above have led to the development of the Bilingual Interactive Activation+ model (BIA+; Dijkstra & van Heuven, 2002) This model is the successor of the original BIA model (Dijkstra & van Heuven, 1998), which is a bilingual adaptation of the interactive activation model (McClelland & Rumelhart, 1981). The BIA+ model is a language nonselective model of lexical access, which entails an integrated bilingual lexicon (see Figure 1). To account for differences in word recognition depending on tasks and other nonlinguistic variables (e.g., instructions and expectations of the participants) the BIA+ model consists of a word identification system and a task/decision system. Like in the (B)IA

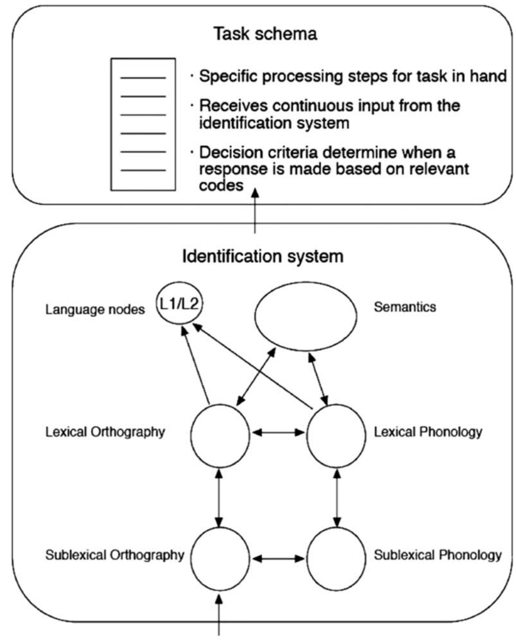

Figure 1. Schematic of the BIA+ model (taken from Dijkstra & van Heuven, 2002)

model, a set of orthographic candidates is activated through bottom-up activation when a written word is encountered. Depending on their similarity to the printed word and their resting-level activation, determined by the word frequency, these representations are partly activated. As L2 items tend to be lower in subjective frequency for unbalanced bilinguals, their representations are activated somewhat slower than L1 items. The activation spreads from the orthographic candidates to the connected phonological and semantic representations. Every word in the lexicon is connected to one of the available language nodes, representing the language membership of that word. In BIA+, these nodes also represent the global lexical activity of a language. These nodes do not feed activation back to the orthographic or phonological level so they cannot function as a language selection mechanism. This architecture for the bilingual lexicon implies that both intra- and cross-lingual orthographic neighbors should prominently influence lexical activation during visual word recognition.

Unfortunately, there are no simulations of neighborhood effects within the BIA+ model. However, because the BIA+ was then not formulated yet, Dijkstra, van Heuven, and Grainger (1998) and van Heuven et al. (1998) explained cross-lingual neighborhood effects using simulations of their results in the BIA model. Because BIA+ is basically the combination of the orthographic system of the BIA model with new (nonimplemented) task-scheme, phonology and semantic systems, and because neighborhood effects mainly rely on orthographic representations, these BIA simulations remain very informative about how BIA+ would model such effects. Dijkstra et al. (1998) operationalized simulated data as the amount of cycles the model needed to run for each item. The average amount of cycles for each condition (Target Language × *N* Language × *N* Density) could then be compared with the RT means of experimental data. Dijkstra et al. determined the degree of the correspondence between simulated and experimental data by qualitative (visual inspection of the response patterns) and quantitative (chi-square tests) measures. For each of the experimental conditions, the authors reported a reliable model fit on each measure. In the BIA simulations, inhibition from L1 neighbors for L1 word recognition is explained by the mechanism of lateral inhibition on the lexical level. Words with more neighbors suffer from the inhibition of their coactivated neighbors, thus taking longer to reach the identification threshold. The facilitation of within-language (L2) *N* density for bilinguals in English is explained by the relative activation of the two languages depending on word frequency in combination with asymmetric top-down inhibition from the language nodes implemented in the BIA-model. More specifically, the coactivated Dutch neighbors of the English word will exert inhibition on the target word through the Dutch language node. van Heuven et al. argue that inhibition will be larger toward words with a small compared with a larger *N* density, creating a relative facilitation effect for words with a larger within-language *N* density. Finally, inhibition from nontarget language neighbors (both in L1 and L2) is again explained by lateral inhibition. As words from different languages are integrated in one lexicon in the BIA model, the coactivated cross-lingual neighbors also inhibit the target word. The similarity between simulation outcomes and experimental data proved to be quite high, as these authors reported that there was no difference between the two on statistical tests.

The facilitatory effect of within-language *N* density on L2 English word recognition in bilinguals was also found by van Heuven et al. (1998) for L1 English monolinguals. To further explain these results, they refer to Grainger and Jacobs (1996), who showed that facilitatory effects of large *N* could be simulated with the IA model with the help of read-out criteria. This means that Dijkstra et al. (1998) explained the monolingual and bilingual facilitatory *N* density effects in English in two different ways, which is not very parsimonious. Another challenge for Dijkstra et al.'s interpretation is that the top down activation from language nodes is not implemented in the BIA+ model. As such, it is unclear how the authors would explain the facilitatory effects of target and nontarget *N* density within the BIA+ model.

Another complicating factor is that it has become clear that defining *N* densities by only including substitution neighbors is insufficient. For example, Davis et al. (2009) found an additional effect of *addition neighbors* (by adding a letter to a word, e.g., *frog* is an addition neighbor of *fog*) and *deletion neighbors* (by deleting a letter from a word, e.g., *rash* is a deletion neighbor of *trash*) above and beyond the effect of substitution neighbors. Word recognition models with fixed letter positions such as the IA and BIA+ have problems explaining these effects, because in these models lexical competition only occurs between representations of identical word length (see Davis and Bowers (2006) for an overview). Alternatively, there are monolingual models of word recognition with a relative positional nature that can account for effects of addition and deletion neighbors (e.g., the SOLAR model [Davis & Bowers, 2004], the SERIOL model [Whitney, 2001], the Overlap model [Gomez, Ratcliff, & Perea, 2008]). In the study by van Heuven et al. (1998), the *N* densities were calculated by counting the number of Dutch and English substitution neighbors of the target word using the CELEX database (Baayen, Piepenbrock, & Van Rijn, 1993). We might get a more accurate picture of cross-lingual *N* effects when we include addition and deletion neighbors in the *N* density measure. This new measure might be more sensitive in detecting cross-lingual effects in L1, which did not show very strong effects in van Heuven et al.

To conclude, despite the development of the BIA+ model and the abundance of papers addressing other effects of cross-lingual activation in visual word recognition, such as cognate effects (e.g., Bultena et al., 2013; Dijkstra et al., 1999; Duyck et al., 2007; Peeters et al., 2013; Van Assche, et al., 2011), there has been only one behavioral study that has provided direct evidence for parallel activation of lexical representations in bilingual word recognition by showing neighborhood effects. This study by van Heuven et al. (1998) used lexical decision instead of natural reading, and found no indications of L2 activation during pure L1 reading. We therefore begin by attempting to replicate van Heuven et al.'s generalized lexical decision task, both using their categorization of stimuli and an optimized *N* density measure, including addition and deletion neighbors. Next, we investigate whether these cross-lingual *N* effects are present in a large database of bilingual eye movements of natural reading (Cop et al., 2016) of parallel access to target language and nontarget language representations of the bilingual lexicon. This conservative test, in which unilingual running text is read, assess the generalizability of the cross-lingual effects obtained in experimental conditions with isolated words.

Experiment 1

In Experiment 1 we attempted to replicate the generalized lexical decision task of van Heuven et al. (1998), using the exact same stimuli as them to investigate cross-lingual *N* density effects in a new group of bilingual Dutch–English participants. Based on their findings, we expect within-language facilitation and cross-language inhibition for L2 reading and only a small within-language inhibitory effect and no cross-lingual effect for L1 reading.

We present linear mixed effects analyses (Baayen, Davidson, & Bates, 2008) including English and Dutch *N* frequency variables. By using a more inclusive measure of *N* density, we expect to find stronger effects of *N* density for L2 words and we might detect cross-lingual effects for L1 words.

Because of the architecture of the BIA+ model we expect larger effects of *N* density and frequency for low-frequent target words (Perea & Pollatsek, 1998; Williams, Perea, Pollatsek, & Rayner, 2006). Low-frequent representations in general need more time to accumulate sufficient activation to rise above the threshold of activation than high-frequent ones, so that they can benefit more from (or be hindered by) their neighbors.

Method

Participants. Thirty undergraduates received course credit for their participation in this experiment (19 females, 27 right-handed, M_{age} = 19.07 [2.08]). All students were unbalanced Dutch–English bilinguals. Participants were tested for language proficiency with the Dutch and English version of the LexTALE (Lexical Test for Advanced learners of English; Lemhöfer, & Broersma, 2012) and a self-report questionnaire (see Table B.1 in

Appendix B for detailed proficiency scores). For the questionnaire, participants rated how good they were at listening, speaking, reading and writing in both languages on a 5-point Likert scale.

Materials. The 160 words (80 Dutch and 80 English) and 160 nonwords were identical to those of Experiment 3 of van Heuven et al. (1998) (see Table 1 for word characteristics; see Appendix C for all stimuli). Importantly, we updated the *N* densities of the stimuli. It has become clear that *N* densities are inconsistently identified in the literature (Marian et al., 2012), so that researchers use different language databases to determine how large the neighborhoods of their stimuli are. This makes it difficult to compare results across experiments. To overcome this problem, Marian et al. (2012) developed the CLEARPOND database (Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities), which provides *N* densities and also allows comparing *N* densities across languages by including comparable corpora of multiple languages. When using this database to calculate *N* densities, we may replace the dichotomous neighborhood density classification that van Heuven et al. (1998) made with a more sensitive measure. In the current study, we used CLEARPOND (Marian et al., 2012) to determine a more accurate *N* density and frequency value, including within and cross-language substitution, addition and deletion neighbors. Furthermore, we calculated some additional word characteristics because they were not provided in the original study (e.g., bigram frequency) or because more up-to-date, and improved, measures exist nowadays (e.g., SUBTLEX frequencies, (SUBTLEX-NL, Keuleers, Brysbaert, & New, 2010; SUBTLEX-U.K., van Heuven, Mandera, Keuleers, & Brysbaert, 2014) instead of CELEX (Baayen et al., 1993) word frequencies).

Each participant saw each stimulus once, which resulted in 320 trials. All stimuli were presented in black against a white background. The font was Courier New, size 18 bold. Instruction language (Dutch or English) and response mapping (pressing the left button for a word, right for a nonword or vice versa) were counterbalanced across participants.

The experiment was programmed in E-Prime 2.0 software (Schneider, Eschman, & Zuccolotto, 2012). Stimuli were presented on a Benq XL2411Z 24-in. LED monitor (BenQ Corporation, Taipei, Taiwan). The computer used for the experiment was a Dell Optiplex 3020 minitower with a 3.2GHz Intel Core i5-4570 processor (Dell Inc., Round Rock, Texas). Participants had to respond by pressing left and right buttons on a RB-730 Cedrus response box.

Procedure. The procedure as was based on Experiment 3 of van Heuven et al. (1998). Participants were seated in a comfortable chair at approximately 45–60 cm from the screen. All instructions were presented on the screen. They were told they had to judge whether a presented letter string was either a word (in Dutch or English) or a nonword by pressing the according button. They were instructed to decide as quickly and accurately as possible. After the instructions, participants had to perform a practice block with 10 trials (five words and nonwords each), which was repeated if their accuracy was below 80%. None of the stimuli used in the practice block were used in the experimental block. Afterward the experimental block followed, with a presentation of the stimuli in a pseudorandom order (no more than four consecutive words or nonwords were presented). Halfway the experiment participants could take a short break.

A trial started with the presentation of a fixation cross (800 ms), followed by a blank screen of 300 ms. The stimulus was presented until the participant responded or for a maximum duration of 2,500 ms. The intertrial interval was kept constant at 700 ms.

After finishing the experiment, participants were presented with the English and Dutch version of the LexTALE and the self-reported questionnaire. The entire session lasted about 45 min.

Table 1
Descriptive Statistics for the Stimuli Used in Experiment 1 by Language and Neighborhood Density (Standard Deviations in Parentheses)

	Neighbors[a]		Number of neighbors[b]		Higher frequent neighbor[c]				
Language	Dutch	English	Dutch	English	Dutch	English	Word frequency[d]	Average bigram frequency[e]	CLD[f]
Dutch	Large	Large	7 (2.49)	7.1 (4.17)	.85 (.36)	.90 (.30)	2.257 (0.61)	1,828.47 (854.07)	.30 (.24)
	Large	Small	7.95 (2.94)	3.95 (2.8)	.90 (.30)	.75 (.43)	2.457 (0.86)	2,533.67 (1,665.93)	.25 (.24)
	Small	Large	4.05 (2.5)	6.6 (2.78)	.85 (.36)	.90 (.30)	2.364 (0.92)	1,947.32 (961.72)	.38 (.27)
	Small	Small	3.45 (2.27)	4.6 (4.46)	.65 (.48)	.75 (.43)	2.368 (0.45)	2,194.75 (1,227.3)	.26 (.25)
English	Large	Large	5.35 (2.8)	8.15 (3.44)	.70 (.46)	.90 (.30)	3.576 (0.57)	1,370.93 (541.67)	.36 (.23)
	Large	Small	2.15 (1.71)	8.3 (3.69)	.50 (.50)	.80 (.40)	3.758 (0.35)	1,300.00 (608.32)	.29 (.27)
	Small	Large	5.9 (6.2)	5.4 (2.58)	.70 (.46)	.55 (.50)	3.434 (0.65)	1,324.74 (668.89)	.30 (.33)
	Small	Small	1.9 (1.7)	4.15 (2.85)	.30 (.46)	.50 (.50)	3.505 (0.62)	1,282.15 (653.78)	.26 (.27)
Nonwords	Large	Large	5.675 (2.59)	6.2 (2.94)					
	Large	Small	4.975 (2.19)	3.675 (2.41)					
	Small	Large	3.125 (2.27)	6.475 (3.14)					
	Small	Small	2.35 (1.92)	3.375 (1.84)					

[a] *N* densities as defined by van Heuven et al. (1998). [b] Total CLEARPOND *N* densities (Marian, Bartolotti, Chabal, & Shook, 2012). [c] The proportion of words with a higher frequent neighbor. [d] Log10 Subtlex frequencies: SUBTLEX-NL for Dutch words (Keuleers et al., 2010), SUBTLEX-UK for English words (van Heuven et al., 2014). [e] Summated bigram frequencies (calculated using WordGen; Duyck, Desmet, Verbeke, & Brysbaert, 2004) were normalized for corpus size and then divided by word length to obtain average bigram frequencies. Bigram frequencies could not be calculated for the nonwords: because van Heuven et al. (1998) did not specify which of the nonwords were matched with which language, we could not determine which language corpus to use to calculate bigram frequencies. [f] Corrected Levenshtein distance (CLD) was calculated as a measure of orthographic overlap with the formula in Appendix A by comparing the word with its closest translation in NIM (Guasch, Boada, Ferré, & Sánchez-Casas, 2013).

Results

All analyses were performed in R version 3.1.2 (R Core Team, 2014). Models were fitted using the lme4-package in R (Bates, Mächler, Bolker, & Walker, 2014).[1] Several predictors were included in the analysis. Word frequency was included because of its indisputable role in (bilingual) visual word recognition (Baayen, Feldman, & Schreuder, 2006; Keuleers et al., 2010). Bigram frequency was added because word characteristics showed that there was a lot of variation for this variable between conditions. We also added a measure of orthographic overlap (i.e., corrected Levenshtein distance, the distance between the target word and its translation; Schepens, Dijkstra, & Grootjen, 2012) to the analysis (see Appendix A for the formula). We included this predictor because Van Assche et al., (2011) showed that an increased amount of cross-lingual overlap causes a continuous facilitatory effect in word recognition.

For all analyses, RTs, word frequencies and average bigram frequencies were log transformed with base 10 to normalize their distribution. All continuous variables were centered to reduce collinearity between main effects and interactions. For the analysis, stimuli that did not reach 70% accuracy were excluded (5.31% of the data). Furthermore, responses that were more than 2.5 standard deviations above or below participants' mean RT (4.99% of the data) were excluded. Additionally, for the RTs analysis incorrect responses were excluded (4.96% of the data). Separate analyses were carried out for each language (Dutch and English) and for the nonwords, both for RTs and error rates. The fixed factors in the models were Dutch *N* density (continuous), English *N* density (continuous), Dutch *N* Frequency ("Yes" indicated that the word had a more frequent neighbor in Dutch, "No" if it did not), English *N* Frequency ("Yes" indicated that the word had a more frequent neighbor in English, "No" if it did not), word frequency (continuous), average bigram frequency (continuous) and orthographic overlap (continuous). We included a random intercept per subject in all initial models. This ensured that differences between subjects concerning genetic, developmental or social factors were modeled. We also included a random intercept per word, to be able to generalize to other nouns, because our stimuli sample is not an exhaustive list of all nouns in a language. First a full model, including the two random clusters and all of the 2-way interactions between the neighborhood variables and word frequency, word length and bigram frequency, was fitted. The optimal model was discovered by backward fitting of the fixed effects, then forward fitting of the random effects and again backward fitting of the fixed effects (Barr, Levy, Scheepers, & Tily, 2013). Finally, the condition number or κ was calculated for each model to check if collinearity was an issue. According to Belsley, Kuh, & Welsch (1980), condition indexes around 5 to 10 are associated with weak dependencies between predictors; values of 30 and higher indicate moderate to strong collinearity.

We report the analysis of the Dutch and English words below. The analysis of the nonwords is reported in the online supplementary materials (see Table S.2 of the supplementary materials).

Results Dutch words. Results of the analysis of RTs and error rates are presented in Table D1 and D2 of Appendix D. The condition indexes for the final models were 4.463 for RTs and 4.255 for error rates. We did not find any main effects of within- or cross-lingual neighborhood density or neighborhood frequency on RTs or error rates.

However, for error rates the interaction between cross-lingual *N* density and word frequency approached significance ($\beta = 0.13$, $SE = 0.07$, $t = 1.93$, $p = .053$; see Figure D1 in Appendix D). Post hoc contrasts revealed that cross-lingual *N* density had a facilitatory effect for low-frequent words (<1.73 log word frequency, $\chi^2 = 3.84$, $df = 1$, $p < .05$) and a small inhibitory effect for high-frequent words (>4.19 log word frequency, $\chi^2 = 2.71$, $df = 1$, $p < .1$).

Results English words. Results of the analysis of RTs and error rates are presented in Table D3 and D4 of Appendix D. For the RT model, $\kappa = 8.501$; for the error rates model, $\kappa = 5.725$.

For RTs, again no main effect of any neighborhood variable was found. Nevertheless, there was a significant interaction between cross-lingual *N* density and bigram frequency ($\beta = -0.013$, $SE = 0.0066$, $t = -2.04$, $p < .05$; see Figure 2). Post hoc contrasts revealed an inhibitory effect of Dutch *N* density for words with a low bigram frequency (<2.953 log average bigram frequency, $\chi^2 = 3.85$, $df = 1$, $p < .05$). Reaction times for English words with a low bigram frequency were slower with increasing Dutch *N* density.

The interaction between the presence of a more frequent English neighbor and word frequency was significant ($\beta = -0.040$, $SE = 0.016$, $t = -2.51$, $p < .05$). There was inhibition of a more frequent neighbor for low-frequent words (<3.29 log word frequency, $\chi^2 = 3.84$, $df = 1$, $p < .05$) and a trend toward a facilitatory effect for high-frequent words (>3.87 log word frequency, $\chi^2 = 2.71$, $df = 1$, $p < .1$). The contrasts of the marginally significant interactions between English *N* density and word frequency ($\beta = 0.0044$, $SE = 0.0025$, $t = 1.75$, $p = .86$) and English *N* frequency and bigram frequency ($\beta = -0.056$, $SE = 0.032$, $t = -1.79$, $p = .78$) did not yield significant effects.

For error rates, the main effect of cross-lingual *N* density was significant ($\beta = 0.10$, $SE = 0.040$, $t = 2.32$, $p < .05$; see Figure D2 in Appendix D). More errors were made when the English noun had more Dutch neighbors. No other main effects of neighborhood were significant.

The marginal interaction between English *N* density and bigram frequency ($\beta = -0.29$, $SE = 0.17$, $t = -1.65$, $p = .099$) showed significant facilitation for English *N* density, but only for low bigram frequency words (>3.1055 log average bigram frequency, $\chi^2 = 3.84$, $df = 1$, $p < .05$). Contrasts for the marginally significant interaction between English *N* frequency and word frequency ($\beta = -1.01$, $SE = 0.55$, $t = -1.84$, $p = .065$) showed that there was inhibition for words with a more frequent neighbor, but only for low-frequent words (<3.665 log frequency, $\chi^2 = 3.84$, $df = 1$, $p < .05$).

[1] In an additional analysis, we analyzed the data by means of F1 (by participant) and F2 (by item) analyses of variance according to the procedure of van Heuven et al. (1998). By doing so, we were able to directly compare our results to those of the original study. This analysis yielded no significant within- nor between- language effects, both in RTs and Error rates.

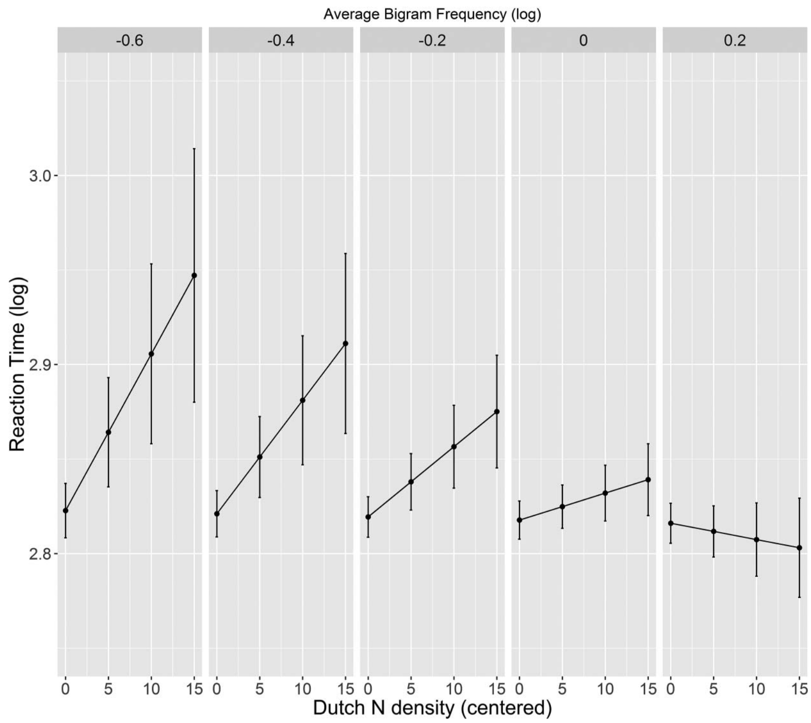

Figure 2. Reaction times (log transformed on the *y*-axis) for English words by Dutch *N* density (on the *x*-axis) and bigram frequency of the word (panels) for a generalized lexical decision task.

Discussion

In the present experiment, we attempted to replicate van Heuven et al.'s (1998) findings of cross-lingual effects of *N* density in a generalized lexical decision task for English words. A detailed pattern of neighborhood effects was discovered by analyzing the data by means of linear mixed models.

Cross-Lingual Neighborhood Effects

L1 lexical decision. For L1 (Dutch) reading we found a near-significant effect on error rates: a lower error rate with an increased cross-lingual *N* density for low-frequent words, but a trend for a reversed pattern for high-frequent words.

L2 lexical decision. In L2 (English) reading, the cross-lingual *N* effects were all inhibitory: we found slower RTs for low bigram frequency words and more errors for all L2 words when cross-lingual *N* density increased.

Within-Language Neighborhood Effects

L1 lexical decision. We found no L1 (Dutch) within-language effect of *N* density or frequency in the current study.

L2 lexical decision. When within-language L2 (English) *N* density increased, fewer errors were made toward words with a low bigram frequency. We also found slower RTs and more errors for low-frequent words when the noun had a within-language more frequent neighbor.

Concerning cross-lingual *N* effects, van Heuven et al. (1998) found an inhibitory effect of L1 *N* density for RTs on L2 words in a generalized lexical decision task. For L1 words the effect of L2 *N* density did not reach significance. In our lmer analyses of the replication, we found similar results for L2 and L1 words: inhibition with an increasing L1 *N* density for RTs (for words with a low bigram frequency) and for error rates. For L1 words, the effect of L2 *N* density did not reach significance for RTs, in the error rates there was only a trend.

Our RTs were in general slower than those of van Heuven et al. (1998). Instruction format can make a difference in lexical decision tasks when dealing with *N* density effects (Sears et al., 2006), but we emphasized both speed and accuracy (as van Heuven et al., 1998, did) so this is an unlikely cause of the slower RTs. Furthermore, the language proficiency of our participants could be different from those of van Heuven et al. (1998), causing the difference in RTs. Unfortunately, van Heuven, et al. (1998) did not provide proficiency scores for their participants so we cannot make a comparison. There was however a small procedural differences between our generalized lexical decision task and van Heuven et al.'s (1998). Our participants were allowed more time to answer, which indirectly might have slowed down the responses of our participants.

The dominant finding in the monolingual literature is facilitation of *N* density in lexical decision tasks (e.g., Andrews, 1989, 1992; Carreiras et al., 1997; Pollatsek et al., 1999). We observed this in L2 but not in L1. The inhibitory within-language effect of a more frequent neighbor (e.g., Carreiras et al., 1997; Davis & Taft, 2005; Grainger & Jacobs, 1996; Perea & Pollatsek, 1998) usually found in monolingual lexical decision was also only present in L2. Taking in account all of these results, we can conclude that the results for the English L2 words are fairly comparable to the existing monolingual literature on neighborhood effects. The discrepancy for L1 words could be explained by the fact that a generalized lexical decision task was used, which creates a bilingual context that is different from a normal unilingual lexical decision task (e.g., van Heuven et al.'s (1998) English lexical decision task also yielded no L2 within-language effect for bilingual participants, whereas this effect was present in the generalized lexical decision task).

Experiment 2

In Experiment 2, we investigated *N* density and *N* frequency effects in a large publicly available database of natural reading (Cop et al., 2016). We analyzed the eye movements of late unbalanced Dutch–English bilinguals when reading L1 and L2 nouns. Because of the discussion on lexical decision tasks as a marker for lexical access (Balota & Chumbley, 1984; Rayner & Pollatsek, 1989), it is very interesting to assess whether cross-lingual *N* effects obtained with isolated word stimuli generalize to natural text reading. Because cross-lingual neighborhood effects have only been investigated in lexical decision (van Heuven et al. (1998) and our Experiment 1), if we find cross-lingual neighborhood effects in these analyses, this would provide the first direct evidence in a completely unilingual context for the existence of activation of nontarget language lexical representations.

Because of the low correlations between RTs on lexical decision times and eye movements (Kuperman et al., 2013) and because it has been shown that neighborhood effects are very task dependent (e.g., Andrews, 1997; Carreiras et al., 1997) it is also difficult to make predictions based upon the results of the cross-lingual *N* effects found in lexical decision tasks. Some of the previous monolingual reading research has corroborated the idea that inhibition from neighbors might arise later in the reading process than facilitation (Perea & Pollatsek, 1998; Pollatsek et al., 1999.

We do expect that cross-lingual neighborhood effects should perfectly parallel within-language neighborhood effects, because in the BIA+ model lexical representations from both languages are included in the same integration lexical system, without distinction between both. Because top down inhibition from the language nodes is also absent (Dijkstra & van Heuven, 2002), cross-lingual neighbors should therefore behave exactly as intralingual neighbors. We also expect that for Dutch L1 reading, the cross-lingual effects will be smaller than for English L2 reading. This because for unbalanced bilinguals, most L2 representations are expected to be of lower frequency, thus having lower resting level activation.

Because our materials constitute an entire, long text, we analyze words varying in both neighborhood density and frequency. This means that we can examine the two effects at the same time. This will be very informative about the net effect of the neighborhood variables in bilingual natural language reading. The fact that our materials are not selected on certain lexical variables, also means that we investigate a database of nouns from a full range of word frequency, word length and bigram frequency. Because some results have shown that such lexical variables can modulate the neighborhood effects, we do expect to identify some important conditions in which neighborhood effects are stronger.

Method

Participants and materials. We selected all nouns (1 745 unique English and 1 777 unique Dutch nouns) from the GECO eye-tracking corpus of Cop et al., (2016). This corpus consists of eye movements recorded from 19 unbalanced Dutch–English bilingual (17 female, M_{age} = 21.2 [2.2]) and 13 English monolingual undergraduates (seven female, M_{age} = 21.8 [5.6]) who read the entire novel "The mysterious affair at Styles" by Agatha Christie (Title in Dutch: "De zaak Styles"). Participants' proficiency was tested with a proficiency battery including the LexTALE (Lemhöfer & Broersma, 2012), a lexical decision task and spelling tests (GL&SCHR for Dutch [De Pessemier and Andries (2009)], WRAT4 for English [Wilkinson and Robertson (2006)]). See Table B.2 in Appendix B for detailed proficiency scores. All nouns that had an identical cognate in the other language were excluded from the dataset (8% for Dutch, 9.1% for English). The final dataset consisted of 1 576 unique Dutch and 1 447 unique English nouns. See Table 2 for characteristics of these nouns.

Procedure. Each participant read the entire novel silently in a self-paced reading task over four separate sessions. They read half of the novel in Dutch, the other half in English. The order was counterbalanced. After each chapter, multiple-choice questions were asked to check whether participants were reading for comprehension. For further details on the procedure Cop et al., (2016) or Cop, Keuleers, Drieghe, and Duyck, see (2015).

Analyses eye movements. We analyzed two eye movement measures that reflect early language processes: Probability of first pass skipping of a word and single fixation duration, the first

Table 2
Descriptive Statistics for the Nouns Analyzed in Experiment 2, Averaged Over Stimuli Per Language (Standard Deviations in Parentheses)

	Neighborhood density[a]		Neighborhood frequency[b]						
Language	Dutch	English	Dutch	English	Word frequency[c]	Log average bigram frequency[d]	Average word length	CLD[e]	Rank of occurrence
Dutch	4.17 (5.35)	2.40 (5.16)	.30 (.44)	.18 (.35)	3.19 (0.97)	3.47 (0.23)	6.69 (2.65)	.32 (.26)	15.87 (30.42)
English	2.65 (4.60)	6.56 (7.44)	.25 (.42)	.53 (.50)	3.98 (0.91)	3.22 (0.24)	5.92 (2.19)	.35 (.29)	13.92 (20.13)

[a] Total CLEARPOND *N* densities (Marian et al., 2012). [b] The proportion of words with a higher frequent Neighbor. [c] Log10 Subtlex frequencies: SUBTLEX-NL for Dutch words (Keuleers et al., 2010), SUBTLEX-US for English words (Brysbaert & New, 2009). [d] Log10 summated bigram frequencies (calculated using WordGen, (Duyck et al., 2004) were normalized for corpus size and then divided by word length to obtain average bigram frequencies. [e] Corrected Levenshtein distance (CLD) was calculated as a measure of orthographic overlap with the formula in Appendix A by manually comparing the word with its closest translation.

fixation duration on a word that is fixated exactly once. We analyzed a measure reflecting intermediate language processing: Gaze duration, the sum of all fixation durations during first passage before the eyes move out of the word. Finally, we analyzed two measures that reflect later, higher-order, language processes such as semantic integration: total reading time, the sum of all fixation durations on the target word, including refixations and finally regression probability, the probability of making a regression back toward the target word.

Reading time measures and skipping probabilities were fitted in (general) linear mixed models using the lme4 (Version 1.1–7) and the lmerTest package (version 2.2–20) of R (Version 3.1.2; R Core Team, 2014). All of the initial models contained the fixed factors of English *N* Density (continuous), English *N* Frequency (*Yes* or *No*), Dutch *N* Density (continuous) and Dutch *N* Frequency (*Yes* or *No*). As in Experiment 1, Word Frequency (continuous), Bigram Frequency (continuous) and Orthographic Overlap (continuous) were included as predictors. Here, also Word Length (continuous) was included because this variable was not constant, as it was in Experiment 1. Furthermore, we included "rank of occurrence" as a predictor because some of the nouns occurred more than once in the novel, which could of course gradually facilitate their recognition. This factor simply consisted of the specific number of the presentation of the noun throughout the novel (i.e., "1" for the first occurrence, "2" for the second . . .). All predictors were calculated the same way as in Experiment 1. Model fitting was done in the same way as in Experiment 1.

Results

Dutch L1 reading. In the following section, the outcomes of the linear mixed effects analyses for L1 reading are presented.

Early measures. The outcome of the final model for skipping probabilities and single fixation durations is presented in Table E1 and E2 in Appendix E. For skipping probability, a logistic linear mixed model was fitted. For the single fixation analyses, only the nouns that received one fixation were selected (56.1%). Single fixation durations that differed more than 2.5 standard deviations from the subject means were excluded (2.20%). The condition index for the final skipping probability model was 10.708, for single fixation duration it was 5.169.

Cross-lingual* N *effects. We found no main effects of cross-lingual neighborhood density or neighborhood frequency for the early measures. The interaction between English *N* frequency and word frequency was marginally significant for skipping rates ($\beta = 0.078$, $SE = 0.043$, $z = -1.790$, $p < .1$). The probability of skipping a word was higher when this noun had a more frequent English neighbor, but only when the noun was high frequent (>3.89 log word frequency, $\chi^2 = 3.85$, $df = 1$, $p < .05$). For single fixation durations we found no cross-lingual neighborhood effects.

Within-language* N *effects. For skipping rates, we found a significant interactions of Dutch neighborhood density with word frequency ($\beta = -0.011$, $SE = 0.003$, $z = -3.266$, $p < .01$) and also with word length ($\beta = -0.007$, $SE = 0.002$, $z = -2.918$, $p < .01$). Post hoc contrasts showed that when nouns were low frequent (<1.90 log word frequency, $\chi^2 = 3.84$, $df = 1$, $p < .05$) or 5 characters or less ($\chi^2 = 10.48$, $df = 1$, $p < .01$), a larger amount of Dutch neighbors makes it more likely that the noun is skipped. For long words (14 characters or more, $\chi^2 = 3.96$, $df = 1$, $p < .05$) a larger neighborhood density made it less likely the noun was skipped.

For single fixation durations, we found an interaction of Dutch neighborhood density with word frequency ($\beta = 0.001$, $SE = 0.0002$, $t = 3.595$, $p < .001$). As the number of Dutch neighbors increased, single fixations became shorter for words with a log word frequency lower than 2.53 ($\chi^2 = 3.86$, $df = 1$, $p < .05$) and longer for high-frequent nouns (>4.23 log word frequency, $\chi^2 = 3.86$, $df = 1$, $p < .05$).

To sum up, in L1 reading we only observed a trend for cross-lingual *N* effects in skipping rates, an indicator of early language processing. The presence of a more frequent cross-lingual L2 neighbor yielded skipping of high-frequent L1 nouns. There was also within-language *N* density facilitation for low-frequent and short words, and inhibition for long words early in the word recognition process.

Intermediate measures. The outcome of the final model for gaze durations is presented in Table E3 in Appendix E. Gaze durations that differed more than 2.5 standard deviations from the subject means were excluded (2.55%). The condition index for the final model was 6.844.

Cross-lingual* N *effects. None of the main or interaction effects including cross-lingual neighborhood variables reached significance.

Within-language* N *effects. There were no main effects of within-language *N* density or *N* frequency. Again, the interaction between Dutch *N* density and word frequency was significant ($\beta = 0.001$, $SE = 0.0003$, $t = 3.662$, $p < .001$). Post hoc contrasts showed that for high-frequent nouns (>4.39 log word frequency, $\chi^2 = 3.86$, $df = 1$, $p < .05$), the effect was inhibitory whereas the effect was facilitatory for words with a log word frequency lower than 2.90 ($\chi^2 = 3.86$, $df = 1$, $p < .05$). The interaction between Dutch *N* frequency and word frequency was also significant ($\beta = 0.006$, $SE = 0.003$, $t = 2.017$, $p < .05$). Post hoc contrasts showed that fixations were shorter if a noun had a more frequent neighbor, but only when it had a log frequency lower than 3.05 ($\chi^2 = 4.02$, $df = 1$, $p < .05$).

Late measures. The outcome of the final model for total reading times and regression rates is presented in Table E4 and E5 in Appendix E. Total reading times that differed more than 2.5 standard deviations from the subject means were excluded (2.90%). For regression rate a logistic linear mixed model was fitted. For the total reading time model, $\kappa = 6.561$*; for the regression rate model,* $\kappa = 4.194$.

Cross-lingual* N *effects. Participants were marginally less likely to make a regression if a Dutch noun had a more frequent English neighbor ($\beta = -0.169$, $SE = 0.087$, $z = -1.915$, $p < .1$). Furthermore, for total reading times there was a marginally significant interaction between English *N* frequency and bigram frequency ($\beta = -0.030$, $SE = 0.017$, $t = -1.754$, $p < .1$). Post hoc contrasts for this interaction did not result in any significant effects.

Within-language* N *effects. There was a main effect of Dutch *N* density for regressions ($\beta = 0.019$, $SE = 0.008$, $z = 2.384$, $p < .05$): participants were more likely to make a regression to a word with an increasing number of neighbors. For total reading times, again the interaction between Dutch *N* density and word frequency was significant ($\beta = 0.001$, $SE = 0.0004$, $t = 3.281$, $p < .01$). Dutch *N* density had a facilitatory effect for low-frequent nouns

(<2.64 log word frequency, $\chi^2 = 3.86$, $df = 1$, $p < .05$) and an inhibitory effect for high-frequent nouns (>4.34 log word frequency, $\chi^2 = 3.85$, $df = 1$, $p < .05$). We also found a significant interaction between Dutch *N* frequency and word frequency ($\beta = 0.009$, $SE = 0.004$, $t = 2.394$, $p < .05$) and a marginal significant one with bigram frequency ($\beta = 0.027$, $SE = 0.015$, $t = 1.859$, $p < .1$). For words with a high word frequency there was an inhibitory effect of having a more frequent neighbor (>4.02 log word frequency, $\chi^2 = 3.85$, $df = 1$, $p < .05$), but there was facilitation for words with a low word frequency (<1.08 log word frequency, $\chi^2 = 3.84$, $df = 1$, $p < .05$). Contrasts for the interaction between Dutch *N* frequency and bigram frequency showed that there was inhibition of having a more frequent neighbor, but only for words with a high average log bigram frequency (>3.80, $\chi^2 = 3.84$, $df = 1$, $p < .05$).

In sum, for L1 reading, having a more frequent L2 neighbor makes it marginally less likely that a regression will be made to the target word. Again, we found a facilitatory effect of within-language *N* density for low-frequent words and an inhibitory effect for high-frequent words. There was also an effect of within-language *N* frequency on total reading times (inhibitory for high-frequent words, facilitatory for low frequent).

English L2 reading. In the following section, the outcomes of the linear mixed effects analyses for L2 reading are presented.

Early measures. The outcome of the final model for skipping probabilities and single fixation durations is presented in Table E6 and E7 in Appendix E. We fitted a logistic linear mixed model for skipping probability. For the single fixation analyses, only the nouns that received one fixation were selected (53.7%). Single fixation durations that differed more than 2.5 standard deviations from the subject means were excluded (2.14%). For the final skipping probability and single fixation models, $\kappa = 4.999$ *and* $\kappa = 8.350$, *respectively.*

Cross-lingual* N *effects. For skipping probabilities, there was a significant interaction between Dutch *N* frequency and average bigram frequency ($\beta = 0.256$, $SE = 0.127$, $z = 2.022$, $p < .05$). Post hoc contrasts for this interaction did not result in any significant effects. The main effect of cross-lingual *N* density was significant for single fixation durations ($\beta = -0.002$, $SE = 0.001$, $t = -2.508$, $p < .05$). The interaction of Dutch *N* density and word length was also significant for single fixation durations ($\beta = -0.001$, $SE = 0.0004$, $t = -2.736$, $p < .01$; see Figure 3). This interaction showed that there was a facilitatory effect of *N* density for words 5 characters long or longer ($\chi^2 = 4.72$, $df = 1$, $p < .05$).

Within-language* N *effects. The main effect of within-language *N* density was significant for skipping rates ($\beta = 0.009$, $SE = 0.003$, $z = 2.730$, $p < .01$). Targets with more neighbors were more likely to be skipped. Furthermore, there were significant interactions between English *N* density and average bigram frequency ($\beta = 0.018$, $SE = 0.009$, $z = 1.986$, $p < .05$), and between English *N* frequency and average bigram frequency ($\beta = -0.300$, $SE = 0.134$, $z = -2.239$, $p < .01$). Post hoc contrasts revealed that having a larger *N* density resulted in a higher skipping probability for nouns with a log average bigram frequency of 3.10 or more ($\chi^2 = 3.89$, $df = 1$, $p < .05$). The effect of *N* frequency was also facilitatory, but only for nouns with a bigram frequency lower than 3.13 ($\chi^2 = 3.90$, $df = 1$, $p < .05$). There was no effect of within-language neighborhood measures for single fixation durations.

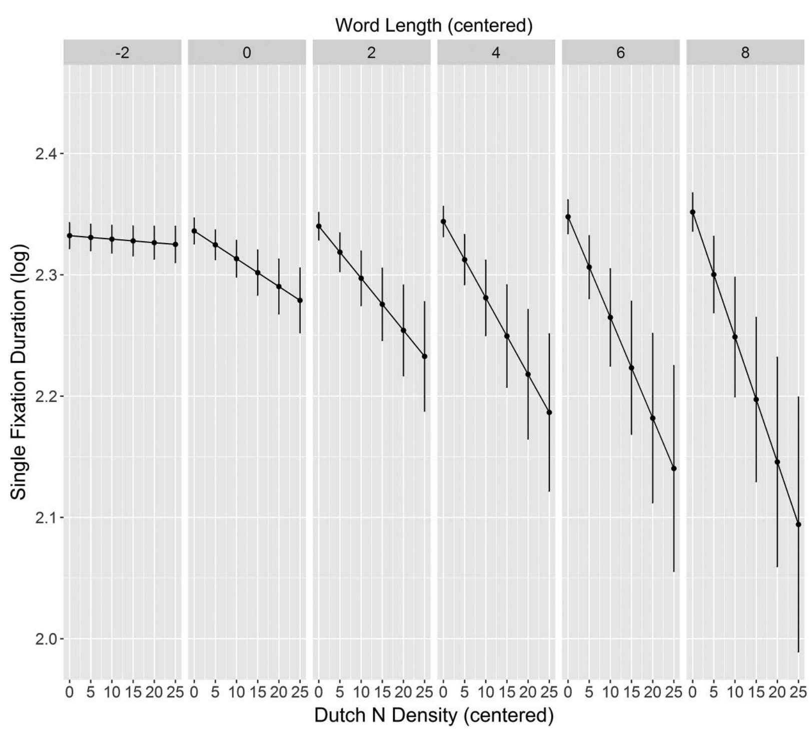

Figure 3. Single fixation durations (log transformed on the *y*-axis) for nouns dependent on Dutch cross-lingual *N* density (centered, on the *x*-axis) and target word length (panels) for English L2 reading.

In sum, for L2 reading, we found facilitatory effects of cross-lingual L1 *N* density on early language processing in single fixation duration. Within-lingual *N* density and *N* frequency also had a facilitatory effect in L2 reading, depending on the bigram frequency of the nouns.

Intermediate measures. The outcome of the final model for gaze durations is presented in Table E8. Gaze durations that differed more than 2.5 standard deviations from the subject means were excluded (2.55%). The condition number for the final model was 8.845.

Cross-lingual* N *effects. We found a marginally significant main effect of cross-lingual *N* density on gaze durations, which was facilitatory ($\beta = -0.002$, $SE = 0.001$, $t = -1.871$, $p < .1$). This measure interacted significantly with word length ($\beta = -0.001$, $SE = 0.0005$, $t = -2.174$, $p < .05$; see Figure 4). For nouns with a length of 7 characters or more there was facilitation with an increasing cross-lingual *N* density ($\chi^2 = 4.12$, $df = 1$, $p < .05$).

Within-language* N *effects. There were no significant effects of within-language *N* measures for gaze durations.

Late measures. The outcome of the final model for total reading times and regression rates is presented in Table E9 and E10 in Appendix E. Total reading times that differed more than 2.5 standard deviations from the subject means were excluded (2.84%). For regression rate a logistic linear mixed model was fitted. For the total reading times model, $\kappa = 5.898$; for the regression rates model, $\kappa = 4.954$.

Cross-lingual* N *effects. We found a significant facilitatory main effect of cross-lingual *N* density on total reading times ($\beta = -0.003$, $SE = 0.002$, $t = -2.066$, $p < .05$). This variable interacted significantly with word length ($\beta = -0.001$, $SE = 0.001$, $t = -1.984$, $p < .05$; see Figure 5). Post hoc contrasts showed that the effect of cross-lingual neighborhood density was significantly facilitatory for words with 6 characters or more ($\chi^2 = 4.36$, $df = 1$, $p < .05$). We failed to find any effects of cross-lingual neighborhood measures on regression rates.

Within-language* N *effects. There were no significant effects of any within-language *N* variables for regressions or total reading times.

In sum, for L2 reading, we found L1 *N* density facilitation for words of 6 letters and longer in late recognition processes, whereas there were no effects of L2 *N* density or *N* frequency.

English monolingual reading. To validate our neighborhood variables, we analyzed the eye movement toward nouns of monolinguals reading the same novel. These monolinguals were specifically selected as having no knowledge of any other language than English. None of the eye movement measures showed significant or marginally significant main effects of Dutch neighborhood density or frequency. Neither did any of the interactions between these measures and word frequency, word length or bigram frequency. We did find early and late facilitatory effects of English neighborhood density. For English neighborhood frequency, there was only a significant interaction with word length on skipping probability. For full analyses see Appendix E.

Discussion

Cross-lingual neighborhood effects. As the main interest of the current analysis was discovering cross-lingual N effects, we first present an overview of these effects for L1 and L2 reading.

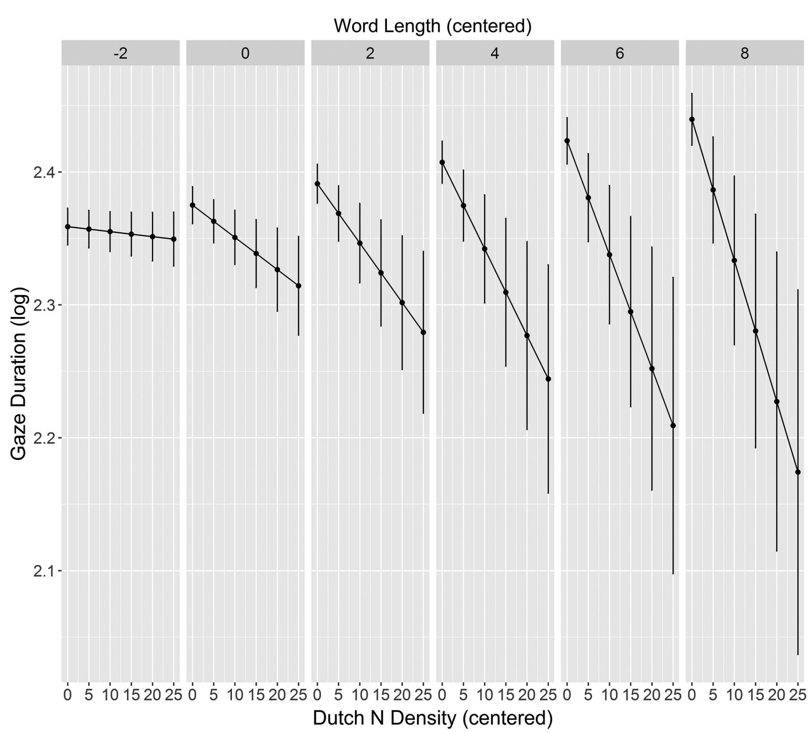

Figure 4. Gaze durations (log transformed on *the y*-axis) for nouns dependent on cross-lingual Dutch *N* density (centered, on the *x*-axis), and target word length (panels) for English L2 reading.

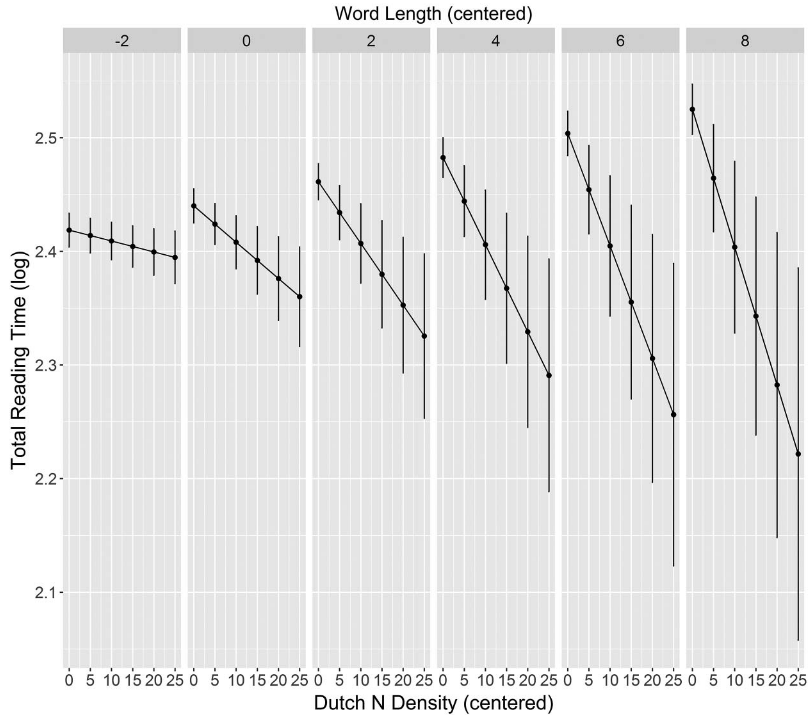

Figure 5. Total reading times (log transformed on the *y*-axis) for nouns dependent on cross-lingual Dutch *N* density (centered, on the *x*-axis) and target word length (panels) for English L2 reading.

L1 reading. For L1 (Dutch) reading, effects of cross-lingual neighbors were rather limited. Only marginally significant effects showed up in the analysis of skipping probabilities and regression rates. In both these measures of early and late language processing, a trend toward facilitation of cross-lingual *N* emerged. None of the timed measures showed effects of cross-lingual *N* density of *N* frequency.

L2 reading. For L2 (English) reading, we found early facilitatory effects of cross-lingual *N* density: when nouns were fixated only once, these fixations were shorter. This facilitatory effect was also found for gaze durations. The fact that the effects on single fixation duration and gaze duration were stronger for long words, might be an indication that lexical access was indeed facilitated by feedback from activated neighbors to letter representations, thus speeding up the identification especially for longer words. For total reading times we also find also facilitation for nouns with increasing L1 *N* density. Again this effect was again stronger for longer words. This could also be due to feedback toward letter representations.

In summary the most important finding is that even when reading natural text, cross-lingual effects of neighbors were present, which is an indication of nonselective lexical activation. These effects were especially clear in L2 reading, whereas they were less convincing or absent in L1 reading. This was in line with our expectations, as the lower resting level of L2 representations could experience a larger influence of their L1 neighbors. The cross-lingual effects found in our L2 bilingual reading data were facilitatory, leading to shorter reading times for nouns with more L1 neighbors.

It is important to note that the absence of an effect of cross-lingual neighbors in the monolingual data show that these effects are not confounds, but due to the knowledge of the second language of the participants.

Within-language neighborhood effects. The current dataset also allowed us to analyse within-language effects of neighborhood density and neighborhood frequency for both L1 and L2 reading.

L1 reading. For L1 reading we found within language effects of *N* density for early (skipping rates and single fixation durations), intermediate (gaze durations) and late (total reading times) eye movement measures. The direction of these effects was largely determined by the word frequency of the target noun. For low-frequent words, a larger *N* density seemed to facilitate the processing of that word. For high-frequent words the opposite was the case: an increasing neighborhood density slowed down the reading of the target word. Also, short words were skipped more often with increasing *N* density, whereas long words were skipped less. Words with a more frequent neighbor received longer total fixation times when they were high-frequent, bot shorter fixation times when they were low-frequent.

L2 reading. For English L2 reading, we only found an early facilitatory effect of *N* density. Nouns with a high bigram frequency were skipped more when they had a larger *N* density. For *N* frequency, again only an early effect was found: less skips were made of nouns with a more frequent within-language neighbor, except for nouns with a high bigram frequency.

Monolingual reading. The analysis of English monolingual reading showed facilitatory effects of *N* density for early measures (skipping probability and single fixation durations). For late measures, there was a facilitatory effect of *N* density on total reading times, as well as an inhibitory effect of *N* density for regression rate. For *N* frequency, we found more skipping with a more frequent neighbor for short words.

Our results for within-language neighborhood density are largely consistent with the results reported by Pollatsek et al. (1999). After controlling for the number of more frequent neighbors, they found early facilitatory effects of neighborhood density. Our early effects of neighborhood density were facilitatory for low-frequent and short words, but inhibitory for high-frequent, long words. Pollatsek et al.'s target nouns had a rather low word frequency (2.60 average log word frequency) and were rather short (average 4.5 characters). For the nouns with similar characteristics we also found facilitatory effects in our data.

To some extent we did replicate the late inhibitory effects of neighborhood frequency found in Davis et al. (2009); Perea and Pollatsek (1998), or Slattery (2009) in our bilingual reading data. In the Dutch L1 reading data total reading times were longer for words with a more frequent neighbor, but this was only true for high-frequency words. For our English monolinguals, we did not find an inhibitory effect of neighbor frequency for regression rates. These monolingual English data support the hypothesis, brought forward by Andrews (1997) and Sears et al. (2006), that there would be no inhibition from neighborhood frequency for English thus separating it from other alphabetic languages, like Spanish and Dutch.

Many of the *N* effects are situated in the skipping rates. Facilitatory effects in skipping rates of neighborhood density or frequency have been explained by misidentification of the target word with its more frequent neighbor (Pollatsek et al., 1999; Slattery, 2009) instead of as a real reflection of faster lexical access. When we look at our Dutch L1 reading results, we observe similar effects in single fixation durations, gaze durations and total reading times. We indeed find a higher correlation between skips and regressions ($r = .55$) for nouns with a more frequent neighbor than we do for nouns without one ($r = .45$; $z = 11.16$, $p < .001$). But we did not find a positive correlation between the skipping rate for nouns with a more frequent neighbor and the total reading time for these nouns ($r = -0.043$, $t = -4.12$, $df = 9252$, $p = 1$). These results show that it might be the case that a fraction of nouns was misidentified but these misidentifications do not have a significant effect on the total time spent on nouns with a high-frequent neighbor.

In general our bilingual and our monolingual within-language reading data show, in accordance with Pollatsek et al. (1999) that there might be early facilitation from activation of letters/bigrams of lexical candidates and to some extent late inhibition in the later word selection phase, although the facilitation/inhibition mechanism seems to interact strongly with word frequency of the target.

General Discussion

In this paper we investigated the effects of cross-lingual orthographic neighbors on bilingual language processing in two experiments. In Experiment 1, word recognition by Dutch–English bilinguals in a generalized lexical decision task was investigated, replicating van Heuven et al. (1998). In Experiment 2, a large database of eye movements during natural reading of a similar group (Cop et al., 2016) was analyzed.

For the data of Experiment 1, using LMM's and updated measures for neighborhood density and frequency (Marian et al., 2012), we did find longer RTs and more errors for L2 (English) words with increasing cross-lingual neighborhood density. For L1 (Dutch) words, error rates were higher for low-frequent words with increasing cross-lingual *N* density, but there was a trend in the opposite direction (lower error rates) for high frequent. We can conclude that only with this more refined analysis did we replicate the most important result of van Heuven et al. (1998), namely the cross-lingual effect of neighbor density in a generalized lexical decision task for L2 words. We additionally found a trend toward a cross-lingual *N* density effect on L1 words in the error rates. This suggests that activation of cross-lingual lexical candidates may not be confined to the processing of L2 words, although this effect was not statistically reliable. Therefore, just as van Heuven et al., the present isolated word experiment offers strong evidence for an L1 influence on L2 processing, but not vice versa.

Despite these cross-lingual effects in the generalized lexical decision task, we also investigated whether these effects would be found in a more unilingual context, because such a context might provide a cue to restrict lexical search and access to the target language (e.g., Van Assche et al., 2012), similar to the way in which readers use syntactic and semantic constraints in order to facilitate processing of upcoming words. In the current study, we therefore assessed the neighborhood effect with (a) words embedded in a completely unilingual language context and (b) a new paradigm, using eye tracking during natural language reading. In Experiment 2, a large database of bilingual eye movements (Cop et al., 2016) was analyzed to find evidence for activation of cross-lingual representations. The eye movements showed effects of cross-lingual neighborhood in early and late eye movement measures for L2 reading and trends for L1 reading. The pattern of results provides strong evidence that during natural reading, both in the early phase of lexical access as in the later language processes, written words activate not only orthographically similar words belonging to the target language, but also representations belonging to the nontarget language. The absence of any cross-lingual neighborhood effects for English monolinguals strongly suggests that it was indeed the knowledge of a second language that produced these cross-lingual neighborhood effects in the bilingual participants, and not some unknown lexical variable we failed to control. In summary, both the results of Experiment 1 and 2 provide evidence for parallel activation of lexical representations in bilingual word recognition and add strength to the argument of the existence of an integrated bilingual lexicon with language independent lexical access implemented in the BIA+ model (Dijkstra & van Heuven, 2002). We expected to see an asymmetry in cross-lingual effects (stronger effects in L2 than in L1 processing) because within the framework of the BIA+ model, L2 words should have a lower resting level of activation than L1 words, at least in our population of unbalanced bilinguals. This should imply that L2 words need more time to be activated, which makes them more sensitive to influences of other activated lexical candidates (Dijkstra & van Heuven, 2002). And indeed, both in the lexical decision and the eye movement results, the cross-lingual *N* effects were more pervasive in L2. In the lexical decision task, L1 words only showed a marginally significant cross-lingual effect in error rates, whereas L2 words showed effects in both error rates and RTs. For the eye movements we see that cross-lingual *N* only marginally influenced skipping rates and regression rates for L1 reading, whereas for L2 reading cross-lingual *N* significant effects were also present in reading tomes.

Next to effects of *N* density, we investigated the role of *N* frequency (i.e., was there an effect of having a more frequent neighbor). In Experiment 1, for L2 words the effect of a more frequent within-language neighbor was inhibitory for low-frequent words and facilitatory for high-frequent words, but we found the reverse pattern in Experiment 2. Apparently in natural reading, a more frequent within-language neighbor speeds up low-frequent word processing, while it slows down high-frequent word processing. This is in contrast to what was found in the monolingual reading studies of Davis et al. (2009); Perea and Pollatsek (1998) and Slattery (2009). This was also the first study investigating the effect of a more frequent cross-lingual neighbor on word recognition. In Experiment 1 we found no effect of cross-lingual *N* frequency above and beyond *N* density. In Experiment 2, we only found a marginally significant effect of cross-lingual neighborhood frequency in our L1 reading data. In L2 reading we only found effects of cross-lingual neighborhood density, not of *N* frequency. For Dutch L1 reading, the L2 neighbors seem to have to be of higher objective frequency than the target word before they are even known to our unbalanced participants. For English L2 reading the neighbors should not have to be of high frequency to have an effect, because the L1 neighbors will already be on average of higher (subjective) frequency than the L2 target words (Dijkstra & van Heuven, 2002). To address this issue of subjective frequency, an idea for future research might be to include *N* frequency as a continuous variable instead of the dichotomous variable we included in the current study, as this might better capture the influence of *N* frequency across languages on the reading process.

Lexical Decision Compared With Eye Tracking

In Experiment 1, the cross-lingual *N* effects were mostly inhibitory: for L2 words RTs were slower and error rates were higher with increasing cross-lingual *N* density. Only for the error rates for low-frequent L1 words was this effect facilitatory. In contrast, the cross-lingual *N* effects in the reading data were facilitatory, even in late language processing. This indicates that while performing a generalized lexical decision task, the cross-lingual activation generated by activated nontarget language neighbors, slows performance, whereas in general, natural language reading benefits from this cross-lingual activation.

When interpreting the difference between the results of our experiments we have to keep in mind that the lexical decision task entails a decision component that might provoke different kinds of strategies in participants, masking the real nature of lexical access (e.g., Paap & Johansen, 1994; Rayner & Liversedge, 2011). Lexical decision results have thus been shown to be very sensitive to blocking manipulations (e.g., van Heuven et al., 1998) and the selection of nonword stimuli (e.g., Lupker & Pexman, 2010; Stone & Van Orden, 1993). The fact that we found mostly inhibitory effects of neighborhood in Experiment 1, whereas the results of most lexical decision experiments have found facilitatory effects of *N* density for monolingual participants illustrates this sensitivity.

Supporting the possibility that the results of the generalized lexical decision task might be influenced by processes not directly related to lexical access alone, the precise direction and interactions of effects differed substantially between Experiments 1 and 2, for both within-language and cross-lingual neighborhood effects.[2] In Experiment 1 for example, we found no within-language *N* effects for Dutch words, whereas there was a marginally cross-lingual *N* effect in error rates. For English words, we found an inhibitory effect of cross-lingual *N* density for error rates while this was facilitatory for within-language *N* density. A language system with an integrated lexicon, such as the BIA+ model, does not make a qualitative distinction between L1 and L2 lexical representations (Dijkstra & van Heuven, 2002). The partly activated neighbors from the target and nontarget language should then have similar effects on target language word recognition or reading. In line with these expectations, for natural reading most of the cross-lingual effects resemble closely, although not exactly, the effects of within-language neighborhood effects. Where the patterns do diverge we see that this difference is driven by word frequency. In the current setting, natural reading might be a better approximation of lexical access than lexical decision.

Neighborhood Effects in the BIA(+) Model

Within the BIA+ architecture, orthographic neighbors, both of the target and the nontarget language, should influence lexical access to the target word by a complex interplay between inhibitory and excitatory connections at the word and letter level (Dijkstra & van Heuven, 2002). Simulations with the BIA model have confirmed that cross-lingual neighborhood density effects could be inhibitory (Dijkstra et al., 1998). Indeed, inhibition of neighbors of the nontarget language could be achieved by means of lateral inhibition. Within the BIA+ framework lateral inhibition from neighbors might be hidden by excitatory activation between representations for letters and words (as shown for the IA framework, Coltheart & Rastle, 1994; Pollatsek et al., 1999). Our data indeed shows that both inhibitory and facilitatory effects from neighbors are at play at the same time during word recognition.

The BIA+ architecture further predicts effects of the frequency of the target word and the frequency of the neighbor words. Because the subjective frequency of representations determines the resting activation of these representations, this could change the complex interactions between excitatory and inhibitory effects of activated neighbors. In our analyses of L1 and L2 language processing, we used corpus word frequencies that are supposed to reflect the frequency of exposure to words for monolinguals (Kuperman & Van Dyke, 2013). We especially expected word frequency effects to turn up in Experiment 2, because we investigated natural reading including a large range of noun characteristics (such as word frequency) in Experiment 2. In classic experiment designs where stimuli are matched on these variables per condition, it is more difficult to investigate these effects. Nevertheless, in both experiments we found that the frequency of the target word modulates the neighborhood effects. In Experiment 1, the effect of within-language *N* frequency on L2 RTs was modulated by word frequency. In Experiment 2, the effects of within-language *N* density on early and late language processes in L1 reading are modulated by word frequency. In both experiments the effect of increasing *N* density was facilitatory for low-frequent words and inhibitory for high-frequent words.

[2] We do note that there is little overlap in the stimuli: Only 15 of the Dutch and 17 of the English nouns of Experiment 1 were also present in the more than 1,700 stimuli of Experiment 2.

Considering our own findings as well as other studies finding effects of addition, deletion and transposition neighbors (e.g., Blythe, Johnson, Liversedge, & Rayner, 2014; Davis et al., 2009), we believe it important that the BIA+ model should be modified to accommodate a more flexible letter position coding mechanism. A mechanism lending itself for this purpose is the one proposed in the *overlap model* of Gomez et al. (2008). This model proposes that the representation of a letter is distributed across ordinal positions in the letter string. Every letter position has a specific standard deviation as free parameter in the model. This model expressively only models the letter coding mechanism, not any other higher order word recognition processes. This makes the overlap model easy to implement in other models, such as the BIA+ model. The effects of average bigram frequency in our data might also suggest that some kind of *open bigram coding* (Grainger & van Heuven, 2003), also implemented in the SERIOL model (Whitney, 2001), might be a good fit for these effects. Here words are coded by all of the ordered letter pairs that occur in that word. For example the word *hand* would be determined by the bigrams [ha, hn, hd, an, ad, nd]. In our opinion, the main architectural elements of the BIA+ model have promise in accommodating our most important results, namely the cross-lingual neighborhood effects found in natural reading, as long as a more flexible letter coding mechanism is implemented.

Neighborhood Effects in Models of Eye Movements

Following the large amount of eye-tracking research in reading, several models of eye movements of reading have been proposed in the last decades. As *N* effects never have been considered by such models, our findings could be of interest here. A first example of a model of eye movement control is the E-Z reader model (Pollatsek, Reichle, & Rayner, 2006; Rayner, Ashby, Pollatsek, & Reichle, 2004; Reichle, Rayner, & Pollatsek, 1999; Reichle, Warren, & McConnell, 2009). Although E-Z reader was designed for monolingual reading, Cop, Drieghe, and Duyck (2015) showed that L2 reading resembled childlike reading, which has been successfully simulated with the model (Reichle et al., 2013), thus making it likely this model can be applied to bilingual reading. The E-Z reader model assumes that lexical processing of words occurs serially and in two stages. In the early stage, which is called the *familiarity check*, orthographic and phonological information of the word is processed and presumably the possible lexical candidates become active. When this stage is complete, the oculomotor system starts programming a saccade toward the next word. After completing the second stage, the *verification stage* in which full lexical identification is accomplished, attention is shifted to the next word. The duration of the two stages is assumed to be determined by the frequency of the word and its predictability. However, it seems that the neighborhood of the word could be an additional factor of contribution to their duration (this was also hypothesized by Williams et al., 2006), given the role of *N* density and *N* frequency on changes in skipping probabilities and timed measures. For example, in L2 reading we find facilitation of the cross-lingual neighborhood in early and late measures of the reading process, meaning that the familiarity and verification stages are executed faster when L2 words have a larger L1 *N* density. Importantly, this means that not only characteristics of the target words, but also of their neighbors determine the duration of these stages. Indeed, the facilitation could for example be due to the higher subjective word frequencies of the L1 neighbors for L2 *N* density effects.

Another model of eye movements is SWIFT (Engbert, Longtin, & Kliegl, 2002; Engbert, Nuthmann, Richter, & Kliegl, 2005). It also proposes two lexical processing stages (preprocessing and lexical completion). The largest difference with E-Z reader is that SWIFT assumes that parallel processing of target words is possible, whereas the former proposes serial processing. Again, *N* density or *N* frequency could influence fixation times through the duration of lexical processing stages. Indeed, in simulation studies of SWIFT, Engbert et al. (2005) found for example a smaller frequency effect in simulated data than in experimental data. They suggest that certain variables that were not modeled, such as *N* frequency, are probably needed for a larger correspondence between their simulated data and experimental observations.

Conclusion

In conclusion, our lexical decision and natural reading data both provide convincing evidence for the existence of cross-lingual activation of lexical candidates during bilingual visual word recognition. Further research should focus on the lexical variables that modulate the size or the direction of these effects, such as the word frequency, both of the target word and its neighbors.

References

Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition, 24,* 477–492. http://dx.doi.org/10.3758/BF03200936

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 802–814. http://dx.doi.org/10.1037/0278-7393.15.5.802

Andrews, S. (1992). Frequency and neighborhood effects on lexical access or orthographic redundancy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 234–254. http://dx.doi.org/10.1037/0278-7393.18.2.234

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review, 4,* 439–461. http://dx.doi.org/10.3758/BF03214334

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language, 55,* 290–313. http://dx.doi.org/10.1016/j.jml.2006.03.008

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical data base on CD_ROM*. Philadelphia, PA: Linguistic Data Consortium.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340–357. http://dx.doi.org/10.1037/0096-1523.10.3.340

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting linear mixed-effects models using lme4*. Retrieved from http://arxiv.org/abs/1406.5823

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Detecting and assessing collinearity, in regression diagnostics: Identifying influential data and sources of collinearity*. Hoboken, NJ: John Wiley & Sons, Inc. http://dx.doi.org/10.1002/0471725153.ch3

Blythe, H. I., Johnson, R. L., Liversedge, S. P., & Rayner, K. (2014). Reading transposed text: Effects of transposed letter distance and consonant-vowel status on eye movements. *Attention, Perception, & Psychophysics, 76,* 2424–2440. http://dx.doi.org/10.3758/s13414-014-0707-2

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, Instruments, & Computers, 41,* 977–990. http://dx.doi.org/10.3758/BRM.41.4.977

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,* 1350–1377. http://dx.doi.org/10.1080/01690965.2012.718353

Caramazza, A., & Brones, I. (1979). Lexical Access in Bilinguals. *Bulletin of the Psychonomic Society, 13,* 212–214. http://dx.doi.org/10.3758/BF03335062

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 857–871. http://dx.doi.org/10.1037/0278-7393.23.4.857

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud—Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1197–1211. http://dx.doi.org/10.1037/0096-1523.20.6.1197

Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (2016). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior Research Methods*. Advance online publication. http://dx.doi.org/10.3758/s13428-016-0734-0

Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLoS One, 10,* e0134008. http://dx.doi.org/10.1371/journal.pone.0134008

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review, 22,* 1216–1234. http://dx.doi.org/10.3758/s13423-015-0819-2

Davis, C. J., & Bowers, J. S. (2004). What do letter migration errors reveal about letter position coding in visual word recognition? *Journal of Experimental Psychology: Human Perception and Performance, 30,* 923–941. http://dx.doi.org/10.1037/0096-1523.30.5.923

Davis, C. J., & Bowers, J. S. (2006). Contrasting five different theories of letter position coding: Evidence from orthographic similarity effects. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 535–557. http://dx.doi.org/10.1037/0096-1523.32.3.535

Davis, C. J., Perea, M., & Acha, J. (2009). Re(de)fining the orthographic neighborhood: The role of addition and deletion neighbors in lexical decision and reading. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1550–1570. http://dx.doi.org/10.1037/a0014253

Davis, C. J., & Taft, M. (2005). More words in the neighborhood: Interference in lexical decision due to deletion neighbors. *Psychonomic Bulletin & Review, 12,* 904–910. http://dx.doi.org/10.3758/BF03196784

De Pessemier, P., & Andries, C. (2009). *GL&SCHR*. Leuven, the Netherlands: Garant.

Dijkstra, T. (2007). The multilingual lexicon. In M. Gaskell (Ed.), *Handbook of psycholinguistics* (pp. 251–265). New York, NY: Oxford University Press.

Dijkstra, T., Grainger, J., & van Heuven, W. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language, 41,* 496–518. http://dx.doi.org/10.1006/jmla.1999.2654

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,* 284–301. http://dx.doi.org/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. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 189–225). Mahwah, NJ: 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,* 175–197. http://dx.doi.org/10.1017/S1366728902003012

Dijkstra, T., van Heuven, W. J. B., & Grainger, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica, 38,* 177–196.

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

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, & Computers, 36,* 488–499. http://dx.doi.org/10.3758/BF03195595

Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research, 42,* 621–636. http://dx.doi.org/10.1016/S0042-6989(01)00301-7

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112,* 777–813. http://dx.doi.org/10.1037/0033-295X.112.4.777

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257–287). Amsterdam, the Netherlands: North-Holland.

Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: Absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 696–713. http://dx.doi.org/10.1037/0278-7393.22.3.696

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological Review, 115,* 577–600. http://dx.doi.org/10.1037/a0012667

Grainger, J. (1990). Word-frequency and neighborhood frequency-effects in lexical decision and naming. *Journal of Memory and Language, 29,* 228–244. http://dx.doi.org/10.1016/0749-596X(90)90074-A

Grainger, J., & Jacobs, A. M. (1993). Masked partial-word priming in visual word recognition: Effects of positional letter frequency. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 951–964. http://dx.doi.org/10.1037/0096-1523.19.5.951

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103,* 518–565. http://dx.doi.org/10.1037/0033-295X.103.3.518

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception & Psychophysics, 45,* 189–195. http://dx.doi.org/10.3758/BF03210696

Grainger, J., O'Regan, J. K., Jacobs, A. M., & Segui, J. (1992). Neighborhood frequency effects and letter visibility in visual word recognition. *Perception & Psychophysics, 51,* 49–56. http://dx.doi.org/10.3758/BF03205073

Grainger, J., & Segui, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception & Psychophysics, 47,* 191–198. http://dx.doi.org/10.3758/BF03205983

Grainger, J., & van Heuven, W. J. B. (2003). Modeling letter position coding in printed word perception. In P. Bonin (Ed.), *Mental lexicon: Some words to talk about words* (pp. 1–23). Hauppauge, NY: NOVA Science.

Grossi, G., Savill, N., Thomas, E., & Thierry, G. (2012). Electrophysiological cross-language neighborhood density effects in late and early English–Welsh bilinguals. *Frontiers in Psychology, 3,* 408. http://dx.doi.org/10.3389/fpsyg.2012.00408

Guasch, M., Boada, R., Ferré, P., & Sánchez-Casas, R. (2013). NIM: A web-based Swiss Army Knife to select stimuli for psycholinguistic studies. *Behavior Research Methods, 45,* 765–771. http://dx.doi.org/10.3758/s13428-012-0296-8

Holcomb, P., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. Cognitive Neuroscience. *Journalism, 14,* 938–950.

Huntsman, L. A., & Lima, S. D. (1996). Orthographic neighborhood structure and lexical access. *Journal of Psycholinguistic Research, 25,* 417–429. http://dx.doi.org/10.1007/BF01727000

Huntsman, L. A., & Lima, S. D. (2002). Orthographic neighbors and visual word recognition. *Journal of Psycholinguistic Research, 31,* 289–306. http://dx.doi.org/10.1023/A:1015544213366

Jacobs, A. M., & Grainger, J. (1992). Testing a semistochastic variant of the interactive activation model in different word recognition experiments. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1174–1188. http://dx.doi.org/10.1037/0096-1523.18.4.1174

Johnson, N. F., & Pugh, K. R. (1994). A cohort model of visual word recognition. *Cognitive Psychology, 26,* 240–346. http://dx.doi.org/10.1006/cogp.1994.1008

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

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,* 563–580. http://dx.doi.org/10.1080/17470218.2012.658820

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 802–823. http://dx.doi.org/10.1037/a0030859

Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12,* 119–131. http://dx.doi.org/10.1016/S0022-5371(73)80001-5

Laxon, V., Coltheart, V., & Keating, C. (1988). Children find friendly words friendly too—Words with many orthographic neighbors are easier to read and spell. *The British Journal of Educational Psychology, 58,* 103–119. http://dx.doi.org/10.1111/j.2044-8279.1988.tb00882.x

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

Lupker, S. J., & Pexman, P. M. (2010). Making things difficult in lexical decision: The impact of pseudohomophones and transposed-letter nonwords on frequency and semantic priming effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1267–1289. http://dx.doi.org/10.1037/a0020125

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS One, 7,* e43230. http://dx.doi.org/10.1371/journal.pone.0043230

Mathey, S. (2001). Influence of orthographic neighborhood in word recognition tasks. *Canadian Journal of Experimental Psychology/Revue Canadienne de Pschologie Expérimentale, 55,* 1–23. http://dx.doi.org/10.1037/h0087349

McClelland, J., & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception. An account of basic findings. *Psychological Review, 88,* 375–407. http://dx.doi.org/10.1037/0033-295X.88.5.375

Midgley, K. J., Holcomb, P. J., van Heuven, W. J. B., & Grainger, J. (2008). An electrophysiological investigation of cross-language effects of orthographic neighborhood. *Brain Research, 1246,* 123–135. http://dx.doi.org/10.1016/j.brainres.2008.09.078

Paap, K., & Johansen, L. (1994). The case of the vanishing frequency effect—A retest of the verification model. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1129–1157. http://dx.doi.org/10.1037/0096-1523.20.6.1129

Paap, K., Newsome, S., McDonald, J., & Schvaneveldt, R. (1982). An activation verification model for letter and word recognition—The word-superiority effect. *Psychological Review, 89,* 573–594. http://dx.doi.org/10.1037/0033-295X.89.5.573

Peereman, R., & Content, A. (1995). Neighborhood size effect in naming—Lexical activation or sublexical correspondences. *Journal of Experimental Psychology-Learning Memory & Cognition, 21,* 409–421. http://dx.doi.org/10.1037/0278-7393.21.2.409

Peeters, D., Dijkstra, T., & Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals: RT and ERP effects. *Journal of Memory and Language, 68,* 315–332. http://dx.doi.org/10.1016/j.jml.2012.12.003

Perea, M., & Pollatsek, A. (1998). The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 767–779. http://dx.doi.org/10.1037/0096-1523.24.3.767

Perea, M., & Rosa, E. (2000). Repetition and form priming interact with neighborhood density at a brief stimulus onset asynchrony. *Psychonomic Bulletin & Review, 7,* 668–677. http://dx.doi.org/10.3758/BF03213005

Pollatsek, A., Perea, M., & Binder, K. S. (1999). The effects of "neighborhood size" in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1142–1158. http://dx.doi.org/10.1037/0096-1523.25.4.1142

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology, 52,* 1–56. http://dx.doi.org/10.1016/j.cogpsych.2005.06.001

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 720–732. http://dx.doi.org/10.1037/0096-1523.30.4.720

Rayner, K., & Liversedge, S. P. (2011). Linguistic and cognitive influences on eye movements during reading. In S. P. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 751–766). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780199539789.013.0041

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. New York, NY: Prentice Hall.

R Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2013). Using E-Z Reader to examine the concurrent development of eye-movement control and reading skill. *Developmental Review, 33,* 110–149. http://dx.doi.org/10.1016/j.dr.2013.03.001

Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: accounting for initial fixation locations and refixations within the E-Z Reader model. *Vision Research, 39,* 4403–4411. http://dx.doi.org/10.1016/s0042-6989(99)00152-2

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review, 16,* 1–21. http://dx.doi.org/10.3758/PBR.16.1.1

Scarborough, D., Gerard, L., & Cortese, C. (1984). Independence of lLexical access in vilingual word recognition. *Journal of Verbal Learning and Verbal Behavior, 23,* 84–99. http://dx.doi.org/10.1016/S0022-5371(84)90519-X

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

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime user's guide*. Pittsburgh, PA: Pschology Software Tools.

Sears, C. R., Campbell, C. R., & Lupker, S. J. (2006). Is there a neighborhood frequency effect in English? Evidence from reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1040–1062. http://dx.doi.org/10.1037/0096-1523.32.4.1040

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 876–900. http://dx.doi.org/10.1037/0096-1523.21.4.876

Slattery, T. J. (2009). Word misperception, the neighbor frequency effect, and the role of sentence context: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1969–1975. http://dx.doi.org/10.1037/a0016894

Snodgrass, J. G., & Mintzer, M. (1993). Neighborhood effects in visual word recognition: Facilitatory or inhibitory? *Memory & Cognition, 21,* 247–266. http://dx.doi.org/10.3758/BF03202737

Stone, G. O., & Van Orden, G. C. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 744–774. http://dx.doi.org/10.1037/0096-1523.19.4.744

Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language, 64,* 88–107. http://dx.doi.org/10.1016/j.jml.2010.08.006

Van Assche, E., Duyck, W., & Brysbaert, M. (2013). Verb processing by bilinguals in sentence contexts. *Studies in Second Language Acquisition, 35,* 237–259. http://dx.doi.org/10.1017/S0272263112000873

Van Assche, E., Duyck, W., & Hartsuiker, R. J. (2012). Bilingual word recognition in a sentence context. *Frontiers in Psychology, 3,* 174. http://dx.doi.org/10.3389/fpsyg.2012.00174

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,* 923–927. http://dx.doi.org/10.1111/j.1467-9280.2009.02389.x

van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language, 39,* 458–483. http://dx.doi.org/10.1006/jmla.1998.2584

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67,* 1176–1190. http://dx.doi.org/10.1080/17470218.2013.850521

Whitney, C. (2001). Position-specific effects within the SERIOL framework of letter-position coding. *Connection Science, 13,* 235–255. http://dx.doi.org/10.1080/09540090110083659

Wilkinson, G. S., & Robertson, G. (2006). *Wide Range Achievement Test 4 professional manual*. Lutz, FL: Psychological Assessment Resources.

Williams, C. C., Perea, M., Pollatsek, A., & Rayner, K. (2006). Previewing the neighborhood: The role of orthographic neighbors as parafoveal previews in reading. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1072–1082. http://dx.doi.org/10.1037/0096-1523.32.4.1072

Appendix A

Formula for the Used Measure of Orthographic Overlap

The Formula for the Corrected Levenshtein Distance (From Schepens, Dijkstra, & Grootjen, 2012)

$$OrthographicOverlap = 1 - \frac{Distance}{Length}$$

Distance = min (number of insertions, deletions and substitutions needed to edit target word into translation word); length = max (length of target word, length of translation word).

(*Appendices continue*)

Appendix B

Proficiency Scores

Table B1

Average Percentage Scores (Standard Deviations in Parentheses and Range in Brackets) on the LexTALE; Average Rating on the Self-Report Questionnaire (Standard Deviations in Parentheses)

Proficiency measure	Dutch	English	*t* L1-L2
LexTALE-score (%)	87.58 (7.03) [70.00–96.25]	73.04 (9.08) [57.50–88.75]	6.519***
Self-report			
Listening	4.9 (0.4)	4.0 (0.58)	5.141***
Speaking	4.87 (0.34)	3.5 (0.612	7.628***
Reading	4.9 (0.3)	3.93 (0.63)	5.604***
Writing	4.8 (0.48)	3.43 (0.72)	6.899***
Average	4.87 (0.29)	3.72 (0.47)	7.523***

Note. Participants had on average a higher proficiency for Dutch then English, both on the LexTALE, $t(29) = 7.518$, $p < .001$, and the average self-proficiency ratings, $t(29) = 10.891$, $p < .001$.
*** $p < .001$.

Table B2

Average Percentage Scores (Standard Deviations in Parentheses and Ranges in Brackets) on the LexTALE, Spelling Test, and Lexical Decision Task for the Bilingual and Monolingual Group in Experiment 2

Proficiency measure	Monolinguals	Bilinguals L1	Bilinguals L2	*t* L1-L2	*t* L1-mono
LexTALE- score (%)	91.07 (8.92) [71.25–100]	92.43 (6.34) [73.75–100]	75.63 (12.87) [51.25–98.75]	7.59***	.49
Spelling score (%)	80.78 (7.26) [73.81–90.48]	83.16 (7.80) [67.00–93.00]	69.92 (8.74) [52.00–83.00]	8.15***	.99
Lexical decision score (%)	77.89 (12.01) [54.61–95.23]	80.47 (5.45) [68.87–88.76]	56.75 (11.01) [38.46–75.86]	9.87***	.67

Note. The Dutch (L1) proficiency of the bilinguals was matched with the English proficiency of the monolinguals indicating that both groups were equally proficient in their first language. Neither the LexTALE ($t = .488$, $df = 22.254$, $p = .630$), the spelling test ($t = .989$, $df = 29.282$, $p = .331$), nor the lexical decision tasks ($t = .667$, $df = 17.092$, $p = .514$) yielded significant differences for these two groups performing in L1. The bilingual L2 LexTALE scores were significantly lower than their L1 scores ($t = 7.587$, df = 18, $p < .001$). The bilingual L2 Spelling scores were lower than the L1 scores ($t = 8.154$, $df = 18$, $p < .001$). The performance of the bilinguals on the classic lexical decision task was significantly better in L1 ($t = 9.873$, $df = 18$, $p < .001$) than in L2.
*** $p < .001$.

Appendix C

Stimulus Materials Used in Experiment 1

Dutch Words

Large Dutch *N*, Large English *N*. *Bons, borg, bril, dolk, hiel, klam, knie, oord, plek, rund, sein, spar, takt, tolk, vork, wolk, worp, woud, wrak, zalf*

Large Dutch *N*, Small English *N*. *Berg, beul, bouw, deun, dief, eter, fuik, kelk, kies, knal, kous, rede, snik, teug, touw, twee, unie, vals, verf, vies*

Small Dutch *N*, Large English *N*. *Brug, bult, draf, drie, fris, galg, hemd, heup, lach, meid, melk, munt, nota, pret, prik, smid, stug, vete, welp, wilg*

Small Dutch *N*, Small English *N*. *Akte, ambt, blad, erwt, ezel, gesp, gids, gips, inkt, joch, muts, ober, pech, pion, rots, snor, stro, toga, trui, veld*

English Words

Large Dutch *N*, Large English *N*. *Aunt, blue, farm, hawk, knit, left, loan, loud, maid, monk, moon, path, quit, shoe, suit, tool, verb, weak, wrap, zero*

Large Dutch *N*, Small English *N*. *Army, atom, bias, bird, diet, edge, germ, huge, butt, jerk, keen, knee, liar, lion, myth, noon, nude, obey, poem, poor*

Small Dutch *N*, Large English *N*. *Bath, bomb, busy, clue, coin, desk, dial, dirt, dish, firm, gray, hurt, iron, joke, lamb, limb, loss, milk, prey, rude*

Small Dutch *N*, Small English *N*. *Deny, duty, earl, envy, evil, folk, frog, guts, idol, kiss, okay, oral, oval, soup, true, twin, ugly, used, vein, view*

(*Appendices continue*)

Nonwords

Large Dutch *N*, Large English *N*. *Aril, aunk, blag, boul, boup, braf, bret, dris, duef, elap, fram, frip, furk, gonk, heud, jeef, knat, knub, koup, loem, meem, merd, mots, oram, peit, pern, piot, pral, pred, rama, sluf, sluk, snus, sols, stui, tess, trum, tult, vene, zork*

Large Dutch *N*, Small English *N*. *Alof, besp, bito, bouf, daus, drot, epoe, etel, feik, goep, grul, heut, irok, jees, jeul, jund, jurf, kalp, kelf, kerd, keun, loga, morp, muig, mups, nazz, noge, nont, noto, obel, oune, pris, puif, reug, reun, slen, smir, viem, woup, zuls*

Small Dutch *N*, Large English *N*. *Aute, bele, bulf, ceot, chah, cham, clet, dolo, drid, dulp, feul, foug, fran, genk, girs, jant, jero, jert, liry, lurd, lurp, lusp, naul, nirk, nudo, orim, pani, prad, prog, puet, raut, reud, rion, ruze, seto, snam, tirk, tran, vich, vorn*

Small Dutch *N*, Small English *N*. *Aler, anas, arns, aurd, baun, cafa, chof, deim, dilm, drio, durs, enip, fenk, feup, frig, frus, giep, heif, hilp, jalp, jofe, kach, kiot, knaf, luet, maup, moug, nige, omil, paby, ridi, siom, taur, torp, tuni, twol, unar, vota, zous, zuke*

Appendix D

Results of the Linear Mixed Effects Analysis of the Generalized Lexical Decision Data (Experiment 1)

Table D1

Estimates, Standard Errors, and t and p Values for the Fixed and Random Effects of the Final General Linear Mixed Effects Model for Reaction Times for Dutch Words

Dutch words	Estimate	*SE*	*t*	*p*
Fixed effects				
Intercept	2.806	.0156	179.90	<.001
Dutch *N* density	−.00006	.0014	−.04	.969
English *N* density	.00001	.0012	.01	.991
Dutch *N* frequency	−.0019	.0119	−.16	.876
English *N* frequency	−.0003	.0125	−.02	.981
Word frequency	−.0391	.0069	−5.66	<.001
Average bigram frequency	.0163	.0176	.93	.357
Orthographic overlap	−.0073	.0167	−.44	.664
	Variance	*SD*		
Random effects				
Word				
(Intercept)	.001	.032		
Subject				
(Intercept)	.003	.051		

$p < .1$.

(*Appendices continue*)

Table D2
Estimates, Standard Errors, and z and p Values for the Fixed and Random Effects of the Final General Linear Mixed Effects Model for Error Rate for Dutch Words

Dutch words	Estimate	*SE*	*z*	*p*
Fixed effects				
Intercept	−3.30	.47	−7	<.001
Dutch *N* density	.05	.05	1.19	.233
English *N* density	−.01	.04	−.19	.849
Dutch *N* frequency	−.22	.39	−.58	.56
English *N* frequency	.06	.39	−.58	.88
Word frequency	−1.22	.24	−5.14	<.001
Average bigram frequency	−.5	.55	−.93	.355
Orthographic overlap	−.15	.54	−.28	.781
English *N* Density × Word Frequency	.13	.07	1.93	.053
	Variance	*SD*		
Random effects				
Word				
(Intercept)	.443	.666		
Subject				
(Intercept)	.389	.624		

$p < .1$.

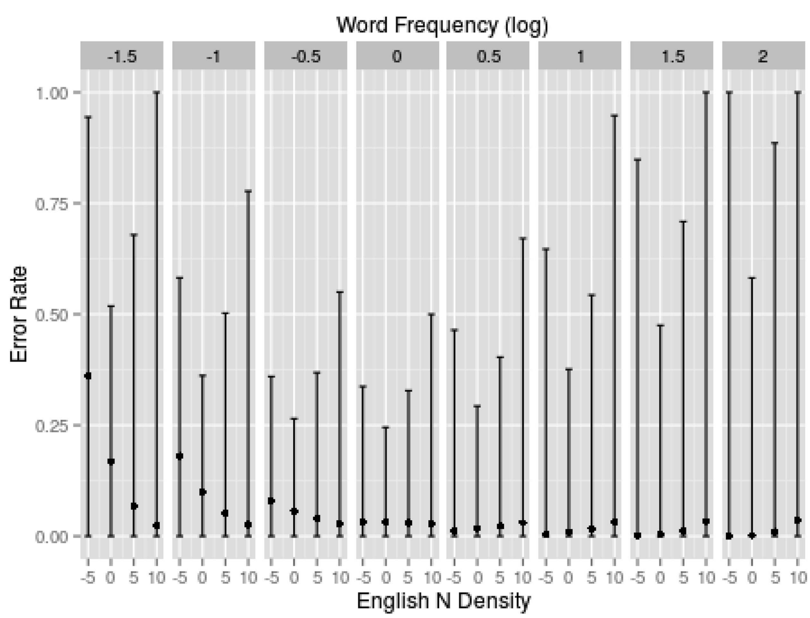

Figure D1. Error rate (on the *y*-axis) for Dutch words dependent on English *N* density (on the *x*-axis) dependent on word frequency of the word (panels) for a generalized lexical decision task.

(*Appendices continue*)

Table D3

Estimates, Standard Errors, and t and p Values for the Fixed and Random Effects of the Final General Linear Mixed Effects Model for Reaction Times for English Words

English words	Estimate	*SE*	*t*	*p*
Fixed effects				
Intercept	2.807	.0128	218.66	<.001
Dutch *N* density	.0013	.0011	1.15	.254
English *N* density	.00007	.0011	.07	.946
Dutch *N* frequency	.0087	.0079	1.10	.277
English *N* frequency	.0081	.087	.94	.353
Word frequency	−.0076	.0129	−.59	.556
Average bigram frequency	.0250	.0258	.97	.337
Orthographic overlap	−.0117	.0124	−.95	.349
Dutch *N* Density × Average Bigram Frequency	−.0134	.0066	−2.04	.046
English *N* density × Word Frequency	.0044	.0025	1.75	.086
English *N* Frequency × Word Frequency	−.0402	.016	−2.51	.015
English *N* Frequency × Average Bigram Frequency	−.0564	.0315	−1.79	.078
	Variance	*SD*		
Random effects				
Word				
(Intercept)	.0005	.021		
Subject				
(Intercept)	.0027	.052		

$p < .1$.

Table D4

Estimates, Standard Errors, and z and p Values for the Fixed and Random Effects of the Final General Linear Mixed Effects Model for Error Rates for English Words

English words	Estimate	*SE*	*z*	*p*
Fixed effects				
Intercept	−3.50	.40	−8.71	<.001
Dutch *N* density	.10	.04	2.32	.021
English *N* density	−.07	.05	−1.35	.177
Dutch *N* frequency	−.06	.33	−.18	.857
English *N* frequency	.228	.40	.57	.566
Word frequency	−.48	.43	−1.10	.270
Average bigram frequency	−1.36	.67	−2.01	.044
Orthographic overlap	.05	.50	.11	.914
English *N* Density × Average Bigram Frequency	−.29	.17	−1.65	.099
English *N* Frequency × Word Frequency	−1.01	.55	−1.84	.065
	Variance	*SD*		
Random effects				
Word				
(Intercept)	.542	.737		
Subject				
(Intercept)	.487	.698		

$p < .1$.

(*Appendices continue*)

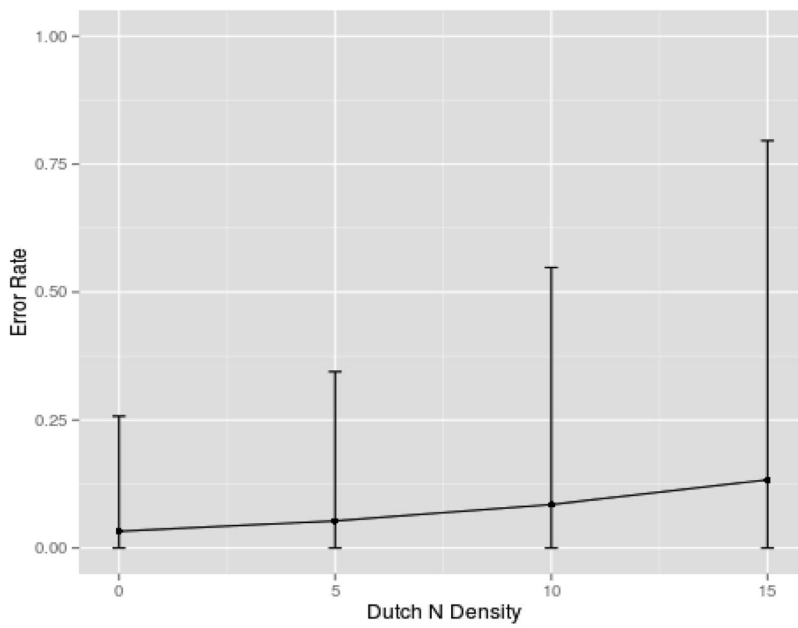

Figure D2. Error rates (on the *y*-axis) for English words dependent on Dutch *N* density (on the *x*-axis) in a generalized lexical decision task.

(*Appendices continue*)

Appendix E

Results of the Linear Mixed Effects Analysis of the Natural Reading Data (Experiment 2)

Table E1

Estimates, Standard Errors, and z and p Values for the Fixed and Random Effects of the Final General Linear Mixed Effects Model for Skipping Rates for Bilingual L1 Reading

Bilingual L1	Estimate	*SE*	*z*	*p*
Fixed effects				
(Intercept)	−.903	.111	−8.130	<.001
Dutch *N* density	.003	.007	.446	.655
English *N* density	.0005	.004	.111	.9111
Dutch *N* frequency	−.011	.037	−.289	.773
English *N* frequency	.050	.043	1.141	.254
Word frequency	.099	.021	4.736	<.001
Word length	−.227	.013	−17.078	<.001
Average bigram frequency	−.031	.067	−.456	.648
Orthographic overlap	−.028	.051	−.548	.583
Rank of occurrence	.0005	.0005	.911	.362
Dutch *N* Density × Word Frequency	−.011	.003	−3.266	.001
Dutch *N* Density × Word Length	−.007	.002	−2.918	.004
English *N* Frequency × Word Frequency	.078	.043	1.790	.074
	Variance	*SD*		
Random effects				
Word				
(Intercept)	.039	.198		
Subject				
(Intercept)	.205	.453		

$p < .1$.

Table E2

Estimates, Standard Errors, and t and p Values for the Fixed and Random Effects of the Final General Linear Mixed Effects Model for Single Fixation Durations for Bilingual L1 Reading

Bilingual L1	Estimate	*SE*	*t*	*p*
Fixed effects				
(Intercept)	2.306	.010	222.955	<.001
Dutch *N* density	−.0002	.0003	−.657	.511
English *N* density	.0001	.0003	.437	.662
Dutch *N* frequency	−.001	.003	−.396	.693
English *N* frequency	.005	.004	1.424	.155
Word frequency	−.010	.001	−7.262	<.001
Word length	.004	.001	6.354	<.001
Average bigram frequency	.001	.005	.243	.808
Orthographic overlap	−.002	.004	−.550	.583
Rank of occurrence	<−.0001	<.0001	−1.056	.291
Dutch *N* Density × Word Frequency	.001	.0002	3.595	<.001
	Variance	*SD*		
Random effects				
Word				
(Intercept)	.0003	.019		
Subject				
(Intercept)	.002	.043		

$p < .1$.

(*Appendices continue*)

Table E3

Estimates, Standard Errors, and t and p Values for the Fixed and Random Effects of the Final General Linear Mixed Effects Model for Gaze Durations for Bilingual L1 Reading

Bilingual L1	Estimate	*SE*	*t*	*p*
Fixed effects				
(Intercept)	2.334	.013	182.702	<.001
Dutch *N* density	−.001	.0004	−1.460	.145
English *N* density	.0001	.0004	.157	.875
Dutch *N* frequency	.001	.003	.285	.776
English *N* frequency	.004	.004	.901	.368
Word frequency	−.016	.002	−8.547	<.001
Word length	.008	.001	11.919	<.001
Average bigram frequency	−.001	.006	−.129	.897
Orthographic overlap	−.005	.004	−1.027	.305
Rank of occurrence	<−.0001	<.0001	−.662	.508
Dutch *N* Density × Word Frequency	.001	.0003	3.662	<.001
Dutch *N* Frequency × Word Frequency	.006	.003	2.017	.044
	Variance	*SD*		
Random effects				
Word				
(Intercept)	.0006	.025		
Subject				
(Intercept)	.0029	.054		

$p < .1$.

Table E4

Estimates, Standard Errors, t and p Values for the Fixed and Random Effects of the Final General Linear Mixed Effects Model for Total Reading Times for Bilingual L1 Reading

Bilingual L1	Estimate	*SE*	*t*	*p*
Fixed effects				
(Intercept)	2.381	.014	175.051	<.001
Dutch *N* density	−.0005	.0005	−.950	.342
English *N* density	.0001	.001	.203	.839
Dutch *N* frequency	.004	.004	.888	.375
English *N* frequency	−.002	.005	−.415	.678
Word frequency	−.022	.002	−9.553	<.001
Word length	.010	.001	12.979	<.001
Average bigram frequency	.003	.010	.350	.727
Orthographic overlap	.0004	.005	.067	.947
Rank of occurrence	<−.0001	<.0001	−1.369	.171
Dutch *N* Density × Word Frequency	.001	.0004	3.281	.001
Dutch *N* Frequency × Word Frequency	.009	.004	2.394	.017
Dutch *N* Frequency × Average Bigram Frequency	.027	.015	1.859	.063
English *N* Frequency × Average Bigram Frequency	−.030	.017	−1.754	.080
	Variance	*SD*		
Random effects				
Word				
(Intercept)	.001	.032		
Subject				
(Intercept)	.003	.057		

$p < .1$.

(*Appendices continue*)

Table E5

Estimates, Standard Errors, z and p Values for the Fixed and Random Effects of the Final General Linear Mixed Effects Model for Regressions for Bilingual L1 Reading

Bilingual L1	Estimate	*SE*	*z*	*p*
Fixed effects				
(Intercept)	−2.143	.098	−21.859	<.001
Dutch *N* density	.017	.008	2.155	.031
English *N* density	.008	.008	1.000	.317
Dutch *N* frequency	−.023	.068	−.333	.739
English *N* frequency	−.169	.087	−1.951	.051
Word frequency	−.060	.031	−1.919	.055
Word length	−.054	.013	−3.992	<.001
Average bigram frequency	.163	.126	−1.299	.194
Orthographic overlap	.057	.095	.601	.548
Rank of occurrence	.0004	.0008	.469	.639
	Variance	*SD*		
Random effects				
Word				
(Intercept)	.262	.512		
Subject				
(Intercept)	.146	.382		

$p < .1$.

Table E6

Estimates, Standard Errors, z and p Values for the Fixed and Random Effects of the Final General Linear Mixed Effects Model for Skipping Rates for Bilingual L2 Reading

Bilingual L2	Estimate	*SE*	*z*	*p*
Fixed effects				
(Intercept)	−1.074	.126	−8.527	<.001
Dutch *N* density	.006	.005	1.409	.159
English *N* density	.009	.003	2.730	.006
Dutch *N* frequency	.008	.034	.228	.820
English *N* frequency	.038	.030	1.269	.205
Word Frequency	.139	.178	7.813	<.001
Word Length	−.190	.010	−18.677	<.001
Average bigram frequency	.039	.100	.387	.698
Orthographic overlap	.120	.045	2.676	.007
Rank of occurrence	.002	.001	3.164	.002
English *N* Density × Average Bigram Frequency	.018	.009	1.986	.047
Dutch *N* Frequency × Average Bigram Frequency	.256	.127	2.022	.043
English *N* Frequency × Average Bigram Frequency	−.300	.134	−2.239	.025
	Variance	*SD*		
Random effects				
Word				
(Intercept)	.029	.171		
Subject				
(Intercept)	.291	.540		

$p < .1$.

(*Appendices continue*)

Table E7
Estimates, Standard Errors, t and p Values for the Fixed and Random Effects of the Final General Linear Mixed Effects Model for Single Fixation Durations for Bilingual L2 Reading

Bilingual L2	Estimate	*SE*	*t*	*p*
Fixed effects				
(Intercept)	2.336	.011	217.144	<.001
Dutch *N* density	−.002	.001	−2.508	.013
English *N* density	<.0001	.0003	.140	.888
Dutch *N* frequency	.002	.003	.689	.491
English *N* frequency	.002	.003	.842	.400
Word frequency	−.016	.001	−11.286	<.001
Word length	.002	.001	1.520	.129
Average bigram frequency	.012	.005	2.440	.015
Orthographic overlap	−.004	.004	−1.166	.244
Rank of occurrence	−.0001	.0001	−1.819	.069
Dutch *N* Density × Word Length	−.001	.0004	−2.736	.006
	Variance	*SD*		
Random effects				
Word				
(Intercept)	.0003	.018		
Subject				
(Intercept)	.002	.045		

$p < .1$.

Table E8
Estimates, Standard Errors, and t and p Values for the Fixed and Random Effects of the Final General Linear Mixed Effects Model for Gaze Durations for Bilingual L2 Reading

Bilingual L2	Estimate	*SE*	*t*	*p*
Fixed effects				
(Intercept)	2.375	.014	169.101	<.001
Dutch *N* density	−.002	.001	−1.871	.062
English *N* density	.0002	.0004	.406	.685
Dutch *N* frequency	.002	.004	.503	.615
English *N* frequency	.001	.003	.352	.725
Word frequency	−.018	.002	−10.867	<.001
Word length	.008	.001	5.487	<.001
Average bigram frequency	.017	.006	2.826	.005
Orthographic overlap	−.003	.004	−.742	.458
Rank of occurrence	−.0001	.0001	−1.019	.308
Dutch *N* Density × Word Length	−.001	.0005	−2.174	.030
	Variance	*SD*		
Random effects				
Word				
(Intercept)	.0009	.030		
Subject				
(Intercept)	.004	.059		

$p < .1$.

(*Appendices continue*)

Table E9

Estimates, Standard Errors, and t and p Values for the Fixed and Random Effects of the Final General Linear Mixed Effects Model for Total Reading Times for Bilingual L2 Reading

Bilingual L2	Estimate	*SE*	*t*	*p*
Fixed effects				
(Intercept)	2.376	.029	83.308	<.001
Dutch *N* density	−.003	.002	−2.066	.039
English *N* density	.0002	.0005	.514	.607
Dutch *N* frequency	.001	.005	.187	.852
English *N* frequency	.001	.004	.391	.696
Word frequency	−.028	.002	−14.460	<.001
Word length	.011	.002	6.041	<.001
Average bigram frequency	.020	.007	2.722	.007
Orthographic overlap	−.006	.005	−1.160	.246
Rank of occurrence	−.0002	.0001	−1.865	.062
Dutch *N* Density × Word Length	−.001	.001	−1.984	.048
	Variance	*SD*		
Random effects				
Word				
(Intercept)	.001	.037		
Subject				
(Intercept)	.004	.064		

$p < .1$.

Table E10

Estimates, Standard Errors, and z and p Values for the Fixed and Random Effects of the Final General Linear Mixed Effects Model for Regressions for Bilingual L2 Reading

Bilingual L2	Estimate	*SE*	*z*	*p*
Fixed effects				
(Intercept)	−2.093	.113	−18.530	<.001
Dutch *N* density	.010	.009	1.164	.244
English *N* density	.009	.006	1.390	.165
Dutch *N* frequency	.026	.063	.415	.678
English *N* frequency	−.004	.052	−.078	.937
Word frequency	−.096	.028	−3.447	<.001
Word length	−.066	.015	−4.361	<.001
Average bigram frequency	.087	.104	.843	.399
Orthographic overlap	−.101	.077	−1.312	.189
Rank of occurrence	−.003	.001	−2.002	.045
	Variance	*SD*		
Random effects				
Word				
(Intercept)	.203	.451		
Subject				
(Intercept)	.211	.459		

$p < .1$.

Received August 26, 2015
Revision received September 20, 2016
Accepted September 22, 2016 ■