Frequency Effects in Monolingual and Bilingual Natural Reading.

Uschi Cop[1], Emmanuel Keuleers[1], Denis Drieghe[2] and Wouter Duyck[1]

[1]Ghent University, Ghent, Belgium

[2]University of Southampton, Southampton, UK



*Author Note*

Uschi Cop, Department of Experimental Psychology, Ghent University; Emmanuel Keuleers, 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 supported by a grant from the FWO (Fonds voor Wetenschappelijk Onderzoek).

Correspondence concerning this article should be addressed to Uschi Cop, Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium. E-mail: uschi.cop@ugent.be. Tel: 0032485084481.

**Abstract**

This paper presents the first systematic examination of the monolingual and bilingual frequency effect (FE) during natural reading. We analyzed single fixations durations on content words for participants reading an entire novel. Unbalanced bilinguals and monolinguals show a similarly sized FE in their mother tongue (L1), but for bilinguals the FE is considerably larger in their second language (L2) than in their L1. The FE in both L1 and L2 reading decreased with increasing L1 proficiency, but it was not affected by L2 proficiency. Our results are consistent with an account of bilingual language processing that assumes an integrated mental lexicon with exposure as the main determiner for lexical entrenchment (Diependaele, Lemhöfer, & Brysbaert, 2013; Gollan et al., 2008). This means that no qualitative difference in language processing between monolingual, bilingual L1 or bilingual L2 is necessary to explain reading behavior. We specify this account and argue that not all groups of bilinguals necessarily have lower L1 exposure than monolinguals do and, in line with Kuperman and Van Dyke (2013), that individual vocabulary size and language exposure change the accuracy of the relative corpus word frequencies and thereby determine the size of the FE's in the same way for all participants.

Although word recognition and production are both very complex processes influenced by a wide range of variables, the frequency of occurrence of a word in a language is by far the most robust predictor of language performance (Brysbaert et al., 2011; Murray & Forster, 2004). In both word identification (e.g. Rubenstein, Garfield, & Millikan, 1970; Scarborough, Cortese, & Scarborough, 1977) and word production tasks (e.g. Forster & Chambers, 1973; Monsell, Doyle, & Haggard, 1989) high frequency words are processed faster than low frequency words. This observation is called the word frequency effect (FE), and it is one of the most investigated phenomena in (monolingual) psycholinguistics.

Multiple language models of comprehension (e.g. Dijkstra & Van Heuven, 2002; McClelland & Rumelhart, 1981; Morton, 1970) explain frequency effects using implicit learning accounts. These state that repeated exposure to a certain lexical item raises this item's baseline activation in proportion to their distance to the activation threshold, so that lexical selection of that particular word is faster during recognition (e.g. Monsell, 1991). The maximal speed of lexical access is limited, so once a word has received a certain amount of exposure, no more facilitation will be expected when there is additional exposure to that particular item (Morton 1970).

In the visual domain, word recognition speed increases with the logarithm of word corpus frequency (Howes & Solomon, 1951). A certain number of additional exposures to a low frequency word will result in a large decrease of its lexical access time, while the same number of additional exposures to a high frequency word will result in a much smaller decrease of its lexical access time. This particular characteristic of the relationship between word frequency and processing time causes the size of the frequency effect to be modulated by language exposure.

Bilinguals offer an interesting opportunity to study the relationship between exposure

and lexical access, because of the within-subject difference in language exposure for L1 and L2. We will examine the effect of word frequency in bilingualism on the basis of new natural reading data collected for English monolinguals and Dutch-English bilinguals. We will start by examining the literature on individual differences in the word frequency effect and discuss the relation of these findings to the frequency effect in bilinguals. Following Kuperman and Van Dyke (2013), we will formulate an account of exposure-related differences in the effect of corpus word frequency that originates in the statistical characteristics of word frequency distributions.

**Individual Differences in the FE**

The collection and evaluation of frequency norms based on text corpora is central to psycholinguistic research (e.g., Brysbaert & New, 2009; Keuleers, Brysbaert & New, 2010; Van Heuven, Mandera, Keuleers, & Brysbaert, 2014). The number of exposures to a certain word is often operationalized as the count of word occurrences in language corpora like the Subtlex database (Keuleers et al., 2010). Mostly, corpus frequencies are expressed as *relative* values because these can be used independent of corpus size. These *objective* corpus word frequencies are supposed to reflect the average number of exposures to certain words of an experienced reader. While corpus word frequencies are a tremendously useful proxy measure for relative exposure, it should not be forgotten that the relative frequency of a word in a text corpus is not necessarily equal to the relative frequency of exposure to that word for a particular individual.

Solomon and Howes (1951) already emphasized that word counts from text corpora are based on an arbitrary sample of the language and that there may be individual variation in the relative frequency of exposure to specific words. In other words, corpus word frequencies may under- or overestimate subjective word frequencies, which can lead to a difference in the size of the FE when corpus word frequencies are used in analyses. The differences in the FE

size would disappear when a measure of actual exposure or subjective frequency (e.g., Connine, Mullennix, Shernoff, & Yelen, 1990; Gernsbacher, 1984) is used. Still, in experiments where words from different semantic domains (for example tools or clothing) are used as stimuli, such differences in relative frequency would in principle not lead to systematic differences in the frequency effect between individuals. This is because differences in subjective frequency in particular semantic categories would be cancelled out by the use of stimuli from multiple domains.

Next to the possibility of individual differences in the relative frequency for *specific* words due to differences in experience with a specific vocabulary, it is possible that individuals, who are at different stages in the language acquisition process, or, more broadly, have a differing amount of total language exposure, may have different relative frequencies for words. For this reason, some studies have used familiarity ratings of words as a more accurate reflection of the actual exposure to certain words for a specific group of readers (e.g. Balota, Pilotti, & Cortese, 2001; Kuperman & Van Dyke, 2013). Balota et al. (2001) observed that these subjective norms explained unique variance above and beyond objective corpus frequencies for lexical decision and naming tasks. Kuperman and Van Dyke (2013) confirm that objective corpus frequencies are particularly poor estimates and systematically overestimate the subjective frequencies for low frequent words for individuals with smaller vocabularies.

**Bilingual FE's**

Most research on the frequency effect in language processing has focused on monolingual participants, while more than half of the world population, the 'default' person, is bilingual or multi-lingual. Taking into account that bi- or multilingualism is at least as widespread as monolingualism, it is important to assess how exposure to L1 or L2 affects bilingual person language processing. This is not straightforward because there is now a

consensus that L1 and L2 constantly interact during visual word recognition (e.g. Duyck, Van Assche, Drieghe, & Hartsuiker, 2007;Van Assche, Duyck & Hartsuiker, 2012). These cross-lingual interactions strongly suggest the existence of a unified bilingual lexicon with parallel activation for all items in that lexicon, with items competing for selection within and across languages (for a more comprehensive overview of the evidence for an integrated bilingual lexicon see Brysbaert & Duyck, 2010 and Dijkstra & Vanheuven, 2002). Not only does L1 knowledge influence L2 lexical access, but the knowledge of an L2 also changes L1 visual word recognition (e.g. Van Assche, Duyck, Hartsuiker & Diependaele, 2009). Because these interactions occur in both directions, it is not only important to assess the differential influence of word exposure on lexical access for L1 and L2 reading, but also the possible differences between the frequency effect for monolinguals and bilinguals in L1.

Although the individual differences in frequency distribution described above are relevant for monolingual research, this is even more the case for bilingual research. The integrated bilingual lexicon will contain on average more lexical items than that of a monolingual. For advanced learners of an L2, who have a lexical entry for almost all concepts, we can assume that they would have almost double the amount of words in their lexicon. Inspired by observations of bilingual disadvantages in production tasks (e.g. Ivanova & Costa, 2008; Gollan, Montoya, Fennema-Notestine & Morris, 2005, Gollan et al., 2011), the weaker links theory (Gollan & Silverberg, 2001;Gollan & Acenas, 2004; Gollan et al. 2008, 2011) was proposed. This theory posits the idea that bilinguals necessarily divide their language use across two languages, resulting in lower exposure to all of the words in their lexicon, including L1 words. The lexical representations of bilinguals in both languages will have accumulated less exposure than the ones in the monolingual lexicon. Over time, this pattern of use would lead to weaker links between semantics and phonology for bilinguals, relative to monolinguals (Gollan et al. 2008).

Diependaele et al. (2013) generalize the weaker links account and assume a decrease in lexical exposure for bilinguals, and suggest that this can result in a reduced lexical entrenchment either by reduced lexical precision of those representations (e.g. Perfetti, 1992, 2007), or by reduced word-word inhibition or weaker integration between phonological and semantic codes (e.g. Gollan et al., 2008, 2011).

In short, the mere knowledge of a second language (and being exposed to its words) will reduce the lexical entrenchment of the first language, because this language will receive less exposure. Gollan et al. (2008) suggest a direct relationship between the weaker links and the frequency effect. They make the explicit hypothesis that bilinguals should have a larger frequency effect than monolinguals because a) bilinguals have used words in each language less often than monolinguals have and b) increased use leads to increased lexical accessibility only until a certain ceiling level of exposure, meaning that low frequency words should be more affected by differences in degree-of-use than high frequency words. From this hypothesis, we can also predict that in the case of unbalanced bilinguals, for whom L2 exposure is lower than the L1 exposure, the L2 FE's will also be larger than the L1 FE's. We support the idea posited by the weaker links account that differential FE's in the bilingual domain can be explained without assuming qualitatively different language processing for bilinguals compared to monolinguals and aim to specify the hypotheses put forward by the weaker links account (Gollan et al., 2008).

**Word Frequency Distribution**

Because of the logarithmic relationship between corpus word frequency and lexical access time, it is customary to use logarithmically transformed corpus word frequencies in any analysis where word frequency is a variable in the model. This transformation changes the functional relationship between corpus word frequency and lexical access time from a

logarithmic one to a linear one (See the upper and middle panel of Figure A.1 in Appendix A for an illustration).

When detecting changes in the size of the FE related to language exposure, it is important to note that when these transformed corpus word frequencies are used, the *size* of the word frequency effect is not affected by absolute exposure. In other words, while a participant who has more exposure to a certain language will be faster to process words in that language than a participant who has little exposure to that language, an analysis based solely on transformed corpus word frequency would predict that the difference in processing times for high frequency and low frequency words, in other words the FE, is the same for both participants. Still another way of putting it is that when *x* and *y* are untransformed relative corpus word frequencies (for instance *x*=100 per million and *y*=1 per million), then for a participant who has been exposed to 100 million words the difference in absolute exposure between *x* and *y* is 9,900 (10,000-100) while for a participant who has been exposed to 10 million words, the difference is 990 (1000-10), which would lead to larger frequency effect for the participant with more exposure. When logarithmically transformed frequencies are used, for the participant with exposure to 100 million words the difference between *x* and *y* is 2 (log10 (10,000) – log10 (100) = 4 - 2 = 2), while for the participant with exposure to 10 million words, the difference between *x* and *y* is also 2 (log10 (1000) - log10 (10) = 3 – 1 = 2).

Another element to consider is that word frequency distributions are fundamentally different from normal distributions, which psychologists are used to working with. For instance, a typical characteristic of normal distributions is that the mean of a sample is an estimate that could be higher or lower than the population average and that gets more and more precise as the sample size grows. This characteristic is not shared with word frequency distributions. Instead, one of the characteristics of word frequency distributions is that the

mean predictably increases as the sample, or the corpus, grows (Baayen, 2001). Importantly, Kuperman and Van Dyke (2013) show that relative word frequency is also related to the corpus size. They demonstrate that as corpus size grows, the relative frequency of low frequency words increases while the relative frequency of high frequency words stays almost constant (See Table 1). By dividing words in ten frequency bands, they show that words in the lowest frequency band (1) have an estimate of relative frequency that is twice as large in a corpus of 50 million words than in a corpus of 5 million words (ratio: 2.234); relative frequency estimates for words in the highest frequency band (10), on the other hand, were nearly equivalent (ratio: 1.003).

Table 1

*The ratio of a word's relative frequency in the 50-million token SUBTLEX corpus to its relative frequency in a sample of 5 million tokens (Relative frequencies averaged over 1000 samples). Taken from Kuperman & Van Dyke (2013).*

| Frequency class | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Between-sample ratio | 2.234 | 2.083 | 1.672 | 1.344 | 1.020 | 1.020 | 0.996 | 0.998 | 1.012 | 1.003 |

*Note.* Ratios are averaged per frequency class (1 = lowest frequency; 10 = highest frequency) and are based on a pool of 500 words, with 50 words per frequency class.

It is precisely this characteristic of word frequency distributions that is overlooked in the analysis of the effect of word frequency. If the evolution of relative word frequency with more exposure follows a trajectory that is analogous to the evolution of relative frequency with increase in corpus size, this alone can account for differences in the size of the FE. On these grounds, an interaction of proficiency and corpus frequency is *expected,* but it should not be attributed to qualitative differences between poor and good readers, or between a categorical difference between monolinguals and bilinguals. As we already mentioned, when

assuming lower exposure to all items in the lexicon and using raw corpus word frequencies in the analyses, a larger FE *slope* is expected. When we log transform these word frequencies we do not necessarily expect a larger FE slope as long as the ratios between the relative frequencies stay the same. The importance of changes for low frequency words but not for high frequency words is exactly what a logarithmic transformation accounts for; differences in the frequency effect due to a lower exposure to all words in the lexicon should not be found if a logarithmic transformation is used and if there are no changes in *relative* word frequency. However, if relative subjective frequencies do not stay constant, this difference should lead to a difference in the size or slope of the frequency effect when a logarithmic transformation is applied to the frequencies. It should be noted that the reasoning that differences in the size of the frequency effect are only due to the logarithmic relationship between word frequencies and word processing times, is therefore incomplete (e.g., Duyck, Vanderelst, Desmet & Hartsuiker, 2008; Schmidtke, 2014).

**Language exposure**

The weaker links theory is consistent with the individual differences account of Kuperman and Van Dyke (2013) in the sense that differences in the FE are attributed to the degree of exposure rather than to qualitative differences originating from the acquisition of multiple languages. However, the weaker links theory makes the general claim that a) there is an overall lower (absolute) exposure to language for bilinguals than for monolinguals and b) that this results in a larger FE for bilinguals.

A pure exposure-based account leaves open the possibility that bilinguals may have the same degree of exposure to one (or, in principle, more) of their two languages as monolinguals have and this account can specify the exact locus of the modulation of the size of the FE, namely that it arises from differences in ratios of high and low relative frequencies for individuals with different levels of exposure.

As already discussed, language exposure should be an important determinant of the shape and size of the FE. It is therefore of vital importance to have a good measurement for this variable. Most experiments use subjective measures like questionnaires to assess exposure, some try to quantify exposure by measuring language proficiency. Because there is a direct relation between the obtained measure of vocabulary size and the degree of exposure (e.g., Baayen, 2001), we prefer the use of a vocabulary test to assess language proficiency. By using vocabulary growth curves (see Figure 1), we can see a tight relationship between language exposure (word tokens on the x-axis) and vocabulary size (word types on the y-axis). Word tokens are counted as every word in a language corpus, including repetitions and word types are unique words. As the number of word tokens grows, so does the number of word types.

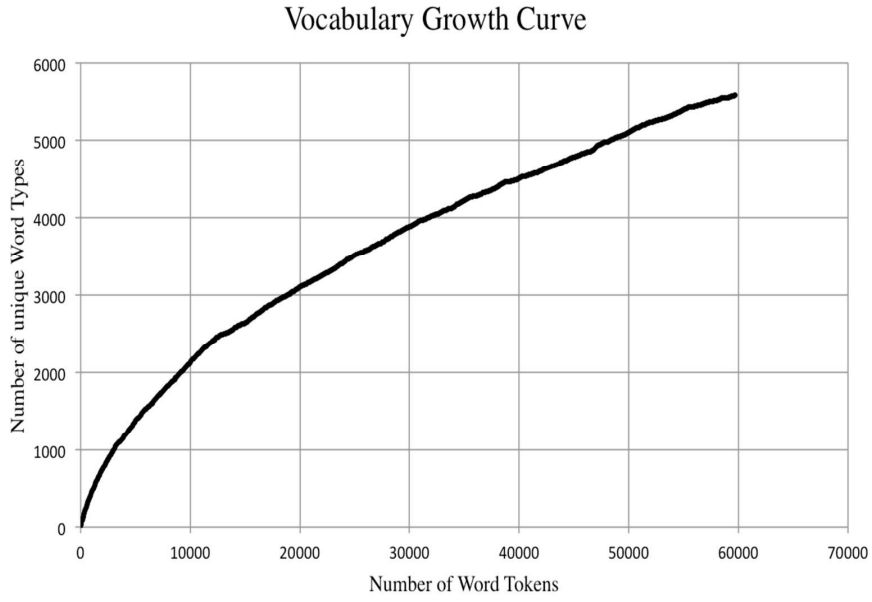


*Figure 1*. An example of a vocabulary growth curve. This plot shows the number of word tokens encountered (on the x-axis) and the amount of encountered word types (on the y-axis) when reading the Dutch version of the novel 'A mysterious affair at Styles'.

When vocabulary size is small, the probability that the next encountered word will be a hitherto unseen type is large, but as exposure grows the probability that the next word will be a new type decreases. As a result, to double vocabulary size requires much more than twice the amount of exposure. Concurrently, the more exposure one has, the smaller the increase in vocabulary size that is associated with additional exposure. Assuming no large differences in the complexity of material that one is exposed too, a similar vocabulary score indicates similar exposure and an increase in vocabulary scores indicates a higher degree of exposure. For subjects with an equal but very high vocabulary score, it becomes more uncertain that they have the exact same amount of language exposure. Nevertheless, on the whole, when participants have equal proficiency scores, we do not expect differential FE's, because language exposure should be quite similar.

Kuperman and Van Dyke (2013) note that robust interactions between language proficiency and word frequency have been found in a wide range of studies concerning individual reading differences: More proficient readers showed a smaller frequency effect on reaction times. (For examples see Chateau & Jared, 2000 and Diependaele et al. 2013)

Although this is indeed a robust finding, it must be noted that some authors have claimed that this finding might be an artifact of the base-rate effect (Butler & Hains, 1979; Faust et al. 1999; Yap et al., 2012). The base-rate effect is the observation that the magnitude of lexical effects correlates positively with reaction latencies. This would mean that the larger frequency effects for participants with a lower language proficiency score would be mainly due to the fact that their reaction times are longer than higher skilled participants. However, Kuperman and Van Dyke (2013) showed that the interaction between word frequency and language skill is still present after z-transforming reaction times per subject, thus eliminating any kind of base rate effect.

**Bilingual research**

As shown by analyses that find larger frequency effects for L2 word recognition when word frequencies are log-transformed (Diependaele et al., 2013; Duyck et al., 2008; Lemhöfer et al., 2008; Whitford & Titone, 2011), exposure does have a systematic relation with the size of the word frequency effect that cannot be accounted for by the logarithmic relation between word processing times and word frequencies alone.

In the tradition of the weaker links account and as evidence for reduced lexical entrenchment for bilinguals compared to monolinguals, the bilingual FE has been compared with the monolingual FE. Indeed, when we look at the experimental findings, we find that when proficiency is equal across groups, no differences in the size of the FE are found: Gollan et al. (2011) found a similar FE in an English lexical decision and a sentence reading task for balanced Spanish-English bilinguals as for English monolinguals; Duyck et al.'s (2008) study did not find a difference between the L1 FE of unbalanced Dutch-English bilinguals and the FE of English monolinguals in lexical decision times either. The studies that did find a larger bilingual FE used bilingual participants with lower proficiency, and thus lower exposure, for the tested language than the monolinguals; also the tested language was acquired later than their other language. This means that the corpus frequencies were probably overestimated for the lower frequent words for the bilingual group, inflating reaction times for the low range. For example, Lehtonen et al. (2012) found a larger FE in a Finnish lexical decision task for balanced Finnish-Swedish bilinguals than for Finnish monolinguals. When we look at the Finnish proficiency scores we see that the bilinguals scored significantly lower than the monolinguals. Also, Lemhöfer et al. (2008) found a larger FE for different groups of bilinguals in English, their L2, than for English monolinguals in a word identification task. Gollan et al. (2011) showed that the L2 FE for Dutch-English

bilinguals in a lexical decision task was larger than for English monolinguals. Naturally the bilinguals had less exposure to their L2 than the monolinguals had for their L1. These two last studies used raw frequencies.

In short, the results of all of these studies are congruent with our expectations, namely that language exposure could account for all differences found between bilingual and monolingual FE's.

Indeed, Diependaele et al. (2013) reinvestigated Lemhöfer et al.'s (2008) English word identification times, using log-transformed word frequencies. They hypothesized that target language proficiency is the determining factor for identification times both in the L1 of the monolinguals and in L2 of the bilinguals, without a qualitative difference between L1 and L2 processing. They found a larger FE for bilinguals' word identification times in L2, than for the monolinguals' word identification times in L1. When they added target language proficiency in their model, the FE modulation by group was no longer significant. Higher target language proficiency reduced the size of the FE and this effect was the same for both groups.

As already discussed, within the unbalanced bilingual's lexicon, we assume lower exposure to L2 words than to L1 items. For this reason, a larger FE for bilinguals reading in L2 is expected compared to reading in L1, even when word frequencies are log-transformed. Duyck et al.'s (2008) data confirm this hypothesis. They used an English and Dutch lexical decision task to test Dutch-English unbalanced bilinguals. Using a dichotomous (low vs. high) corpus frequency manipulation, they found that the L2 FE is about twice as large as the L1 FE. Whitford and Titone (2011) used eye movement measures of L1 and L2 paragraph reading of unbalanced English-French and French-English bilinguals. Bilinguals reading in L2 showed larger FE's in gaze durations and total reading time than they did in L1. On top of that, they found a modulation of the L1 and L2 FE by L2 exposure. Bilinguals with a higher

L2 exposure showed a smaller FE when reading in L2 than the bilinguals with a lower L2 exposure.

In sum, the findings of FE modulation in the bilingual field are compatible with the account that Kuperman and Van Dyke (2013) propose for individual differences in FE's for monolingual participants. Quantitative differences between language exposure, resulting in a different ratio of relative frequencies for low compared to high exposure items, can account for the differences between bilingual and monolingual language processing, but also for the differences found within groups for L1 and L2 processing.

**This Study**

Our study is the first to investigate the difference between the first acquired and dominant L1 FE of unbalanced bilinguals, and the monolingual FE in natural reading. Duyck et al.'s (2008) study compared the same groups (Dutch-English bilinguals and English monolinguals) but merely used an isolated word recognition task. This lexical decision task contained a limited number of 50 target words (25 low frequency and 25 high frequency words) per participant and provided only a small amount of data per participant. On top of that, the *isolated-word method* used in their experiment, represent an oversimplification of the natural way in which words are encountered, limiting ecological validity. When reading in a natural context, word processing takes place while other language processing is going on, e.g. integrating words in context, parsing of syntax, etc. Also a lexical decision task involves a behavioral response, which might require mental processes or strategic factors that are normally not associated with reading.

Until now, only 2 studies compared the frequency effects for L1 and L2 visual word recognition (Duyck et al., 2008; Whitford & Titone, 2011). In Whitford and Titone's (2011) study, comparing L1 with L2 FE's, participants read 2 paragraphs each containing only about 50 content words. In our study, the largest bilingual eye tracking data corpus (Cop, Drieghe

& Duyck, 2014), bilingual and monolingual participants read a whole novel containing around 29 000 content words. Not only is this a much larger and thus more generalizable, assessment of bilingual reading, it is also an even more naturalistic setting than paragraph reading, since people often read text in the context of a coherent story.

This study also attempts to resolve a concern we have with most cited studies, namely a poor measurement of L2 proficiency and a lack of assessment of L1 proficiency. We follow Luk and Bialystok (2014) in their assertions that there are multiple dimensions of bilingualism and follow their recommendation to use both methods of subjective and objective proficiency assessments. By triangulating these different measurements, we calculated a composite proficiency score for both L1 and L2 language proficiency. Both the individual measurements as this composite score can then be used to assess differences in proficiency between the tested groups. The way this composite score was calculated is described in the method section.

Most studies on the bilingual FE use self-reported L2 language exposure as a measure of proficiency (cf. Whitford & Titone, 2011) or do not measure the language proficiency of their participants at all (cf. Duyck et al. 2008). For our analyses we use the LexTALE scores because this test has been validated as an indication of vocabulary size, a central concept in this study. Kuperman and Van Dyke (2013) explain the different individual FE's precisely by vocabulary size. On top of that, the LexTALE score has been used in multiple bilingual studies, ensuring an easy comparison between the results and replication of the effects of this score.

Interestingly, no study has ever investigated the differential effects of L1 vs. L2 proficiency for bilinguals on frequency effects. This is the first study to even add L1 proficiency to the analysis of the FE of bilingual reading data. Neither Whitford and Titone (2011), Duyck et al. (2008) nor Diependaele et al. (2013) used this variable in analyzing the

bilingual data, while it is expected that the proficiency of L1, which is an indication of lexicon size and exposure, is of importance to the actual frequencies of the word forms in the bilingual lexicon.

Concerning proficiency, the weaker links account (Gollan & Acenas, 2004) always assumed a trade off between the two scores: A high L2 exposure will imply a lower L1 exposure. The proliferation of lexical items in bilinguals should necessarily lead to a lower exposure to other items and eventually to weaker links between lexical representations and their word forms. For unbalanced bilinguals we assume that the mentioned trade-off between L1 and L2 exposure will be much more unclear. We might even expect that the L1 and L2 proficiency scores should correlate positively with each other, when we assume that innate language aptitude plays a role in language acquisition. Many studies in the monolingual domain have found that participants with increased vocabulary size show a reduced response time and a higher accuracy rate in lexical decision tasks (Yap, Balota, Sibley, & Ratcliff, 2012) for both familiar and unfamiliar words (e.g. Chateau & Jared, 2000). On top of that Perfetti, Wlotko and Hart (2005) observe that individuals who are better at comprehending text or have a higher reading skill, require fewer exposures to learn new words. This means that a person with a large L1 proficiency score, will be faster at establishing a connection between a new word form and its meaning (Perfetti et al., 2005) and might thus be more likely to also have a larger L2 proficiency score.

For monolingual L1 and bilingual L1 reading, we expect that L1 proficiency should have a large influence on the size of the frequency effect, with smaller L1 FE's for higher L1 proficiency. The relationship between L1 proficiency and the FE should be the same for both groups. For the comparison between the bilingual L2 reading, L1 proficiency might have a similar effect on the size of the FE, within the vocabulary size rationale discussed above. Given the robust effects of L2 proficiency on the size of the FE in previous studies, we might

expect this effect to persist even in the presence of L1 proficiency. If it does, a higher L2 proficiency is expected to reduce the FE in L2 reading but not in L1 reading.

## Method

This method section is partly taken from Cop, Drieghe, and Duyck (2014) because the data in this analysis is a subset from a large eye movement corpus described in Cop et al. (2014).

### Participants

Nineteen unbalanced bilingual Ghent University and fourteen monolingual Southampton University undergraduates participated either for course credit or monetary compensation. The bilingual participants' dominant language was Dutch, their second language English. They had a relatively late L2 age of acquisition (mean=11 [2.46]). The monolingual participants had knowledge of only one language: English. Bilingual participants completed a battery of language proficiency tests including a Dutch and English spelling test (GLETSHER and WRAT4), the LexTALE (Lemhöfer & Broersma, 2011) in Dutch and English, a Dutch and English lexical decision task (for results see Table B.1 in Appendix B) and a self-report language questionnaire (based on the LEAP-Q, Marian, Blumenfeld, & Kaushanskaya, 2007). Monolinguals completed an English spelling test, the English LexTALE and an English lexical decision task. We calculated a composite L1 and L2 proficiency score by averaging the score on the spelling test, the score on the LexTALE and the adjusted score of the lexical decision task. This composite score and the LexTALE scores show that bilinguals score significantly higher on L1 proficiency than they do on L2 proficiency, and that the bilinguals and monolinguals are matched on L1 proficiency. The LexTALE score is used in the analysis. Participants had normal or corrected-to-normal

vision. None of the participants reported to have any language and/or reading impairments. For detailed scores see Table B.1 in Appendix B.

**Materials**

The participants were asked to read the novel "The mysterious affair at Styles" by Agatha Christie (Title in Dutch: "De zaak Styles"). This novel was selected out of a pool of books that was available via the Gutenberg collection. The books were judged on length and difficulty, indicated by the frequency distribution of the words that the book contained. We selected the novel whose word frequency distribution was the most similar to the one in natural language use (Subtlex database). The Kullback–Leibler divergence was used to measure the difference between the two probability distributions (Cover and Thomas, 1991).

In English, the book contains 56 466 words and 5 212 sentences (10.83 words per sentence); in Dutch it contains 60 861 words and 5 214 sentences (11.67 words per sentence). The average word length in Dutch was 4.51 characters and 4.27 characters in English. The average word log frequency of the content words in the book was 3.82 for both books. Only the non-cognate content words of the novel were analyzed. The Dutch novel contained 30 817 content words and the English novel 28 108. From those words, 5 207 Dutch and 4 676 English words were individually distinct types. This means that each participant read ± 5000 different content words. See Table 2 for the description of the content words in Dutch and English. Although both word frequency and word length show minor differences across languages, these variables will be included in the higher order interactions in our linear mixed model.

Table 2

*Summary of the characteristics of the content non-cognate words of the novel: Number of Words, Average Content Word Frequency and Average Word Length. Standard deviations are in brackets.*

| | Dutch | English |
|---|---|---|
| Number of Words | 22 919 | 20 695 |
| Average Content Word Frequency | 3.74 [1.23] | 3.79 [1.20] |
| Average Word Length | 5.95 [2.56] | 5.47 [2.23] |

**Apparatus**

The bilingual and monolingual eye movement data were recorded with the EyeLink 1000 system (SR-Research, Canada) with a sampling rate of 1 kHz. Reading was binocular, but eye movements were recorded only from the right eye. Text was presented in black 14 point Courier New font on a light grey background. The lines were triple spaced and 3 characters subtended 1 degree of visual angle or 30 pixels. Text appeared in paragraphs on the screen. A maximum of 145 words was presented on one screen. During the presentation of the novel, the room was dimly illuminated.

**Procedure**

Each participant read the entire novel in four sessions of an hour and a half, except for one bilingual participant who only read the first half of the novel in English. The other bilinguals read half of the novel in Dutch, the other half in English. The order was counterbalanced.

The participants were instructed to read the novel silently while the eye tracker recorded their eye movements. It was stressed that they should move their head and body as little as possible while they were reading. The participants were informed that they would be presented with multiple-choice questions about the contents of the book after each chapter.

This was done to ensure that participants understood what they were reading and paid attention throughout the session.

The text of the novel appeared on the screen in paragraphs. When the participant finished reading the sentences on one screen, they were able to press the appropriate button on a control pad to move to the next part of the novel.

Before starting the practice trials, a nine-point calibration was executed. After this, the calibration was done every 10 minutes, or more frequently when the experiment leader deemed necessary.

## Results

Words that had an orthographically overlapping translation equivalent in the other language were categorized as identical cognates and were excluded for the frequency analysis (Dutch: 8.1%, English: 13.7%). The first and last word on a line were excluded from the analysis (Dutch: 18.8%, English: 16.9%), because their processing times also reflect sentence wrap-up effects (e.g. Rayner et al., 1989).

In Table 3 we report the average single fixation duration, gaze duration, skipping rates and the frequency effects for monolinguals and bilinguals reading in L1 and L2. A single fixation duration is the duration of the fixation on target words that were fixated only once. The gaze duration is the time spent on the word prior to moving the eye towards the right of that word. This means that first pass refixations are included in this measure. The skipping rate of a word is the likelihood that that word will be skipped the first time it is encountered. For the sake of visualization, words were median-split by frequency to create a low and high frequency set.

In this article we report the analysis of the single fixation durations. We prefer this measure because eye movements are complex and can reflect different processes. For example, first fixation durations are used most commonly as an early measure of lexical

access. However, these can consist of either the single fixation duration but also of the first fixation of multiple fixations on a word. This measure sometimes shows reversed word length effects because the first of a fixation on a longer word will be shorter because of the need to fixate a longer word multiple times (e.g. Rayner, Sereno & Raney, 1996). If there is only a single fixation on a target word, we assume that the target word is processed sufficiently with this one fixation because there is no refixation prior to moving to the next word or after doing so. Thus we prefer the measurement of single fixation duration because this would most accurately reflect lexical access time for the target word. The size of the corpus allows us to exclude words that are refixated whilst maintaining ample amount of statistical power. For the analyses of the other 3 dependent variables, we refer to Tables S1-S6 in the online supplementary materials.

Table 3

*Average Single Fixation Durations, Gaze durations and Skipping Rates for low [0.01-3.98] and high [3.99-5.90] frequent words and the L1 and L2 bilingual and monolingual frequency effects.*

| | Bilingual L1 | | | Bilingual L2 | | | Monolingual | | |
|---|---|---|---|---|---|---|---|---|---|
| | Low Frequency words | High Frequency words | FE | Low Frequency words | High Frequency words | FE | Low Frequency words | High Frequency words | FE |
| Single Fixation duration (ms) | 217.9 | 210.7 | 7.2 | 239.3 | 224.9 | 14.4 | 223.9 | 215.1 | 8.8 |
| Skipping Rate (%) | 27.6 | 48.9 | 21.3 | 23.8 | 44.0 | 20.2 | 29.9 | 51.0 | 21.1 |
| First Fixation Duration (ms) | 216.6 | 210.2 | 6.4 | 233.4 | 223.0 | 10.4 | 221.5 | 214.9 | 6.5 |
| Gaze duration (ms) | 241.8 | 223.9 | 17.9 | 277.9 | 244.6 | 33.3 | 245.3 | 227.4 | 17.9 |

**Bilingual L1 Reading vs. Monolingual Reading**

For the comparison between monolinguals and bilinguals reading in L1, all words that were either not fixated or were fixated more than once were excluded (46,63%). Single fixations that differed more than 2.5 standard deviations from the subject means were excluded from the dataset (2.23%). This left us with 265 756 data points. The dependent variable was log transformed to normalize the distribution as suggested by the Box-Cox method. This transformation did not change the functional relationship between the single fixation durations and the log-transformed word frequencies (see Figure A.1 in Appendix A). This data was fitted in a linear mixed model using the lme4 package (version 1.1-7) of R (version 3.0.2). The model contained the fixed factors of Bilingualism (L1 or mono), log 10 word frequency (continuous), L1 proficiency (continuous) and the control variable of word length (continuous). As proficiency variable we used the score on the L1 LexTALE (Lemhöfer & Broersma, 2011). For the word frequency, the subtitle word frequency measures (English: Brysbaert & New 2009; Dutch: Keuleers, Brysbaert & New, 2010) were log transformed with base 10 to normalize their distribution. All continuous predictors were centered. The maximum correlation between fixed effects in the final model was -0.063.

In the model we included a random intercept per subject. This ensured that differences between subjects concerning genetic, developmental or social factors were modeled. We also included a random intercept per word because our stimuli sample is not an exhaustive list of all words in a language. The model was fitted using restricted maximum likelihood estimation (REML). First a full model, including all of the interactions between the fixed effects and the two random clusters, was fitted. The optimal model was discovered by backward fitting of the fixed effects, then forward fitting of the random effects and finally

again backward fitting the fixed effects. We strived to include a maximal random structure (Barr, Levy, Scheepers & Tily, 2013). For the final model see Table 4.

Our two groups did not differ in single fixation durations: L1 reading was equally fast for mono- and bilinguals (β=-0.019, SE=0.015, t-value=-1.25). We did find an overall frequency effect (β=-0.0082, SE=0.00095, t-value=-8.59), which was not larger for bilinguals than for monolinguals (β=0.00051, SE=0.0013, t-value=0.39).

No main effect of L1 proficiency was found. Proficiency did however interact with word frequency (β=0.00017, SE=0.000077, t-value=2.19). The score on the L1 LexTALE has a larger impact on the single fixation durations on low frequency words than on high frequency words (See Figure 2). This results in a smaller FE for participants with higher L1 proficiency scores.

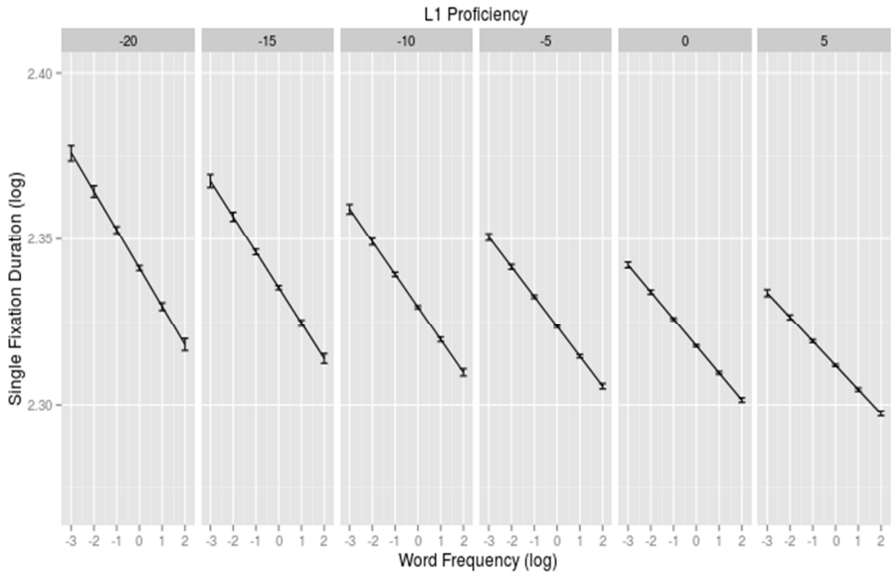


*Figure 2.* The effect of L1 Language Proficiency (centered on panels) and Word Frequency (centered and log-transformed on the x-axis) on Single Fixation Durations (log-transformed

on the y-axis) for monolinguals and bilinguals reading in their L1. This graph is plotted using the model estimates of the relevant effects of the final model for Single Fixation Durations.

What is striking is that the relationship between frequency and single fixation duration is the same for monolinguals and bilinguals reading in L1. Because word length is not matched across languages (0.48 letter difference), we added word length to this higher order interaction. The 3-way interaction was not significant and did not render the significant 2-way interaction between L1 proficiency and frequency insignificant.

Table 4

*Estimates, standard errors and t-values for the fixed and random effects of the final linear mixed effect model for Single Fixation Durations of the comparison between L1 bilingual and monolingual reading.*

| Bilingual L1 vs. Monolingual | | | |
|---|---|---|---|
| | Estimate | SE | t-value |
| Fixed Effects | | | |
| (Intercept) | 2.33 | 0.012 | 194.06 |
| Word Frequency | -0.0082 | 0.00095 | -8.59 |
| Bilingualism | -0.019 | 0.015 | -1.25 |
| L1 Proficiency | -0.0012 | 0.0012 | -0.99 |
| Word Frequency*L1 Proficiency | 0.00017 | 0.000077 | 2.19 |
| Word Frequency * Bilingualism | 0.00051 | 0.0013 | 0.39 |
| Control variables | | | |
| Word Length | 0.0020 | 0.00044 | 4.52 |
| Word Frequency * Word Length | -0.0013 | 0.00021 | -6.16 |
| L1 Proficiency * Word Length | -0.00013 | 0.000049 | -2.55 |
| | | | |
| | Variance | | SD |
| Random Effects | | | |
| Word | | | |
| (Intercept) | 0.00026 | | 0.016 |
| Subject | | | |
| (Intercept) | 0.0024 | | 0.048 |
| Word Frequency | 0.0000087 | | 0.0030 |
| Word Length | 0.0000045 | | 0.0021 |

| | | |
|---|---|---|
| Word Frequency * Word Length | 0.00000078 | 0.00088 |

**Bilingual L1 Reading vs. Bilingual L2 Reading**

Again, all words that were either not fixated or were fixated more than once were excluded from the dataset (50.8%). Single fixations that differed more than 2.5 standard deviations from the subject means were also excluded (2.27%). This left us with 221 953 data points. The dependent variable was log transformed with base 10 to normalize the distribution. As we have already demonstrated, this transformation did not change the functional relationship between the dependent variable and the log-transformed word frequencies (see Figure A.1 in Appendix A). This data was fitted in a linear mixed model using the lme4 package (version 1.1-7) of R (version 3.0.2). The model contained the fixed factors of language (L1 or L2), log 10 word frequency (continuous), L1 and L2 proficiency (continuous) and the control variables of word length (continuous) and age of L2 acquisition (continuous). As proficiency variables we used the score on the L1 and L2 LexTALE (Lemhöfer et al.). We computed the frequency variable the same way as in the previous comparison. Again, all continuous predictors were centered. The maximum correlation in the final model between fixed effects was -0. 643. Again, we included a random intercept per subject and a second random intercept per word. The model was fitted using restricted maximum likelihood estimation (REML). First a full model, including all of the interactions between the fixed effects, was fitted. The optimal model was discovered by backward fitting of the fixed effects, then forward fitting of the random effects and finally again backward fitting of the fixed effects. We strived to include a maximal random structure (Barr, Levy, Scheepers & Tily, 2013). For the final model see Table 5.

Our bilinguals fixated on average longer when reading in L2 than in L1 (β=-0.034, SE=0.0011, t-value=-11.37). We find an overall frequency effect (β=-0.011, SE=0.0011, t-value=-9.89) and a modulation of the FE by language (β=0.0031, SE=0.00099, t-value=3.10).

The FE is larger in L2 than in L1, which is caused by a larger disadvantage for low frequency L2 words (See Figure 3).

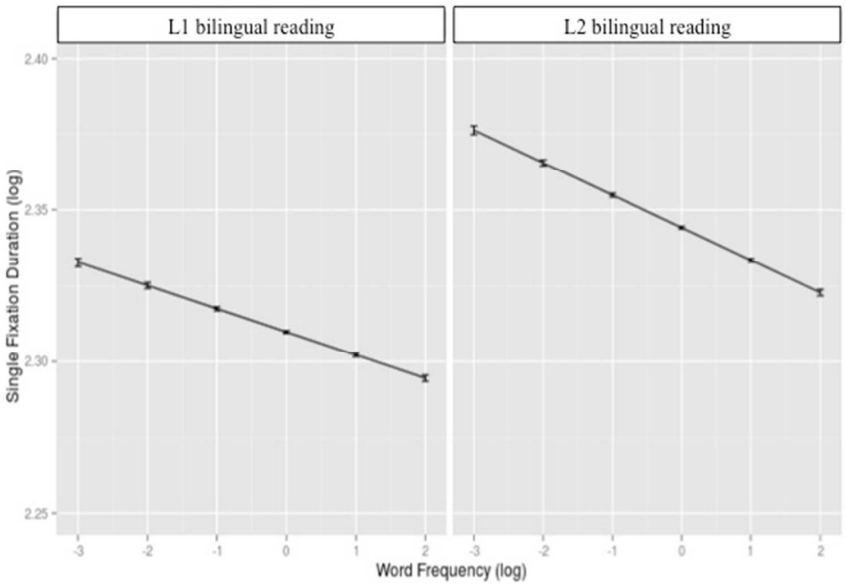


*Figure 3.* Single fixation durations (log-transformed) dependent on word frequency (log transformed and centered on the x-axis) and for bilinguals reading in L1 and L2 (panels). Standard Errors are indicated by whiskers. This graph is plotted using the model estimates of the relevant effects of the final model for Single Fixation Durations.

No main effects of L1 or L2 proficiency were found, but L1 proficiency modulates the frequency effect (β=0.00026, SE=0.00010, t-value=2.48). This modulation is the same when reading in L1 or L2. The FE is smaller when L1 proficiency is higher, both when the bilinguals read in L1 and in L2. We thus replicate the modulation by L1 proficiency of the FE. Figure 4 shows that the modulation of the FE by L1 proficiency is driven by speeded lexical access for low-frequent words both in L1 and L2 reading.

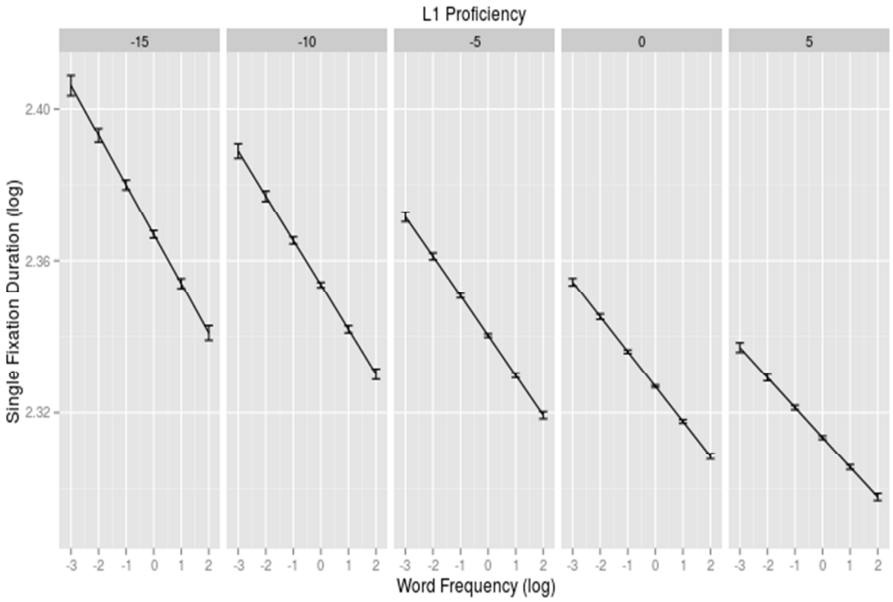


*Figure 4*. The effect of L1 Language Proficiency (centered on the panels) and Word Frequency (log-transformed and centered on the x-axis) on Single Fixation Durations (log-transformed on the y-axis) for bilinguals reading in L1 and L2. This graph is plotted using the model estimates of the relevant effects of the final model for Single Fixation Durations.

L2 proficiency interacted with language (β=0.00082, SE=0.00020, t-value=4.16). This means that when the bilinguals were reading in L2, there was an advantage for participants scoring high on L2 proficiency: they make shorter single fixations. For L1 reading an opposite effect was found: a higher score on L2 proficiency made the single fixation durations longer (See Figure 5).

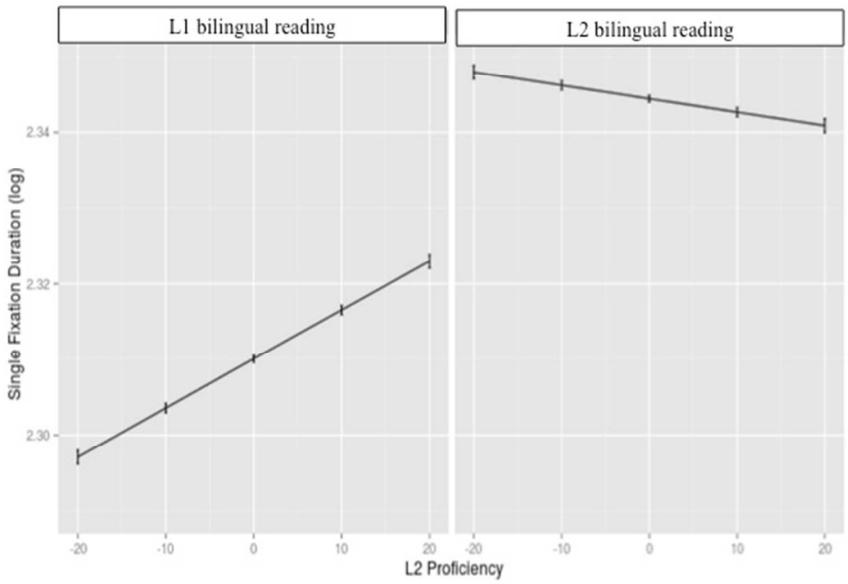


*Figure 5.* Single fixation durations (log-transformed) dependent on L2 Proficiency score (on the x-axis) for bilinguals reading in L1 and L2 (panels). Standard errors are indicated by whiskers. This graph is plotted using the model estimates of the relevant effects of the final model for Single Fixation Durations.

Because word length is not matched across languages, we again added word length to the higher order interactions. These 3-way interactions were not significant and did not render the other 2-way interactions insignificant. This means that the effects described generalize for both short and long words.

Table 5

*Estimates, standard errors and t-values for the fixed and random effects of the final linear mixed effect model for Single Fixation Durations of the comparison between bilingual L1 and L2 reading.*

| Bilingual L1 vs. Bilingual L2 | | | |
|---|---|---|---|
| | Estimate | SE | t-value |
| Fixed Effects | | | |
| (Intercept) | 2.34 | 0.011 | 213.81 |
| Word Frequency | -0.011 | 0.0011 | -9.89 |
| Language | -0.034 | 0.0030 | -11.37 |
| L1 Proficiency | -0.0027 | 0.0019 | -1.41 |
| L2 Proficiency | -0.00019 | 0.00096 | -0.19 |
| Word Frequency * Language | 0.0031 | 0.00099 | 3.10 |
| Word Frequency * L1 Proficiency | 0.00026 | 0.00010 | 2.48 |
| Language * L2 Proficiency | 0.00082 | 0.00020 | 4.16 |
| Control variables | | | |
| Word Length | 0.0046 | 0.00076 | 6.09 |
| Age of Acquisition L2 | -0.0020 | 0.0035 | -0.58 |
| Word Frequency*Word Length | -0.0012 | 0.00014 | -8.25 |
| L1 Proficiency* Word Length | -0.00025 | 0.00010 | -2.42 |
| L2 Proficiency * Word Length | 0.00012 | 0.000050 | 2.35 |
| Language * Word Length | -0.0024 | 0.00049 | -4.88 |

| | Variance | SD |
|---|---|---|
| Random Effects | | |
| Word | | |
| (Intercept) | 0.00025 | 0.016 |
| Subject | | |
| (Intercept) | 0.0023 | 0.048 |
| Language | 0.00015 | 0.012 |
| Word Frequency | 0.000015 | 0.0038 |
| Word Length | 0.0000086 | 0.0029 |
| Language * Word Frequency | 0.0000048 | 0.0022 |
| Language * Word Length | 0.0000015 | 0.0012 |

## General Discussion

This paper compared the monolingual and bilingual (L1 and L2) FE in text reading. Participants read an entire novel containing ± 29 000 content words, of which ± 8 000 were nouns. Bilinguals read the novel half in Dutch (L1), half in English (L2). In the analyses of single fixation durations on non-cognate content words, we found similarly sized FE's for bilinguals and monolinguals reading in their mother tongue. A rise in L1 proficiency reduced the slope of the L1 FE. The bilinguals showed a larger FE when reading in their L2 compared to reading in their L1. We also found a modulation of the bilingual L2 FE by L1 proficiency. A rise in L1 proficiency reduced the slope of the L1 and L2 FE. L2 proficiency did not modulate the FE, but it did have a differential effect across languages. In L2 reading, a rise in L2 proficiency speeds up single fixations, for L1 reading a rise in L2 proficiency does the opposite. This trade-off of reading speed is in line with the idea of 'weaker links'. To account for both these and previous results, we propose an account that fits within the framework of the weaker links hypothesis, suggesting not only a lower exposure to all lexical items but a disproportionate overestimation of corpus word frequencies for low frequency words for smaller vocabularies. Our proposal is consistent with a purely *exposure* based explanation of language processing speed.

### Bilingual vs. Monolingual L1 FE

We find a similarly sized FE for bilinguals reading in L1 and monolinguals reading in their mother tongue. Our findings seem at odds with the weaker links account, which predicts that due to a lower exposure to all items in the bilingual lexicon, bilinguals would show an overall larger FE in both their languages compared to a monolingual. Gollan and Acenas (2008), who mostly tested balanced Spanish-English populations, make the implicit assumption with their weaker links account that the total language exposure is equal for all people. While this maybe the case for their participants, it is definitely not true for all groups

of bilinguals. Our population of unbalanced bilinguals usually acquires a 2nd language in a classroom context, thus increasing their total language exposure, not per se substantially decreasing their L1 exposure. The acquisition of a second language for adults might be more defined by actively seeking more language exposure in a second language, resulting in indeed a larger lexicon, but also a higher total exposure. The hypothesis that bilingual exposure to L1 is not substantially lowered by bilingualism is supported by the fact that the L1 proficiency of our monolinguals was equal to the L1 proficiency of the bilinguals[1]. The similar proficiency scores indicate a similar sized vocabulary and thus a similar exposure to L1 for both groups. This contrast with most studies reporting differential FE's for bilinguals compared to monolinguals which use balanced bilingual populations and/or report lower target language proficiency for bilinguals than for monolinguals (Gollan et al., 2011; Lemhöfer et al., 2008; Lehtonen et al., 2012). To conclude, the weaker links account connects lower language exposure, leading to lower proficiency, to a larger FE. We nuance this rationale by pointing out that not all bilingual groups necessarily have lower L1 exposure than monolinguals do. This means that as long as there are no differences in language exposure as measured by language proficiency, we do not expect differently sized FE's. We would only predict a perceivable disadvantage for bilinguals in L1 compared to monolinguals when vocabulary size, and thus exposure, would be considerably smaller for the bilinguals.

The second important observation in our data is the reduction of the monolingual and bilingual L1 FE as L1 proficiency rises. This is consistent with multiple findings in the literature. For example Ashby, Rayner and Clifton 's (2005) eye tracking experiment found that underperforming adults show a larger frequency effect especially for low frequency words. Also, Kuperman and Van Dyke (2011) showed that individual language skill scores in

[1]All 4 methods measuring L1 proficiency (LexTALE, lexical decision task, spelling test and the proficiency questionnaire) do not yield different scores for the two groups (see Table B.1 in Appendix B for a summary of the objective measures). This makes it highly unlikely that we fail to pick up on existing language proficiency differences between our two groups.

rapid automatized naming and word identification modulated frequency effects for fixation times. Participants scoring high on language skill, showed a smaller frequency effect. Diependaele et al. (2013) showed that both for monolinguals and bilinguals, the rise of target language proficiency makes the size of the FE of word identification times smaller. Kuperman and Van Dyke (2013) observed that the relative amount of exposure to high corpus based frequency words will be virtually identical for individuals with different language experiences, whereas the low corpus frequency words will yield a larger difference in exposure, i.e. lexical entrenchment, for different groups.

In short, a higher L1 proficiency score reflects the size of the lexicon and the exposure to the items in that lexicon. Our results show, consistent with ideas formulated in Diependaele et al. (2013), that target language proficiency explains the size of the FE in both monolingual and bilingual groups and that the relationship between proficiency and FE is exactly the same for these two groups. This implies that we do not need qualitatively different lexical processing mechanisms to explain the size of L1 FE's for monolinguals and unbalanced bilinguals.

When we look at the mechanisms behind this modulation of the FE, we can draw conclusions about the location on the word frequency range this effect takes place. As we see a modulation of the FE by L1 proficiency even when word frequency is log transformed, this means that L1 proficiency does not measure absolute L1 exposure but is more sensitive to the L1 exposure for low frequency L1 items.

**Bilingual L1 vs. Bilingual L2 FE**

Bilinguals show a larger effect of frequency in the processing of L2 text than in the processing of L1 text. This finding is compatible with findings of Duyck et al. (2008) and Whitford and Titone (2011), who also found larger L2 FE's for unbalanced bilinguals, respectively for sentence reading and paragraph reading.

This finding is compatible with accounts of word recognition that implement implicit learning. In unbalanced bilingual populations, L2 words are learned later than L1 words and they have received on average less exposure than L1 words, thus making the threshold for activation for L2 items lower or the representations of these L2 words less accurate. Because we used corpus word frequencies in our analyses, the actual word exposure is overestimated for L2 reading compared to L1 reading. Kuperman and Van Dyke showed that this is especially the case for words with a low corpus frequency. This results in a larger FE in L2 mainly driven by a disproportional slower processing of low frequency words (See Figure 3).

Both in L1 and L2, a larger L1 proficiency reduces the slope of the FE. The effect of L1 proficiency on L1 reading is explained extensively in the above section: the processing time becomes disproportionally faster for low frequency than for high frequency words as exposure rises, causing a smaller FE.

The effect of L1 proficiency on L2 reading is much more surprising. Apparently, increased vocabulary size in the mother tongue facilitates access to low frequency words in a second language. To accommodate this finding, we have to assume that the L1 vocabulary size is measuring something more than exposure to the mother tongue. It is reasonable to assume that the amount of L1 exposure should be approximately the same for subjects with similar SES, education and age. Given that we do find different L1 proficiency scores, we are probably picking up on a more abstract reading skill or general language aptitude by measuring L1 vocabulary size. This assumption makes it more understandable that L1 proficiency modulates the FE in L2 reading in much the same way as it does in L1 reading. This line of reasoning is compatible with the idea proposed by Perfetti et al. (2005) that there is some individual variable that determines the speed of learning connections between word forms and meaning. We seem to capture this variable with our measure of L1 proficiency.

Diependaele et al. (2013) showed that proficiency explained the difference in FE across groups. In our data proficiency modulated the FE, but it did not eliminate the interaction between frequency and group. This means that the size of the FE was not totally explained by proficiency score. This is not that surprising, given that eye movement measures are more complex than identification times. Also, Whitford and Titone (2011) 's results are in line with ours, seeing that they still found differences across groups after proficiency was added to their model.

In our data L2 proficiency did not have an effect on the size of the FE, neither in L1 reading nor in L2 reading. Higher L2 proficiency scores did however reduce L2 reading speed, which validates the measure. For reading in L1, a rise in L2 proficiency made the single fixation durations longer. High L2 proficiency does seem to reduce reading speed in L1, congruent with the idea of weaker links. These are the only effects of L2 proficiency we find in our reading data. It seems that while L1 proficiency has a disproportioned impact on low frequency words in both languages, L2 proficiency has an equally large impact on low and high frequency words, but an opposite effect in both languages. Our results thus show that, despite the high correlation between the two, L1 and L2 proficiency are distinct concepts. L1 vocabulary size seems to be a measure for a general language aptitude, while L2 vocabulary size might be more linked to actual L2 exposure.

Although we tested similar populations (unbalanced bilinguals[2]) in a similar task (natural reading), Whitford and Titone (2011) found that more L2 exposure was linked to a larger L1 FE, but to a smaller L2 FE. So in their data L1 and L2 FE's are a function of L2 exposure, while our data shows that L1 and L2 FE's are a function of L1 proficiency. A large factor to take into account when trying to reconcile our data with those of Whitford and

[2] Note that the languages of the tested populations were different. In our study Dutch-English bilinguals were tested, in Whitford and Titone's (2011) English-French bilinguals were tested.

Titone is that their analysis did not actually include L1 proficiency of the bilinguals. Given that L1 and L2 proficiency are highly correlated, it is plausible that removing one of the factors from the analysis will have an impact on the significance of the other. Another factor is that they use a subjective estimate of L2 exposure in their analysis, while we use an objective vocabulary score to approximate language exposure. When we enter the subjective L2 exposure ratings in our analysis without L1 proficiency, we see that Subjective L2 exposure does have an effect on the slope of the L1 and L2 FE, just as in Whitford and Titone. A higher subjective exposure to L2, reduces the slope of the FE in L1 and L2. Again, a lower exposure, inflates the FE. So, the fact that L2 exposure influences the size of the FE is compatible with Whitford and Titone's results. What is not compatible is that we do not find a differential effect of this subjective L2 exposure on L1 and L2 reading. In our data, the effect of L2 exposure is the same in L1 and L2 reading, with smaller FEs for both languages.

Another possible reason for these different findings is that Whitford and Titone (2011) use gaze durations and total reading time as dependent variables. As already explained we prefer single fixation durations due to the complexity of eye movement variables. In their appendix they do report analyses of first fixation duration and skipping rates, but not single fixation durations. Their results for first fixation durations patterned with their results for gaze durations.

Our results are compatible with the assumption that the interaction between language proficiency and word frequency reported across a number of studies is caused by the use of corpus based word frequencies. Kuperman and Van Dyke (2013) show that in eye movement data the interaction between proficiency and frequency disappears when the objective corpus frequencies are replaced in the analysis by subjective frequencies, acquired by familiarity ratings. These subjective frequencies are supposed to be a closer approximation of the exact number of times a person has been exposed to a word form. For future studies, we

recommend the use of more accurate estimates of actual word frequencies of bilingual populations to study the bilingual and monolingual FE.

A possible criticism to our comparison of English and Dutch text is that the larger FE's for L2 compared to L1 reading could be explained by inherent language differences between English and Dutch, not controlled for in the experimental design. Given that the monolingual (English) - L1 Bilingual (Dutch) comparison did not yield any significant differences across groups, the differences we did find across languages in L1 (Dutch) and L2 (English) are very unlikely to be due to inherent language characteristics. Also the two most important lexical variables, word length and word frequency, were included in all of the higher order interactions in each model. This ensures that the reported effects are not due to any differences between the English and Dutch texts regarding word frequency or word length.

Even so, it could be pointed out that, although the Dutch language is very closely related to English, English has a deeper orthography than Dutch (Aro & Wimmer, 2003). This means that the mapping from orthography to phonology is less transparent for English than for Dutch. This deeper orthography could, according to the orthographic depth hypothesis (Katz & Feldman, 1983) lead to more reliance on the orthographic route of visual word recognition leading to more coarse-grained language processing. In this view, one could assume that this larger reliance on lexical representations for deep orthographies could cause larger word frequency effects on lexical access in those languages. This orthographic depth hypothesis is not without challenge (e.g. Besner & Hildebrandt, 1987; Lukatela & Turvey, 1999; Seidenberg, 1985, 1992;Tabossi & Laghi, 1992). For example Besner and Hildebrandt (1985) compared naming in two Japanese syllabic orthographies and show that Japanese readers always use the orthographic route, regardless of the orthographic depth of the script they are reading. Second, looking at data supporting the orthographic depth hypotheses, no

cross-lingual comparison has found a modulation of the size of the frequency-effect by the orthographic depth of a language (Frost, Katz & Benin, 1987; Seidenberg & Vidanovic, 1985) and, to our knowledge, no study finds effects of orthographic depth on eye movements. As far as we know, the only evidence for a modulation of the frequency effect by depth of orthography comes from a study by Frost (1994). He compared naming of words in two scripts of Hebrew; an unpointed (deep) and a pointed (shallow) variant. He found a frequency effect for unpointed Hebrew words and no frequency effect for pointed Hebrew words. The absence of any frequency effect in the pointed script is probably caused by a) the very transparent nature of the script and the task used, which makes it sufficient to use strict grapheme to phoneme conversion rules without activating the correct lexical representation and/or b) the low frequent use of this particular script. Both of these factors are not applicable to reading Dutch. According to the same orthographic depth hypothesis, language learners rely more on phonology than adult skilled readers, regardless of language (e.g. Katz & Feldman, 1983). This means that L2 reading of English should rely less on the orthographic route, than L1 reading. So this hypothesis would actually predict a smaller frequency effect for L2 readers of English compared to L1 readers of English or Dutch, the opposite of what we observed.



## Conclusion

A systematic exploration of the bilingual and monolingual FE in text reading showed that the FE is modulated by L1 proficiency, both for monolinguals and for bilinguals in L1 and L2.

The size of the FE was comparable for bilinguals and monolinguals when both groups read in their mother tongue. . Bilinguals displayed no disadvantages in any of the L1 proficiency (see Appendix B) or any of the L1 reading measures under investigation (see

results and supplementary materials) compared to monolinguals. A higher score on L1 proficiency reduced the size of the FE equally for both groups. The size of the FE was larger for bilinguals reading in L2 compared to bilinguals reading in L1. Bilinguals showed clear proficiency (see Appendix B) and reading disadvantages (see results and supplementary materials) in L2 compared to L1. The size of the FE was reduced for participants with higher scores on L1 proficiency, both for L1 and L2 reading. Whereas objective L2 proficiency had no effect on the slope of the FE, neither in L1 reading nor in L2 reading, a subjective rating of L2 exposure did modulate the size of the FE. A higher subjective exposure to L2 reduces the slope of the FE in L1 and L2. Because of the log transformation of the word frequency measure, we can attribute the modulation of the frequency effect to a disproportionate lower exposure to words with a low corpus frequency in L2 compared to L1.

These results are easily reconcilable with the weaker links account and a) provide evidence for the assumption that the same qualitative relationship between exposure frequency and word recognition exists for all language users and b) clarify that it is not a lowering of exposure to all items in the lexicon, but a disproportional lowering of the exposure to words with a low corpus word frequency that inflates the FE.

## References

Ashby,J., Rayner, K., & Clifton, C.J. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology, 58A* (6), 1065-1086.

Baayen, R.H. (2001). *Word frequency Distributions.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Balota, D., Pilotti, M., Cortese, M. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory and Cognition, 29*, 639-647

Besner, D., & Hilderbrandt, N. (1987). Orthographic and phonological codes in the oral reading of Japanese kana. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13*, 335-343.

Brysbaert, M. & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition, 13* (3), 359-371.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera 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, 41* (4), 977–990.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bülte, J., & Bühl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimate. *Experimental Psychology, 58* (5), 412-424.

Butler, B., & Hains, S. (1979). Individual differences in word recognition latency. *Memory & Cognition, 7,* 68–76.

Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition, 28*, 143-153.

Christie, A. *The Mysterious Affair at Styles*. N.p.: John Lane, 1920. *Project Gutenberg*. 1 Mar. 1997. Web. 07 Nov. 2012. <http://www.gutenberg.org/>.

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16(*6), 1084-1096.

Cop, U., Drieghe, D., & Duyck, W. (2014). Eye Movements in Natural Reading: a Comparison of Monolingual and Bilingual Novel Reading. Manuscript submitted for publication.

Diependaele, K., Lemhöfer, K. & Brysbaert, M. (2013). The Word Frequency Effect in First- and Second-language word recognition: A lexical entrenchment account. *The Quarterly journal of experimental Psychology, 66*(5), 843-863.

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

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Learning, Memory, and Cognition*, *33*(4), 663-679.

Duyck, W., Vanderelst, D., Desmet, T. & Hartsuiker, R.J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review, 15*(4), 850-855.

Faust, M., Balota, D., Spieler, D., & Ferraro, F. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125,* 777–799.

Forster, K. I. F. & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 12, 627-35

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for Visual Word Recognition and Orthographical Depth: A Multilingual Comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(1), 104-115

Frost, R. (1994). Prelexical and Postlexical Strategies in Reading: Evidence From a Deep and a Shallow Orthography. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(1), 116-129.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113*(2), 256-281.

Gollan, T. H., & Acenas, L. A. (2004). What is a TOT? Cognate and translation effects on tip-of-the-tongue states in Spanish–English and tagalog-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 246–269.

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition (Pre-2011), 33*(7), 1220-1234.

Gollan, T.H., Montaya, R.I., Sera, C., & Sandoval, T.C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism and the weaker links hypothesis. *Journal of Memory and Language, 58*(3), 787-814.

Gollan, T. H., & Silverberg, N. B. (2001). Tip-of-the-tongue states in Hebrew–English bilinguals. *Bilingualism: Language and Cognition, 4,* 63– 84.

Gollan, T.H., Slattery, T.J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental psychology, 140*(2), 186-209.

Howes, D.H., & Solomon, R.L. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, **41**(6), 401-410

Ivanova, I., & Costa, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica, 127,* 277-288.

Katz, L. & Feldman, L. (1983). Relation between Pronunciation and Recognition of Printed Words in Deep and Shallow Orthographies. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9* (1), 157-166

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

Kuperman, V. & Van Dyke, J.A. (2011). Effects of individual differences in verbal skills on eye-movement patterns during sentence reading. *Journal of Memory and Language, 65,* 42-73.

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, 39* (3), 802-823.

Lehtonen, M., Hulten, A., Rodriguez-Fornells, A., Cunillera, T., Tuomainen, J., & Laine, M. (2012). Differences in word recognition between early bilinguals and monolinguals: Behavioral and ERP evidence. *Neuropsychologia, 50,* 1362-1371.

Lemhöfer, K., & Broersma, M. (2011, advance online publication). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*. doi:10.3758/s13428-011-0146-0.

Luk, G. & Bialystok, E. , 2014. Bilingualism is not a categorical variable: Interaction between Language proficiency and usage. *Journal of Cognitive Psychology, 25:5,* 605-621.

Lukatela, G., & Turvey, M.T. (1999). Reading in two alphabets. *American Psychologist*, *53*, 1057–1072.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and

Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. Journal of Speech, Language, and Hearing Research, 50, 940–967.

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.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General, 118,* 43-71.

Monsell, S. (1991). The nature and locus of word frequency effects in reading. In D. Besner & G.W. Humphreys (Eds.), *Basic Process in reading: Visual word recognition* (pp. 148-197). Hillsdale, NJ: Erlbaum.

Morton, J. (1970). A functional model of human memory: In D. A. Norman (Ed.), *Models of human memory* (pp. 203-254). New York: Academic Press

Murray, W.S., & Forster, K.I. (2004). Serial mechanisms in lexical access: the rank hypothesis. *Psychological Review, 111,* 721-756

Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), Reading acquisition (pp. 145–174). Hillsdale, NJ: Lawrence Erlbaum Associates.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. Scientific Studies of Reading, 11, 357–383.

Perfetti, C., Wlotko, E., & Hart, L. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1281-1292.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes*, 4(special issue), 21–49.

Rayner, K., Sereno, S.C. & Raney, G.E. (1996). Eye movements Control in Reading: A Comparison of two types of models. *Journal of Experimental Psychology, 22* (5), pp. 1188-1200.

Rubenstein, H., Garfield, L. & Millikan, J.A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 9,* 487-492.

Scarborough, D.L., Cortese, C., & Scarborough, H.S. (1977). Frequency and repetition effects in lexical memory, *Human Perception and Performance*, *3,* 3-12

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: evidence from pupillometry. *Frontiers in Psychology,5, doi:10.3389/fpsyg.2014.00137*

Seidenberg, M.S. (1985). The time-course of phonological code activation in two writing systems. *Cognition, 19*, 1-30.

Seidenberg, M.S. (1992). Beyond orthographic depth in reading: Equitable division of labor. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 85–118). Amsterdam: Elsevier.

Seidenberg, M.S., & Vidanovic, S. (1985). Word Recognition in Serbo-Croatian and English: Do they differ? Paper presented at the XXVI Annual Meeting of the Psychonomic Society, Boston.

Solomon R. L., & Howes, D. H. (1951). Word frequency, word value, and visual duration thresholds. *Psychological Review*, 58, 256- 270.

Tabossi, P., & Laghi, L. (1992). Semantic Priming in the Pronunciation of Words in Two Writing Systems: Italian and English. *Memory and Cognition, 20,*315-328.

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

Van Assche, E., Duyck, W., Hartsuiker, R., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. *Psychological science*, *20*(8), 923-927.

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* (6), 1176-1190.

Whitford, V., & Titone, D. (2011). Second-language experience modulates first-and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review,* doi: 10.3758/s13423-011-0179-5.

Yap, M.J., Balota, D.A., Sibley, D.E. & Ratcliff, R. (2012) Individual differences in visual word recognition: Insights from the English lexicon project. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 53-79

APPENDIX A: Illustration of the functional relationship between word frequency and single fixation duration

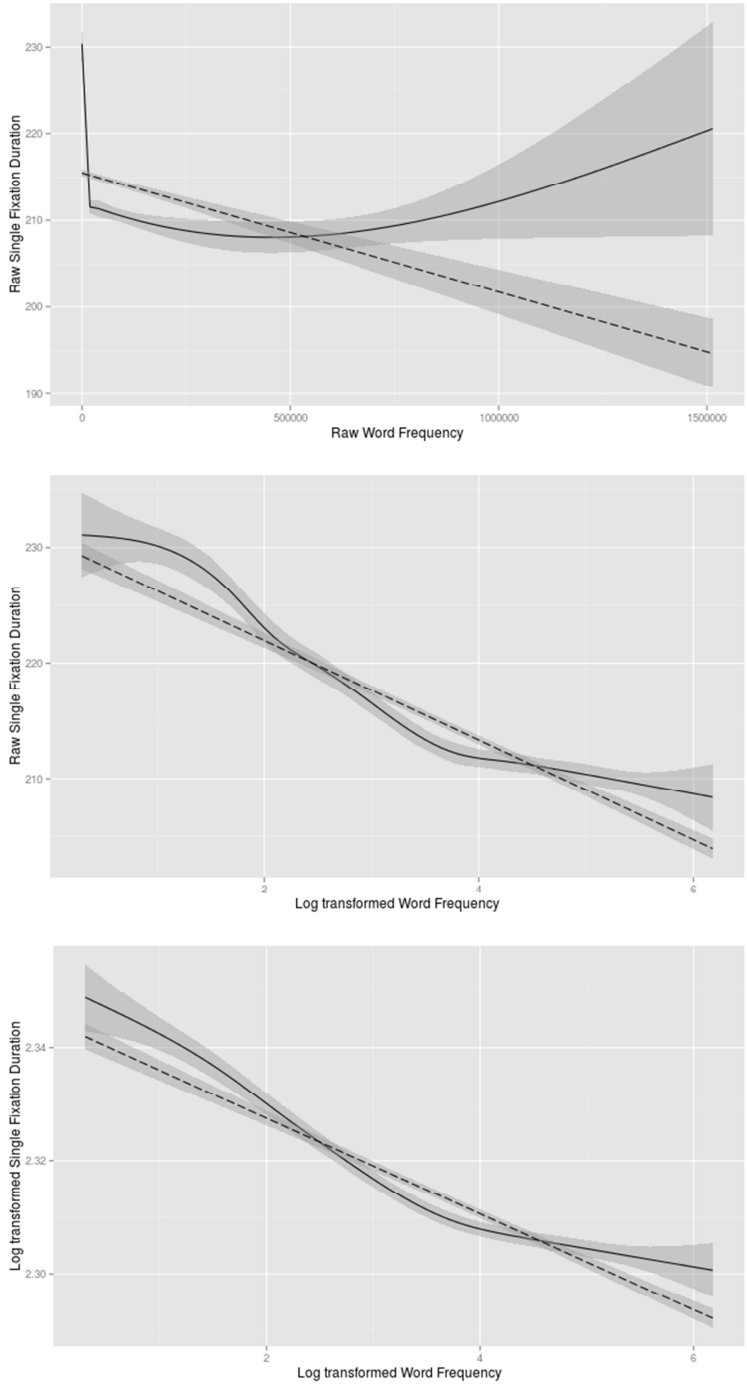

*Figure A.1.* The functional relationship between corpus word frequency and single fixation durations for non-cognate nouns for Dutch-English bilinguals reading the novel in Dutch. Dashed lines show the best linear fit, full lines show the best non-parametric additive fit. The first panel shows the relationship when both variables are untransformed. The second panel shows the relationship between untransformed fixation durations and log10 transformed word frequencies. The third panel shows the relationship when both variables are log-transformed with base 10. The second and third panels look similar, because the transformation of the dependent variable only caused a small change.

## APPENDIX B: Proficiency scores

Due to the lack of a standardized cross lingual spelling test, we tested the English spelling with the spelling list card of the WRAT 4 (Wilkinson & Robertson, 2006) and the Dutch spelling with the GLETSCHR (Depessemier & Andries, 2009). The LexTALE (Lexical Test for Advanced Learners of English) is an unspeeded lexical decision task, which is an indicator of language proficiency for intermediate to highly proficient language users, validated for English, Dutch and German (Lemhöfer & Broersma, 2011). A classical speeded lexical decision task was also administered in Dutch and English. The mean accuracy scores for the LexTALE and the percentage of correct word trials corrected for false alarms for the lexical decision task are reported in Table B.1.

Table B.1

*Average percentage scores (standard deviations between brackets and range between square brackets) on the LexTALE, Spelling test and Lexical Decision task for the bilingual and monolingual group.*

| | Monolinguals | Bilinguals L1 | Bilinguals L2 | t-value L1-L2 | t-value 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 *** | 0.49 |
| 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 *** | 0.99 |
| 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 *** | 0.67 |
| Composite Proficiency Score (%) | 83.25 (8.30) [67.35-94.40] | 85.54 (4.68) [77.87-95.25] | 67.81 (9.72) [52.49-86.76] | 11.78*** | 0.93 |

* p<0.05, **p<0.01, ***p<0.001

The Dutch (L1) proficiency of the bilinguals was matched with the English proficiency of the monolinguals (See Table B.1 in Appendix B), indicating that both groups were equally proficient in their first language. Neither the LexTALE (t=0.488, df=22.254, p=0.630), the spelling test (t=0.989, df=29.282, p=0.331), nor the lexical decision tasks (t=0.667, df=17.092, p=0.514) yielded significant differences for these two groups performing in L1. Consequently, there are no differences between the composite proficiency scores (t=-0.932, df=19.051, p=0.363). The bilingual L2 LexTALE scores were significantly lower than their L1 scores (t=7.587, df=18, p<0.001). The bilingual L2 Spelling scores were lower than the L1 scores (t=8.154, df=18, p<0.001). The performance of the bilinguals on the

classic lexical decision task was significantly better in L1 ($t=9.873$, $df=18$, $p<0.001$) than in L2. Bilinguals have lower composite proficiency scores in L2 than in L1 ($t=11.777$, $df=18$, $p<0.001$).

Supplementary materials - *Frequency Effects in Monolingual and Bilingual Natural Reading.*

**Results for skipping rates, first fixation durations and gaze durations.**

All models comparing monolingual and bilingual L1 reading contained the fixed factors of Bilingualism (bilingual or monolingual), log 10 word frequency (continuous), L1 proficiency (continuous) and the control variable of word length (continuous). All models comparing bilingual L1 and L2 reading contained the fixed factors of Language (L1 or L2), log 10 word frequency (continuous), L1 proficiency (continuous), L2 proficiency (continuous) and the control variable of word length (continuous) and L2 age of acquisition (continuous). As proficiency variables we used the score on the LexTALE (Lemhöfer & Broersma, 2011). For the word frequency, the subtitle word frequency measures (English: Brysbaert & New 2009; Dutch: Keuleers, Brysbaert & New, 2010) were log transformed to normalize their distribution. The procedure of the fitting of the models was exactly the same as for the single fixation durations.

**Skipping rates**

The data was fitted in a linear mixed model with a binomial distribution using the lme4 package (version 1.1-7) of R (version 3.0.2). To reduce collinearity, all continuous predictors were centered.

*Monolingual vs. Bilingual L1.* The maximum correlation between fixed effects in the final model was -.185. For the final model see Table S.1.

The main effect of bilingualism was not significant: skipping rates during L1 reading were equal for mono-and bilinguals. We did find an overall frequency-effect, which was not larger for bilinguals than for monolinguals. High frequency words were skipped more often than low frequency words.

No main effect of L1 proficiency was found. Contrary to the analysis of single fixation durations, the analysis of skipping rates did not show an interaction between proficiency and word frequency.

Table S.1

*Estimates, standard errors and z-values for the fixed and random effects of the final linear mixed effect model for Skipping Rates of the comparison between L1 bilingual and monolingual reading.*

| Bilingual L1 vs. Monolingual | | | |
|---|---|---|---|
| | Estimate | SE | z-value |
| Fixed Effects | | | |
| (Intercept) | -0.68 | 0.093 | -7.352 |
| Word Frequency | 0.13 | 0.010 | 11.98 |
| Bilingualism | 0.13 | 0.14 | 0.99 |
| L1 Proficiency | 0.01 | 0.0010 | 1.31 |
| Control variables | | | |
| Word Length | -0.28 | 0.0094 | -29.98 |
| Word Frequency * Word Length | -0.029 | 0.002 | -14.68 |
| | | | |
| | Variance | | SD |
| Random Effects | | | |
| Word | | | |
| (Intercept) | 0.046 | | 0.22 |
| Subject | | | |
| (Intercept) | 0.19 | | 0.43 |
| Word Frequency | 0.0024 | | 0.049 |
| Word Length | 0.0025 | | 0.050 |

*Bilingual L1 vs. Bilingual L2.* The maximum correlation between fixed effects in the final model was -0.75.

For the final model see Table S.2. The main effect of Language was not significant: skipping rates during L1 reading were equal to the skipping rates in L2 reading. We did not find an overall frequency-effect.

No main effect of L1 proficiency was found. Contrary to the analysis of single fixation durations, the analysis of skipping rates did not show an interaction between proficiency and word frequency.

Table S.2

*Estimates, standard errors and z-values for the fixed and random effects of the final linear mixed effect model for Skipping Rates of the comparison between bilingual L1 and L2 reading.*

| Bilingual L1 vs. Bilingual L2 | | | |
|---|---|---|---|
| | Estimate | SE | z-value |
| Fixed Effects | | | |
| (Intercept) | -1.95 | 0.19 | -10.44 |
| Word Frequency | 0.021 | 0.016 | 1.34 |
| Language | -0.10 | 0.091 | -1.12 |
| L1 Proficiency | 0.030 | 0.027 | 1.10 |
| L2 Proficiency | 0.006 | 0.012 | 0.48 |
| Word Frequency * L1 Proficiency | -0.0016 | 0.0022 | -0.75 |
| Control variables | | | |
| Word Length | -0.23 | 0.024 | -9.30 |
| Age of Acquisition L2 | 0.11 | 0.058 | 1.89 |
| L1 Proficiency*Word Length | 0.0040 | 0.0040 | 1.01 |
| Word Frequency * Word Length | -0.037 | 0.0060 | -6.07 |
| Word Frequency * L1 Proficiency * Word Length | 0.0031 | 0.00066 | 4.68 |
| | Variance | | SD |
| Random Effects | | | |
| Word | | | |
| (Intercept) | 0.12 | | 0.34 |
| Subject | | | |
| (Intercept) | 0.70 | | 0.84 |
| Language | 0.15 | | 0.39 |
| Word Frequency | 0.0027 | | 0.052 |
| Word Length | 0.010 | | 0.10 |
| Word Frequency* Word Length | 0.00042 | | 0.020 |

**First Fixation Durations**

The data was fitted in a linear mixed model using the lme4 package (version 1.1-7) of R (version 3.0.2). First fixation durations were log transformed with base 10 to approximate a normal distribution. To reduce collinearity, all continuous predictors were centered.

*Monolingual vs. Bilingual L1 reading.* The maximum correlation between fixed effects in the final model was .010. See Table S.3 for the final model.

The main effect of bilingualism was not significant: first fixation durations during L1 reading were equally long for mono-and bilinguals. We did find an overall frequency effect, which was not larger for bilinguals than for monolinguals.

No main effect of L1 proficiency was found. Conform the analysis of single fixation durations, the analysis of first fixation durations showed a marginal interaction between L1 proficiency and word frequency. More specifically, the FE slope decreased as L1 proficiency rose (see Figure S.1).

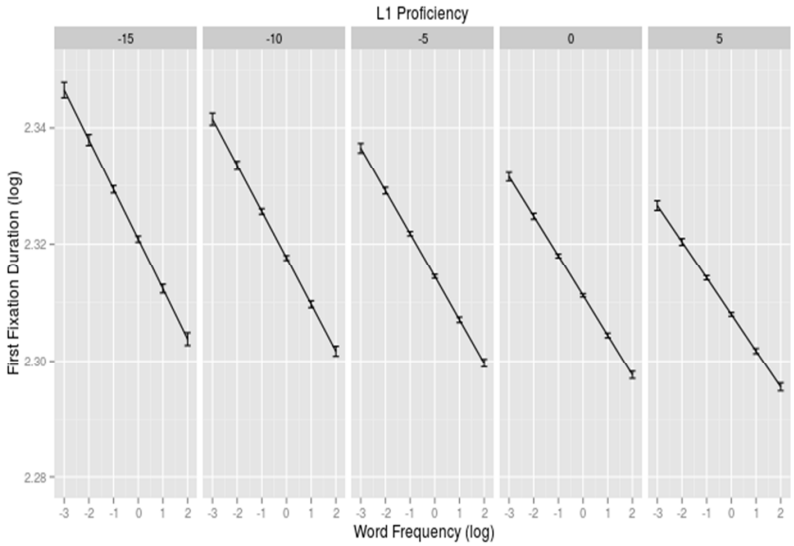


*Figure S.1.* Effect of Word Frequency (log-transformed and centered on the x-axis) on First Fixation Duration (log-transformed on the y-axis) dependent on L1 Proficiency (panels) for monolinguals and bilinguals reading in L1. Standard Errors are indicated by whiskers. This

graph is plotted using the model estimates of the relevant effects of the final model for First Fixation Durations.

Table S.3

*Estimates, standard errors and t-values for the fixed and random effects of the final linear mixed effect model for First Fixation Durations of the comparison between L1 bilingual and monolingual reading.*

| Bilingual L1 vs. Monolingual | | | |
|---|---|---|---|
| | Estimate | SE | t-value |
| Fixed Effects | | | |
| (Intercept) | 2.32 | 0.012 | 187.71 |
| Word Frequency | -0.0066 | 0.00053 | -12.39 |
| Bilingualism | -0.011 | 0.016 | -0.65 |
| L1 Proficiency | -0.00065 | 0.0011 | -0.58 |
| Word Frequency*L1 Proficiency | 0.00012 | 0.000061 | 1.92 |
| Control variables | | | |
| Word Length | 0.00051 | 0.00035 | 1.46 |
| Word Frequency * Word Length | -0.00051 | 0.00018 | -2.84 |
| | | | |
| | Variance | | SD |
| Random Effects | | | |
| Word | | | |
| (Intercept) | 0.00023 | | 0.015 |
| Subject | | | |
| (Intercept) | 0.0022 | | 0.047 |
| Word Frequency | 0.0000047 | | 0.0022 |
| Word Length | 0.0000027 | | 0.0016 |
| Word Frequency * Word Length | 0.00000065 | | 0.00081 |

*Bilingual L1 reading vs. Bilingual L2 reading.* The maximum correlation between fixed effects in the final model was -.67 for L1 and L2 proficiency. See Table S.4 for the final model.

The main effect of language was significant: first fixation durations during L1 reading were shorter than those made during L2 reading. We also found a main effect of frequency.

The FE for first fixations was marginally larger for L2 reading. See Figure S.2. This is compatible with the modulation of the FE found in the analysis of single fixation durations.

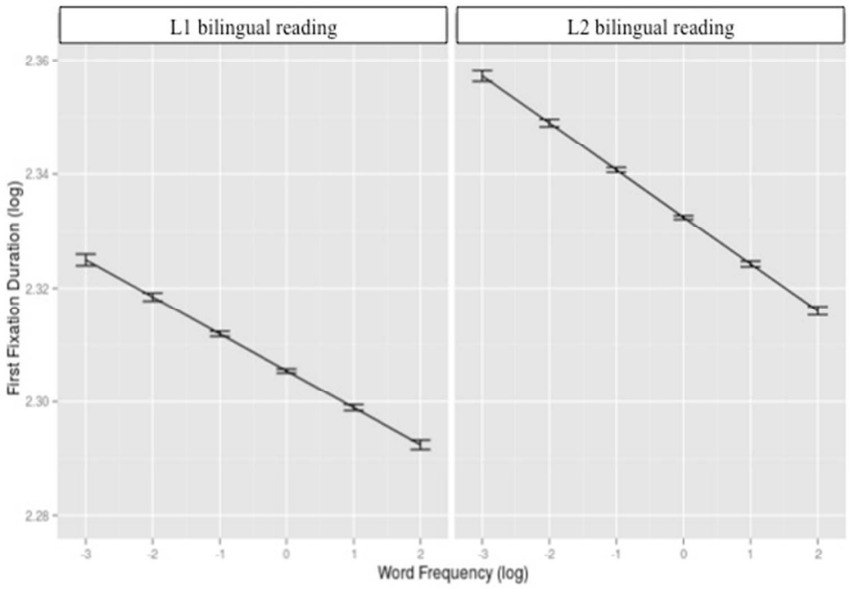


*Figure S.2.* Effect of Word Frequency (log-transformed and centered on the x-axis) on First Fixation Duration (log-transformed on the y-axis) for bilingual reading in L1 and L2 (panels). Standard Errors are indicated by whiskers. This graph is plotted using the model estimates of the relevant effects of the final model for First Fixation Durations.

No main effect of L1 proficiency was found. Contrary to the analysis of single fixation durations, the analysis of first fixation durations showed no interaction between proficiency and word frequency. We did find an interaction of Language and L1 proficiency. The effect of L1 proficiency was significant in both languages. When L1 proficiency score was higher, both the first fixations in L1 and in L2 reading are shorter. This effect was larger in L2. This is maybe due to a floor effect in L1 reading. See Figure S.3.

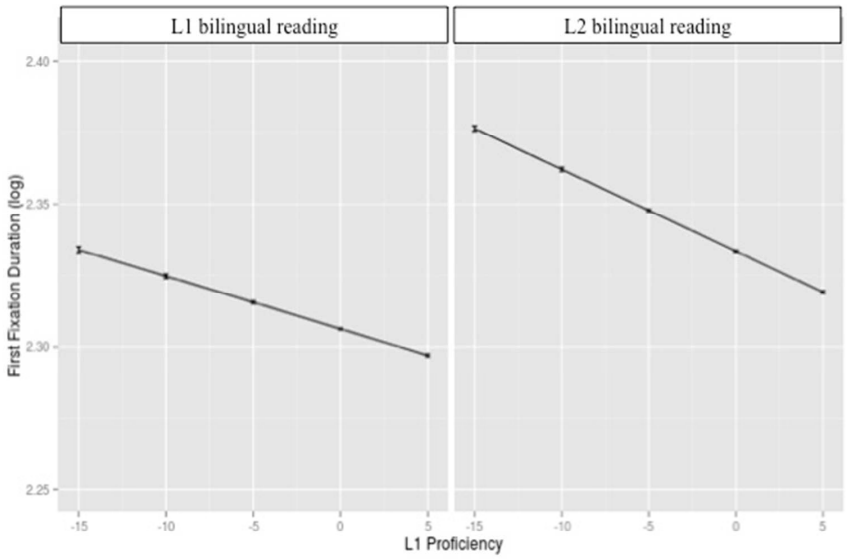


*Figure S.3.* Effect of L1 proficiency (centered) on First Fixation Duration (log-transformed) for bilinguals reading in L1 and L2 (panels). Standard Errors are indicated by whiskers. This graph is plotted using the model estimates of the relevant effects of the final model for First Fixation Durations.

Table S.4

*Estimates, standard errors and t-values for the fixed and random effects of the final linear mixed effect model for First Fixation Durations of the comparison between bilingual L1 and L2 reading.*

| Bilingual L1 vs. Bilingual L2 | | | |
|---|---|---|---|
| | Estimate | SE | t-value |
| Fixed Effects | | | |
| (Intercept) | 2.33 | 0.0095 | 246.35 |
| Word Frequency | -0.0082 | 0.0069 | -11.99 |
| Language | -0.027 | 0.0031 | -8.64 |
| L1 Proficiency | -0.0029 | 0.0020 | -1.43 |
| L2 Proficiency | 0.0011 | 0.0011 | 1.00 |
| Word Frequency * Language | 0.0017 | 0.00094 | 1.83 |
| Language * L1 Proficiency | 0.0010 | 0.00037 | 2.80 |
| Control variables | | | |

| | | | |
|---|---|---|---|
| Word Length | 0.0013 | 0.00041 | 3.24 |
| Age of Acquisition L2 | 0.00037 | 0.0041 | 0.09 |

| | Variance | SD |
|---|---|---|
| Random Effects | | |
| Word | | |
| (Intercept) | 0.00028 | 0.017 |
| Subject | | |
| (Intercept) | 0.0017 | 0.041 |
| Language | 0.00016 | 0.013 |
| Word Frequency | 0.0000046 | 0.0021 |
| Word Length | 0.0000026 | 0.0016 |
| Language * Word Frequency | 0.0000093 | 0.0031 |

**Gaze durations**

The data was fitted in a linear mixed model using the lme4 package (version 1.1-7) of R (version 3.0.2). Gaze Durations were log transformed with base 10 to approximate a normal distribution. All continuous predictors were centered.

*Monolingual vs. Bilingual L1.* The maximum correlation between fixed effects in the final model was -.062. For the final model see Table S.5.

We found a FE that was not modulated by bilingualism. This means that the FE is equally large for monolinguals and bilinguals when they read in their first language. We did not find a main effect L1 proficiency or Bilingualism. Word frequency did interact with L1 proficiency. Participants with a higher score for L1 proficiency showed a smaller frequency effect (see Figure S.4). These findings are consistent with the findings for single fixation durations and first fixation durations.

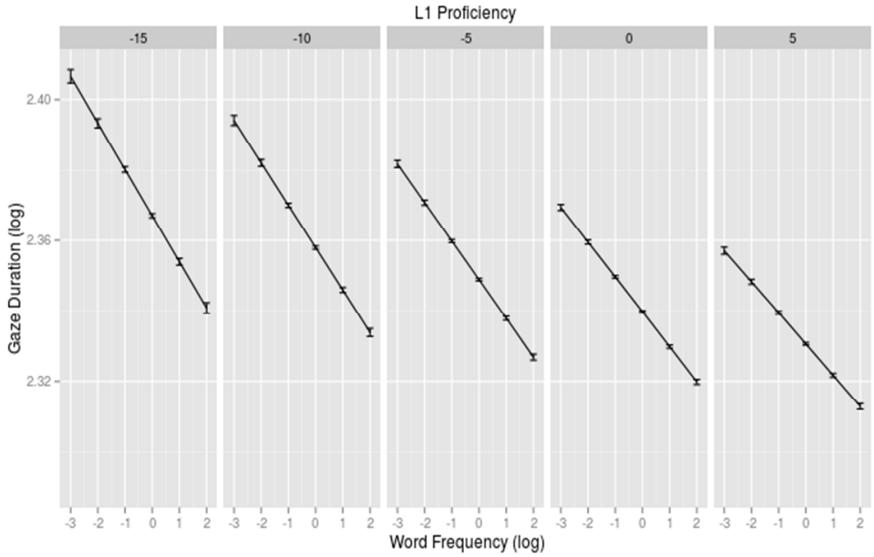


*Figure S.4*. The effect of Word Frequency (log-transformed and centered on the x-axis) on Gaze Duration (log-transformed) dependent on L1 Proficiency (centered on panels) for monolinguals and bilinguals reading in L1. Standard Errors are indicated by whiskers. This graph is plotted using the model estimates of the relevant effects of the final model for Gaze Durations.

Table S.5

*Estimates, standard errors and t-values for the fixed and random effects of the final linear mixed effect model for Gaze Durations of the comparison between L1 bilingual and monolingual reading.*

| Bilingual L1 vs. Monolingual | | | |
|---|---|---|---|
| | Estimate | SE | t-value |
| Fixed Effects | | | |
| (Intercept) | 2.34 | 0.013 | 183.57 |
| Word Frequency | -0.0095 | 0.0010 | -9.31 |
| Bilingualism | -0.014 | 0.015 | -0.96 |
| L1 Proficiency | -0.0017 | 0.0013 | -1.35 |
| Word Frequency * Bilingualism | 0.00025 | 0.0014 | 0.18 |
| Word Frequency * L1 Proficiency | 0.00022 | 0.000085 | 2.56 |
| Control variables | | | |

| | | | |
|---|---|---|---|
| Word Length | 0.0053 | 0.00076 | 7.05 |
| Word Frequency * Word Length | -0.0013 | 0.00023 | -5.69 |
| L1 Proficiency * Word Length | -0.00020 | 0.000098 | -1.99 |

| | Variance | SD |
|---|---|---|
| Random Effects | | |
| Word | | |
| (Intercept) | 0.00054 | 0.023 |
| Subject | | |
| (Intercept) | 0.0029 | 0.054 |
| Word Frequency | 0.000010 | 0.0032 |
| Word Length | 0.000016 | 0.0040 |
| Word Frequency * Word Length | 0.00000096 | 0.00098 |

*Bilingual L1 vs. L2 reading.* The maximum correlation between fixed effects in the final model was -0.60. See Table S.6 for the final model.

We found a main effect of language. Gaze durations were longer in L2 than in L1. Also, a main effect of word frequency was found. Gaze durations for high frequency words were shorter than those for low frequency words. This FE was larger in L2 than in L1 (see Figure S.5). The 3-way interaction between language, word frequency and word length was also significant. A significant interaction between Language and L2 proficiency was found (see Figure S.6). For L2 reading a higher L2 proficiency score facilitated gaze durations, parallel to the effect found for single fixation durations.

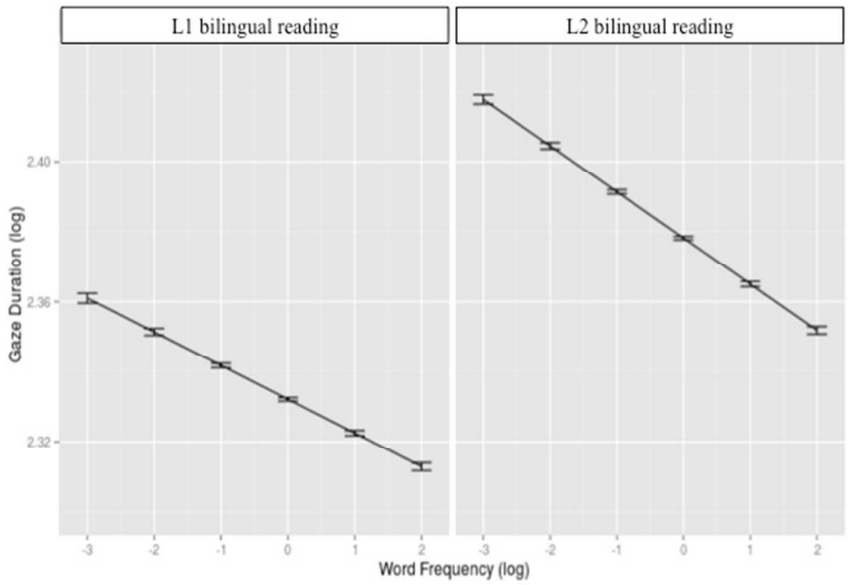


*Figure S.5.* The effect of Word Frequency (log-transformed and centered on the x-axis) on Gaze Duration (log-transformed) for bilinguals reading in L1 and L2 (panels). Standard Errors are indicated by whiskers. This graph is plotted using the model estimates of the relevant effects of the final model for Gaze Durations.

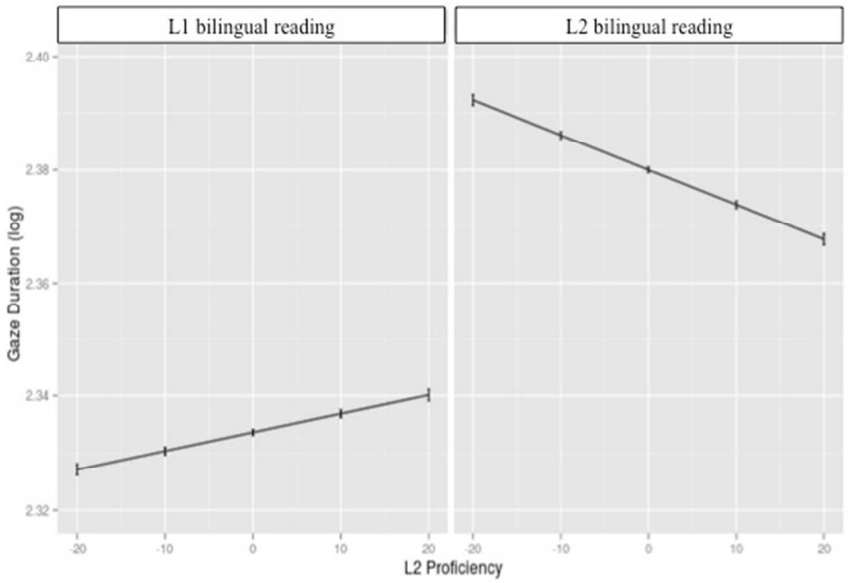


*Figure S.6.* The effect of L2 Proficiency (centered on the x-axis) on Gaze Durations (log-transformed) for bilinguals reading in L1 and L2 (panels). Standard Errors are indicated by whiskers. This graph is plotted using the model estimates of the relevant effects of the final model for Gaze Durations.

Table S.6

*Estimates, standard errors and t-values for the fixed and random effects of the final linear mixed effect model for Gaze Durations of the comparison between bilingual L1 and L2 reading.*

| Bilingual L1 vs. Bilingual L2 | | | |
|---|---|---|---|
| | Estimate | SE | t-value |
| Fixed Effects | | | |
| (Intercept) | 2.38 | 0.014 | 175.21 |
| Word Frequency | -0.013 | 0.0014 | -9.19 |
| Language | -0.044 | 0.0040 | -10.92 |
| L1 Proficiency | 0.00016 | 0.0016 | 0.10 |
| L2 Proficiency | -0.00066 | 0.00094 | -0.71 |
| Word Frequency * Language | 0.0039 | 0.0012 | 3.37 |
| Language * L2 Proficiency | 0.00094 | 0.00022 | 4.31 |

| | | | |
|---|---|---|---|
| Control variables | | | |
| Word Length | 0.010 | 0.0015 | 6.85 |
| Age of Acquisition L2 | -0.0040 | 0.0032 | -1.24 |
| Word Frequency * Word Length | -0.00041 | 0.00025 | -1.66 |
| Language * Word Length | -0.00048 | 0.00082 | -5.78 |
| L2 Proficiency * Word Length | 0.00011 | 0.00005 | 2.23 |
| Word Frequency * Language * Word Length | -0.0011 | 0.00032 | -3.30 |

| | Variance | SD |
|---|---|---|
| Random Effects | | |
| Word | | |
| (Intercept) | 0.00069 | 0.026 |
| Subject | | |
| (Intercept) | 0.0035 | 0.059 |
| Language | 0.00027 | 0.017 |
| Word Frequency | 0.000028 | 0.0053 |
| Word Length | 0.000037 | 0.0061 |
| Language * Word Frequency | 0.0000054 | 0.0023 |
| Language * Word Length | 0.0000062 | 0.0025 |

## References

Brysbaert, M., & New, B. (2009). Moving beyond Kucera 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, 41* (4), 977–990.

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

Lemhöfer, K., & Broersma, M. (2011, advance online publication). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*. doi:10.3758/s13428-011-0146-0.