The Word Frequency Effect in First and Second Language Word Recognition: A Lexical Entrenchment Account

Kevin Diependaele (1)

Kristin Lemhöfer (2)

Marc Brysbaert (1)

1. Ghent University, Ghent, Belgium
2. Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, The Netherlands

Kevin Diependaele

Department of Experimental Psychology

Henri Dunantlaan 2

9000 Ghent, Belgium

kevin.diependaele@gmail.com

Abstract

We investigate the origin of differences in the word frequency effect between native speakers and second language speakers. In a large-scale analysis of English word identification times we find that group-level differences are fully accounted for by the individual language proficiency scores. Furthermore, exactly the same quantitative relation between word frequency and proficiency is found for monolinguals and three different bilingual populations (Dutch-English, French-English and German-English). We conclude that the larger frequency effects for second language processing compared to native language processing can be explained by within-language characteristics and thus need not be the consequence of "being bilingual" (i.e., a qualitative difference). More specifically, we argue that language proficiency increases lexical entrenchment, which leads to a reduced frequency effect, irrespective of bilingualism, language dominance, and language similarity.

Introduction

The frequency with which words occur in the language is arguably the best documented and most robust predictor of word recognition performance. High-frequency words are processed faster and more accurately than low-frequency words, and this typically accounts for a great part of the variance. For instance, when Brysbaert, Buchmeier, Conrad, Jacobs, Bölte, and Böhl (2011) ran a stepwise multiple regression analysis on the lexical decision times of the 40 thousand words from the English Lexicon Project (Balota, et al., 2007), logarithmic word frequency came out as the most important variable accounting for almost 41% of the variance in the latencies. Similarly, Murray and Forster (2004, p. 721) concluded that: "Of all the possible stimulus variables that might control the time required to recognize a word pattern, it appears that by far the most potent is the frequency of occurrence of the pattern ... Most of the other factors that influence performance in visual word processing tasks, such as concreteness, length, regularity and consistency, homophony, number of meanings, neighborhood density, and so on, appear to do so only for a restricted range of frequencies or for some tasks and not others".

In the present study, we address the question why the word frequency effect is stronger in the second language (L2) than in the first language (L1). There have been several attestations of larger frequency effects as a function of multilingualism and language dominance. As far as we have been able to ascertain, the first such effect was reported by Van Wijnendaele and Brysbaert (2002: Figure 1). They asked Dutch-French and French-Dutch bilinguals to name words in L1 and L2. For each group, they observed a steeper word frequency curve in L2 than in L1 in addition to an increase in the intercept (i.e., generally slower naming times in L2 compared to L1).

In the domain of word production (picture-naming), Gollan, Montoya, Cera, and Sandoval (2008) also found a larger frequency effect for English-dominant bilinguals than for monolingual English participants. The same type of bilinguals further showed an even larger frequency effect in their non-dominant language (Spanish). Similar results have been reported in lexical decision (Duyck, Vanderelst, Desmet & Hartsuiker, 2008; Gollan et al., 2011), eye movement recording (Gollan et al., 2011; Whitford & Titone, 2012), and word identification (Lemhöfer et al., 2008).

Two types of explanation for the phenomenon can be put forward on the basis of the existing literature. According to the first, based on traditional interactive-activation type models of visual word recognition, the stronger frequency effect in L2 is caused by language competition in bilinguals. The second explanation attributes the difference in frequency effects to differences in language-specific skill. We first outline the two accounts and then introduce the present study and its role in evaluating the two accounts.

Differences in the Frequency Effect are Caused by Language Competition

There is ample evidence that the two lexicons of a bilingual are not functionally independent. For example, word recognition in a given target language has been shown to be influenced by semantic and/or form overlap with words of the other, non-target language (e.g., Christoffanini, Kirsner, & Milech, 1986; de Groot, Borgwaldt, Bos, & van den Eijnden, 2002; Dijkstra, Miwa, Brummelhuis, Sappelli, and Baayen, 2010; Haigh & Jared, 2007; Lemhöfer & Dijkstra, 2004), and it can be primed by, for example, form-overlapping words from the other language (Brysbaert, Van Dyck & Van de Poel, 1999; Dijkstra, Hilberink-Schulpen, & van Heuven, 2010; Kim & Davis, 2003). It is thus likely that lexical activation usually spreads across words from both languages.

Due to this language-independent lexical activation, it could be argued that bilinguals need to cope with more competition between similar word form representations than monolinguals. Competition between resembling word form representations (orthographic/phonological neighbors) is a central component of computational models based on interactive activation (e.g., Coltheart et al., 2001; Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981; Perry, Ziegler & Zorzi, 2007). When participants are processing the word 'bale', the correct representation must be discriminated from neighbors that also become activated, such as 'sale' and 'bake'; furthermore, L2 speakers need to additionally discriminate it from possible neighbors from their first language, like 'balk' in Dutch. The competition is thought to be particularly time-consuming for low-frequency words with high-frequency neighbors (Segui & Grainger, 1990). Given that L1 words can be regarded as, on average, subjectively high-frequent, it might be argued that the larger frequency effect in L2 is the outcome of increased competition from resembling L1 word form representations.

As a general proof of concept, consider Figure 1. It shows the behavior of a simple interactive activation network as a function of word frequency (logarithmic scale). The effect of a bilingual lexicon is simulated by comparing a full model of 7439 words with one where half of the lexicon is randomly removed. The left and right panels of Figure 1 thus provide a rough approximation of the situation for monolinguals (see 'Small Lexical Space') and bilinguals (see 'Large Lexical Space'). As expected on the basis of increased competition, Figure 1 shows that differences in "lexical search space" can indeed result in a larger frequency effect (see the steeper regression line in the right panel than in the left).

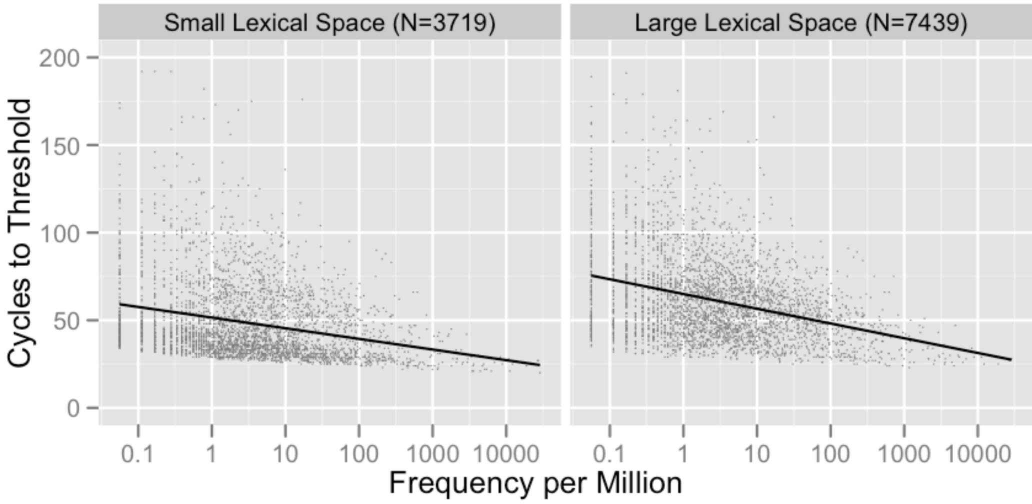

Figure 1. Cycles needed for a word identification response in the orthographic route of the bimodal IAM (Diependaele, Ziegler & Grainger, 2010). The effect of language competition is simulated by comparing target recognition times (i.e., cycles needed for targets to reach a preset decision threshold) across two different lexicon sizes: 3719 and 7439 words (the smaller lexicon was randomly sampled from the bigger one). In both panels, dots represent the individual decision times for the 3719 words contained in the smaller lexicon. Lines show the linear regression onto the corresponding log frequency scores (see Diependaele et al., 2010 for further details; all plots designed with ggplot2, Wickham, 2009).

The language-competition hypothesis makes two interesting predictions. First, the exact degree to which the frequency effect increases should be a function of how many languages an individual knows (i.e., how many cross-language competitors are activated) and how well they know each of these languages (i.e., language proficiency). This aligns with the findings of Gollan et al. (2008), who reported a large frequency effect in L2, followed by a smaller effect in the L1 of bilinguals, and the smallest effect in monolinguals. However, these results were obtained for word production, which might differ from word recognition. With respect to proficiency, the more skilled a bilingual is in one language relative to the other(s), the higher words from this language should be in subjective frequency, and therefore the less susceptible the processing of this language should be to interference from competitors in a non-target language (the greater the activity difference with cross-language competitors in the IA framework, for instance). The fact that language competition effects (e.g., cognate effects, homograph effects, etc.) are usually larger and more reliable in the processing of a non-dominant (i.e., more susceptible) language is also consistent with this idea (e.g., Caramazza & Brones, 1979; de Groot, et al., 2002).

The second critical prediction is that frequency effects will increase as a function of how strongly the known languages resemble each other (i.e., how many cross-language word neighbors there are). The increase in the English frequency effect is thus expected to be stronger for Dutch-English bilinguals than, for instance, for Finnish-English bilinguals. With respect to the language combinations that are present in the dataset analyzed here (L1: Dutch, German, or French; L2: English), French is more orthographically similar to English than are Dutch and German (Schepens, Dijkstra & Grootjen, 2012). So, for comparable proficiency levels, the competition account predicts a larger frequency effect for English L2 in French-English bilinguals than in the other two groups of bilinguals.

At the same time, there are several indications against the viability of the language-competition hypothesis. For a start, simulations with computational models suggest that neighborhood interference effects are only observed when one compares words without neighbors with words that have one or two neighbors (Bowers, Davis & Hanley, 2005; Davis, 2003, 2010, p. 732). Words without neighbors are recognized faster than those with a few neighbors. However, for words with many neighbors, the inhibitory relationship between the number of neighbors and word identification times disappears. In the present context, this means that increased interference due to L1 neighbors will only affect the processing of L2 words that have no neighbors in their own language, but that do have neighbors in L1. For Dutch-English bilinguals, for instance, less than 2% of the L2 words meet this restriction (according to the Celex lexical database; Baayen, Piepenbrock, & Gulikers, 1995).

The above reservations are in line with the elusiveness of the cross-language neighborhood effect. Lemhöfer et al. (2008), for instance, asked native English speakers and L2 speakers with various first languages to respond to a large number of words in the progressive demasking paradigm. Whereas the authors observed strong frequency effects in all groups (accounting for 20-40% of the observed variance), direct measures of cross-language competition yielded only very small effects. In particular, for the bilinguals there was no significant effect of L1 orthographic neighborhood size on L2 performance (see de Groot et al., 2002, for a similar finding in lexical decision).

Finally, the fact that word frequency effects are not stronger in (monolingual) individuals with a large vocabulary than in individuals with a small vocabulary, also argues against the idea that a larger lexical search space automatically leads to more competition. Quite on the contrary, many monolingual studies have reported that the word frequency effect is smaller in individuals who know many words than in individuals with a limited vocabulary (Ashby, Rayner & Clifton, 2005; Chateau & Jared, 2000; Sears, Siakaluk, Chow & Buchannan, 2008; Spielberger & Denny, 1963).

Differences in Frequency Effects are Caused by Lexical Entrenchment

An alternative general account of the stronger frequency effect in L2 is that it is not due to between-language competition, but to the usage-based characteristics of the lexical representations themselves (see e.g., Ellis, 2002). Due to a generally lower proficiency in L2 compared to L1 in unbalanced bilinguals, lexical memory representations in L2 will be 'weaker' than those in L1, in the sense that processing them will require more energy. There are several proposals as to how this reduced lexical entrenchment can be conceptualized. First, probably the most straightforward way to model the L2 lexical disadvantage in the context of an IA-type model is through the resting levels of the word nodes. These will reflect subjective rather than objective frequencies. L2 words are encountered less often than L1 words and this difference will be especially pronounced in the lower frequency range. Thus, subjective frequencies will be lower than objective ones in L2 and disproportionally more so in the lower ranges (see Kuperman & Van Dyke, 2012). The result is that frequency curves are shifted upwards, with a larger shift in the lower range. A lower proficiency or exposure rate will thus result in steeper frequency curves, i.e., larger frequency effects. In an IA model, we can implement this by reducing the L2 resting levels by a constant factor; the lower the proficiency in L2, the higher this multiplicative reduction will be[1]. Figure 2 illustrates the effect of this change: For reduced resting levels (i.e., L2 speakers of lower proficiency), the frequency effect will be larger.

---

[1] It is customary in IA models to scale resting levels between -.92 (minimum ~ lowest word frequency) and 0 (maximum ~ highest word frequency; see McClelland & Rumelhardt, 1981). In the present simulation we lowered the scale by multiplication with 1.5. The range thus became -1.38 – 0.

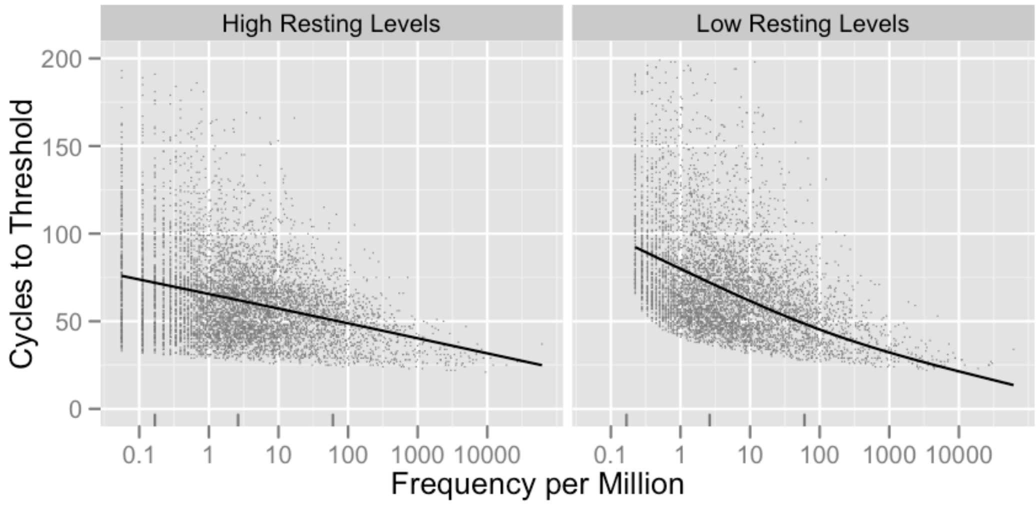

Figure 2. Cycles needed for a word identification response in the orthographic route of the bimodal IAM (Diependaele et al., 2010). Lexical selection as a function of lexical entrenchment is simulated via multiplicative scaling of the resting levels. More difficult selection (i.e., low entrenchment) corresponds to a situation where resting levels (i.e., subjective frequencies) are scaled down multiplicatively (by a factor 1.5 in the present simulation; see footnote 1). Dots represent individual decision times for all known words (N=7439). Lines show the regression with a 3-knot natural spline expansion of Log Frequency (Harrell, 2001, 2011; see Method section for further details). Knot locations are shown by the upward ticks on the x-axis (see Diependaele et al., 2010 for further details).

A second way to think of reduced lexical entrenchment in L2 is that lexical representations can differ in terms of how 'precise' they are, i.e., how well the orthographic, phonological and semantic information is defined and integrated in memory. This approach has been put forward in the form of the 'lexical quality' hypothesis (e.g., Perfetti, 1992, 2007). The idea is that increased word knowledge results in better precision of the corresponding lexical representations and, by consequence, these representations experience less interference from representations of similar words during their activation. L2 lexical representations will, on average, be of a lower precision than those in L1. For example, to an L2 speaker, the English words ‘squirrel’ and ‘quarrel’ may be more similar and thus more confusable than to a native speaker, who can quickly decide whether he is presented with the one or the other. The so-called ‘weaker links hypothesis’ of Gollan et al. (2008, 2011) is a specific example of this approach to lexical entrenchment in the context of bilingual language production. It states that non-native speakers show larger frequency effects in their language production because the limited experience with L2 leads to a reduced level of integration of semantic and phonological codes.

In the IA framework, one way to model lower entrenchment in terms of lower lexical precision is by decreasing the level of word-word lexical (form) inhibition. Less precise lexical representations will have lower ability to inhibit their competitors. Because of the lower level of inhibition, more competitors will reach the activation threshold and thus negatively influence target recognition times (i.e., make target selection more difficult). Figure 3 illustrates the effect of a low and high degree of proficiency on the frequency effect, simulated by, respectively, a low and high value for lateral word-word inhibition. In essence, the behavior is the same as in the competition account simulation above: the more candidates can come into play, the larger (i.e., steeper) the frequency effect. The critical difference is that in the lexical entrenchment approach, there is no need to assume *cross-language* (neighbor) competition as the origin of the effect.

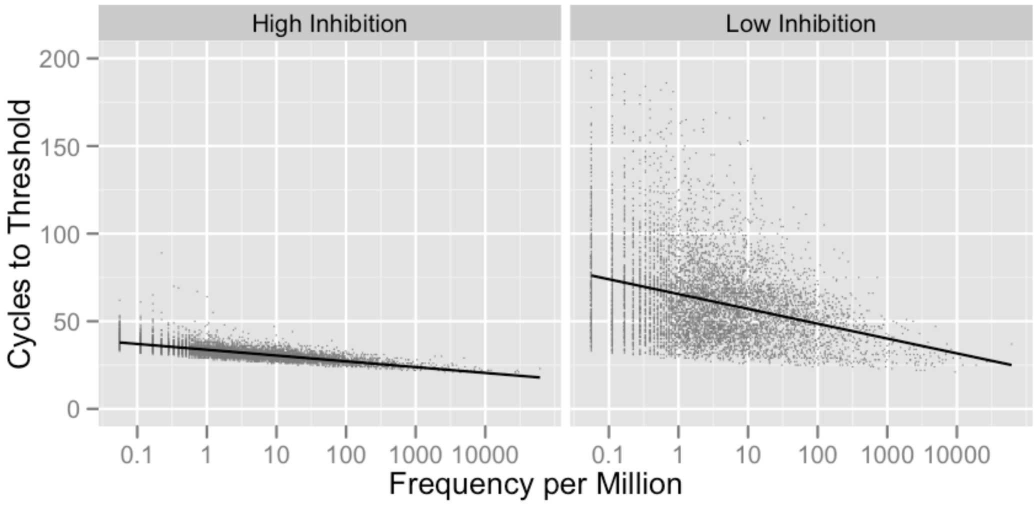

Figure 3. Cycles needed for a word identification response in the orthographic route of the bimodal IAM (Diependaele et al., 2010). Lexical selection as a function of lexical entrenchment is simulated via the word-word inhibition parameter. Lower lexical precision (i.e., low entrenchment) corresponds to low inhibition. Dots represent individual decision times for all known words (N=7439). Lines show the linear regression onto the corresponding log frequency scores (see Diependaele et al., 2010 for further details).

Whatever the exact mechanism, the above simulations show that the larger frequency effects in L2 could arise as a side effect of the overall reduced lexical entrenchment in that language, without having to assume cross-language competition as the basis of the observed differences. A critical prediction of the lexical entrenchment account is that irrespective of bilingualism, language dominance and language similarity, the same *quantitative* relation between proficiency and word frequency should arise. This is not predicted by the competition account because, even if proficiency effects are accounted for, (a) there will always remain larger competition for bilinguals than for monolinguals due to a larger lexical space, and (b) frequency effects in one language will depend on the level of form similarity with other known languages. This prediction is tested here.

The Present Study

We adopt a novel strategy in the study of frequency effects and bilingual processing. We will examine to which extent it is possible to explain the shape of the frequency curve in L2 on the basis of L2 language proficiency and whether this relationship also applies to differences in L1 proficiency. Unlike the language-competition hypothesis, the lexical entrenchment account predicts that the same quantitative relationship should exist between proficiency in the relevant language and frequency effects, irrespective of bilingualism, language dominance, and the similarity of L2 and L1. We will test these predictions by including several groups of speakers of English in our analysis of the relation between proficiency and frequency effects: L2 speakers who differ in their L1 (and hence in the degree of similarity between the respective L1 and English), and native speakers of English.

Before proceeding, it is important to note that although we have illustrated each account quantitatively in the IA framework, the present research purpose is not limited to this framework. Each view can also be translated, for instance, onto the more complex distributed connectionist framework (e.g., Harm & Seidenberg, 2004) or the unimplemented serial search framework (e.g., Murray & Forster, 2004). Hence the findings have wider theoretical conclusions, which we will return to in the discussion section.

To provide a fine-grained analysis, we will adopt a mixed-effects regression approach, in which we *quantitatively* consider the *shape* of *individual frequency curves*. The actual shape of frequency effects, although well-studied for native speakers (e.g., Keuleers, Diependaele & Brysbaert, 2010), is almost never considered in bilingualism research, let alone quantitatively and at an individual level. Researchers mostly prefer to make theoretical predictions in terms of categorical contrasts (e.g., high vs. low, greater vs. smaller, significant vs. non-significant; e.g., Gollan et al., 2008, 2011; Duyck et al., 2008). This of course severely limits the theoretical insights that can be gained and often renders the comparison of results very difficult, due to the heuristic definition of categories, subjective interpretation of effect sizes and - certainly in the domain of frequency - non-linear continuous relationships with performance. A notable exception to this practice (in the context of bilingualism) was recently provided by Whitford and Titone (2012; Figure 1). They recorded eye fixations from a large number of bilinguals while reading paragraphs in their first or second language (English or French). In a mixed-effects regression, they found that the slope of the continuously measured frequency effect on fixation times was steeper in L2 than in L1. Furthermore, the L2 frequency effect became smaller as a function of L2 exposure, whereas the L1 frequency effect grew larger.

Despite the more fine-grained continuous assessment and the opposite effects of L2 exposure (as a correlate of L2 proficiency) on L1 and L2, Whitford and Titone's study unfortunately does not allow us to draw conclusions with respect to the origin of the differential frequency effects. First of all, they did not compare the effect of L2 exposure on L2 frequency to that of L1 exposure on L1 frequency. Even more importantly, they did not compare these effects to those of monolinguals. Hence, both accounts outlined above are able to explain Whitford and Titone's results. According to the competition account, a higher degree of L2 exposure the leads to a lower degree of interference from L1 representations. According to the entrenchment account, L2 exposure leads to better L2 lexical integration. In both cases, smaller frequency effects are predicted with increasing L2 exposure, but only the entrenchment account predicts exactly the same relation between exposure and frequency effects in L1 and for monolinguals.

To investigate the issue of language competition vs. lexical entrenchment properly, we compared the frequency curves of French-English, German-English and Dutch-English (L1-L2) bilinguals in English word identification to those of English monolinguals, and tested to what degree English proficiency allows us to explain the observed individual differences *within* and *across* these groups. Given the considerably different distances between the respective L1's and English in terms of lexical similarity (e.g., Schepens et al., 2012), this can be considered a particularly strong test of the entrenchment account, as only this account predicts the same quantitative relation between proficiency and frequency effects in all groups.

Furthermore, for a detailed quantitative analysis of the frequency curve, it is critical to account for the frequency curve typically observed. Although Whitford and Titone (2012) studied frequency effects continuously and explicitly discuss the asymptotic behavior of frequency effects, they still modeled them by simple linear curves, even though the frequency curve is definitely non-linear (see, e.g., Keuleers et al., 2010). It is impossible to judge to what degree this linearity assumption has affected the quality of Whitford and Titone's (2012) conclusions. In the present study, we avoid this problem by estimating frequency curves using a non-linear expansion of the frequency values (i.e., see the Method section for more details).

The fact that Whitford and Titone (2012) did not assess L1 exposure / proficiency is in fact not surprising. A well-known problem in multilingualism research is to obtain a representative language skill measure that allows differentiating among individuals in the low *and* high range. If the same low-resolution (typically 5-7 points) questionnaires are used for L1 as for L2 (e.g., Marian, Blumenfeld & Kaushanskaya, 2007), these are likely to provide researchers with ceiling scores in the case of L1. Like Whitford and Titone, most researchers therefore limit the language assessment to L2. As a result, it is impossible to evaluate quantitative relationships both within and between L1 and L2 participants and thus to distinguish between the different hypotheses. A strategy that is often used in the L1 individual differences literature is to combine the results of several tests (e.g., Andrews & Lo, 2012). From a methodological perspective, the question whether or not differential frequency effects in a bilingual and monolingual context can be predicted through the same quantitative relation with language skill is thus far from trivial.

We address this difficulty by using the scores of the LexTALE vocabulary test, recently published by Lemhöfer and Broersma (2012). This test consists of a non-speeded English lexical decision task and is specifically targeted at differentiating among highly proficient speakers (hence its name: Lexical Test for Advanced Learners of English). Although explicitly designed as a vocabulary test, Lemhöfer and Broersma (2012) have validated the score as a measure of general English proficiency. Several other studies support that vocabulary size and the ability to learn new words are central components of language skill (e.g., Braze, Tabor, Shankweiler & Mencl, 2007; Perfetti & Hart, 2002; Verhoeven & van Leeuwe, 2008). In a recent study by Andrews and Lo (2012), vocabulary size scores were also found to be critical determinants of lexical inhibition effects. It would thus appear that, at least within the entrenchment account, vocabulary size should be considered as a correlate of lexical selection difficulty and, hence, reduced frequency effects.

To recapitulate, we investigate (a) whether for French-English, German-English and Dutch-English bilinguals, individual frequency curves in an English word identification task reflect the same quantitative relation with LexTALE scores as observed for English monolinguals and (b) whether this quantitative relation fully explains frequency effect differences observed between these groups.

Method

To answer our research questions, we re-analyzed the data of Lemhöfer et al. (2008). These provide a unique opportunity in the present context for several reasons. First of all, this study is important because of its size. A detailed analysis of frequency effect curves requires a sufficient number of degrees of freedom and the dataset of Lemhöfer et al. meets this requirement. Word identification latencies were collected from 83 participants for 1025 monosyllabic English words in a progressive demasking paradigm (Grainger & Segui, 1990). Importantly, the words came from a broad frequency range. For the purpose of theoretical generality, it is also critical that apart from monolinguals (i.e., 20 native English participants), there were three different groups of bilinguals: native German (N=21), native French (N=21) and native Dutch (N=21). The Lemhöfer et al. data thus provide us with the opportunity to compare results in L1 to L2 performance across different types of bilinguals (i.e., bilinguals with different native languages), which is critical in the current context. The original analysis already revealed a larger frequency effect for the bilingual participants, but Lemhöfer et al. did not relate this difference to the proficiency scores of the participants. The data nevertheless provide an ideal situation to do so because LexTALE scores were obtained from both the monolingual and bilingual participants and, importantly, the variation in the proficiency scores (see Figure 4) provides scope to differentiate individuals both within and between the monolingual and bilingual groups. As discussed in the Introduction, the latter property is essential if we want to compare the relationship with frequency curves both within and across the monolingual and bilingual participants. For details on participants, materials, and the experimental procedure, we refer to the original article by Lemhöfer et al. (2008).

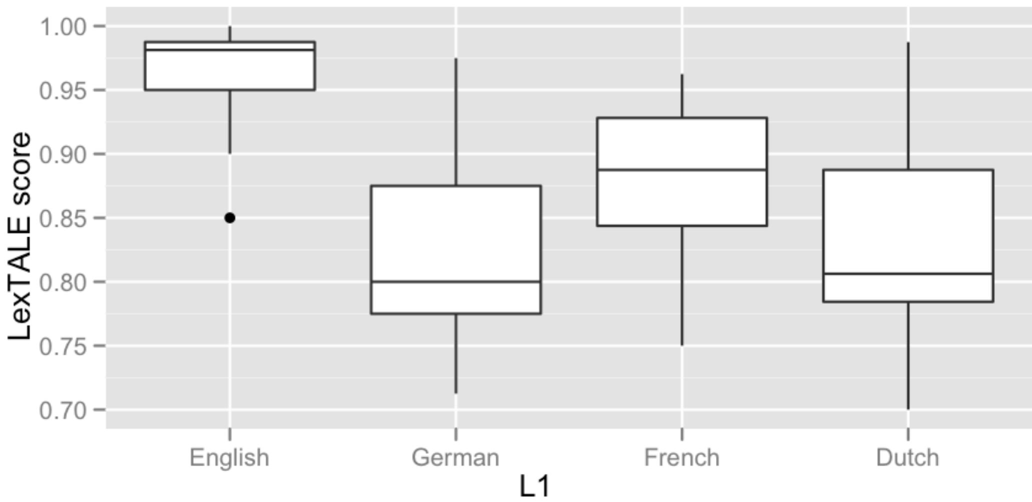

Figure 4. Box and whiskers plots of the LexTALE scores for the participant groups in Lemhöfer et al. (2008). Boxes show the first, second and third quartiles ($Q_{.25}$, $Q_{.50}$, $Q_{.75}$). Whiskers extend to the last values that fall within the range [$Q_{.25}$-1.5IQR : $Q_{.75}$+1.5IQR], with IQR = Inter-Quartile Range = $Q_{.75}$ - $Q_{.25}$.

In all analyses, we fitted non-linear mixed effects models onto the logarithmic reaction times (Log RTs) using the lme4 package (Bates, Maechler & Bolker, 2011) in R (R Development Core Team, 2011). The systematic variance with respect to mean Log RT by participants and words was modeled by estimating two separate Gaussian variances with respect to the intercept of the equation. We introduced three differences with respect to the original analyses of the frequency effect in Lemhöfer et al. (2008). First, we used the film subtitle frequencies of Brysbaert and New (2009) as the objective measure of word frequency. The analyses of Lemhöfer et al. were based on the written and spoken frequencies of the British National Corpus (BNC Consortium, 2001). Film subtitle frequencies provide a better estimate of everyday language use and explain more variance than the written BNC frequencies in lexical decision (see Brysbaert & New, 2009)[2]. These estimates were also used by Whitford and Titone (2012).

The second difference we introduced is that we modeled the typical non-linear (i.e., asymptotic) relation between reaction times and frequency via a natural spline expansion of the logarithmic frequency values. The mathematical details of this expansion are beyond the scope of the present study. We refer the interested reader to Harrell (2001). The design matrix for the current expansion was obtained from the rcspline.eval{Hmisc} function using 3 knots (Harrell, 2011). The main advantage of the natural spline expansion over the more standard (typically 2nd degree) polynomial expansion (see e.g., Lemhöfer et al., 2008) is that its behavior is defined locally, i.e., it provides a piecewise polynomial fit. If the data change in only a small region, this can drastically change the global shape of a regular polynomial fit. Spline functions protect against such behavior and therefore provide much better generalizability, without necessarily increasing the number of model parameters. Using 3 knots in the natural spline expansion, as in the current analyses, introduces no extra complexity (i.e., parameters) in the model compared to the 2nd degree polynomial approach. Given the flexibility of a piecewise approach, it can also provide better insight into the specific form of non-linearity.

The final difference is that, along with the fixed interaction of frequency with L1 and proficiency, we captured residual differences in individual frequency slopes in two Gaussian variance parameters (i.e., by-participant random adjustments to the linear and cubic frequency components). This is essential if we want to assess which part of the individual frequency differences is accounted for by the fixed predictors

---

[2] It is noteworthy that Lemhöfer et al. (2008) found no difference between monolinguals and bilinguals with respect to spoken frequency. These frequencies - taken from the BNC corpus - were calculated on the basis of a relatively small sample, however (i.e., 124 individuals), potentially leading to a high level of idiosyncrasy in the measure.

L1 and proficiency[3]. We also modeled covariance between the by-participant intercept and frequency-effect adjustments. As a result, we took into account the correlation between the participants' overall response speed and frequency slopes when testing the fixed effects of L1 and proficiency. This is important because, in principle, it is possible to predict that larger frequency effects are merely due to higher response thresholds. As shown in Figure 5, simply increasing the overall activity threshold required for word recognition in the IA framework not only slows down the overall response speed, but also leads to larger frequency effects. Especially since L2 and low-proficiency performance is typically associated with slower responses (e.g., Duyck et al., 2008), we thus need to take into account the relation between individual response speed and frequency slope in the current context.

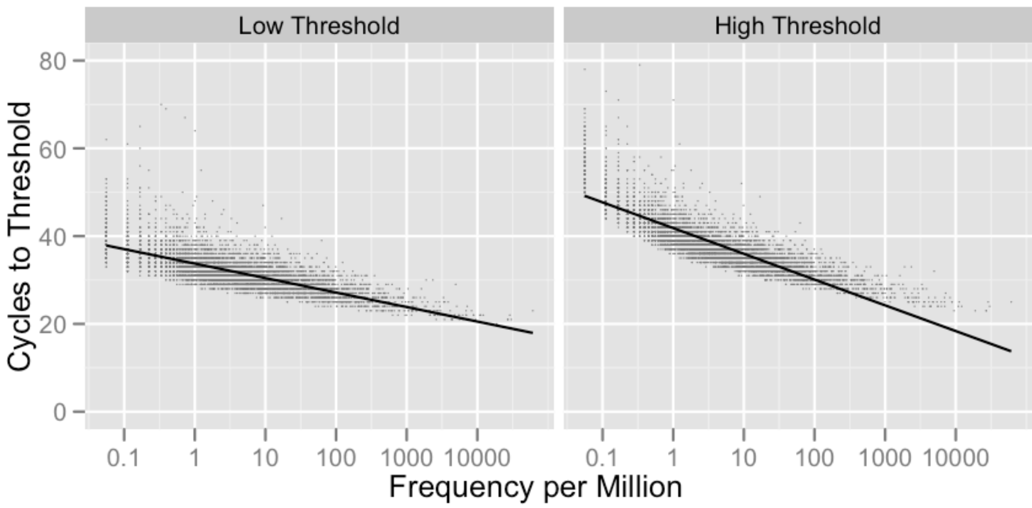

Figure 5. Cycles needed for a word identification response in the orthographic route of the bimodal IAM (Diependaele et al., 2010). Changing the activity threshold that the most active lexical representation needs to reach before a response is given leads to both overall slower responding and a larger frequency effect. Dots represent individual decision times for all known words (N=7439). Lines show the linear regression onto the corresponding log frequency scores (see Diependaele et al., 2010 for further details).

[3] We would like to thank an anonymous reviewer for pointing this out.

Results

Interaction of Frequency and L1

We start by reproducing the interaction between participant group and word frequency, reported by Lemhöfer et al. (2008). Figure 6 visualizes the results. A summary of the model is given in Table 1. To reduce collinearity, Log frequencies were centered to their mean value (i.e., the mean was subtracted from each value). The fitted model clearly replicates the earlier reported finding of larger frequency effects for bilingual participants (i.e., L1 = {Dutch, French, German}). As can be seen from the regression weights and their p-values, for English monolinguals, identification times decreased significantly as a function of frequency (see $\text{frequency}_{\text{linear}}$) and the decrease was significantly higher in the low range (see $\text{frequency}_{\text{cubic}}$). The six interaction terms show that for each bilingual group the linear and nonlinear effects were more pronounced. The fitted curve in Figure 6 further illustrates that the differences of interest (i.e., steeper frequency curves in a nonnative context) are situated in the lower range (<100 per million).

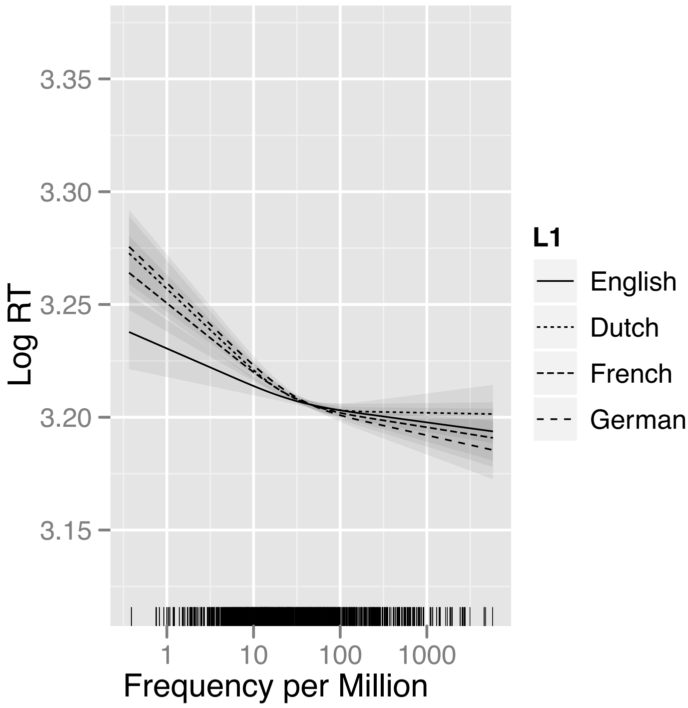

Figure 6. Reproduction of the *Frequency x L1* interaction in Lemhöfer et al. (2008) using movie subtitle frequencies, restricted cubic splines (3 knots – equally spaced between the .1 and .9 quantiles) and random frequency terms. Lines show the predicted frequency curves for the 4 participant groups together with 95% confidence bands.

Table 1. Summary of the *Frequency x L1* model: estimated variances and weights for the random and fixed effects, respectively. Significance values for the fixed effects estimates are based on the t-distribution. These values tend to be anti-conservative, but especially so in the case of small datasets (present *df*=81926). The two frequency parameters represent the first and second component of the 3-knot natural spline expansion of frequency. The components represent an overall linear and local nonlinear (cubic) term respectively. *L1* was coded using a treatment contrast with *English* as the reference level.

Random effects	σ^2			
intercept|word	0.0005	***correlation***		
intercept|participant	0.0048			
$\text{frequency}_{\text{linear}}$|participant	0.0002	*-0.1780*		
$\text{frequency}_{\text{cubic}}$|participant	0.0001	*0.2510*	*-0.9460*	
residual	0.0797			
Fixed effects	β	***SE***	***t(81926)***	***p(>|t|)***
Intercept (English)	3.1891	0.015521	205.47	< 0.0001
$\text{frequency}_{\text{linear}}$ (English)	-0.0167	0.004412	-3.79	0.0002
$\text{frequency}_{\text{cubic}}$ (English)	0.0093	0.00505	1.83	0.0669
Dutch vs. English	-0.0064	0.021854	-0.29	0.7700
French vs. English	0.0196	0.021854	0.90	0.3689
German vs. English	0.0235	0.021854	1.08	0.2823
$\text{frequency}_{\text{linear}}$: Dutch vs. English	-0.0195	0.004981	-3.92	0.0001
$\text{frequency}_{\text{cubic}}$: Dutch vs. English	0.0194	0.005192	3.73	0.0002
$\text{frequency}_{\text{linear}}$: French vs. English	-0.0141	0.004982	-2.83	0.0047
$\text{frequency}_{\text{cubic}}$: French vs. English	0.0106	0.005193	2.05	0.0407
$\text{frequency}_{\text{linear}}$: German vs. English	-0.0200	0.004982	-4.02	0.0001
$\text{frequency}_{\text{cubic}}$: German vs. English	0.0134	0.005193	2.59	0.0097

Introduction of the Proficiency Scores

Following the replication of the *Frequency x L1* interaction in Lemhöfer et al. (2008) using subtitle frequencies, natural spline expansion and random frequency slopes, our next step was to extend the model with additional linear fixed effect terms to capture the potential relation between frequency and individual language skill (as measured in the LexTALE test score). Such an interaction can be expected on the basis of both theoretical approaches that we discussed in the Introduction. The critical question is whether or not this interaction can *fully* explain the group-level differences such that the previously observed *Frequency x L1* interaction is not significant anymore, at least within the lower range. In that case we do not need the assumption that language competition is responsible for the differential frequency effects observed in the previous analysis.

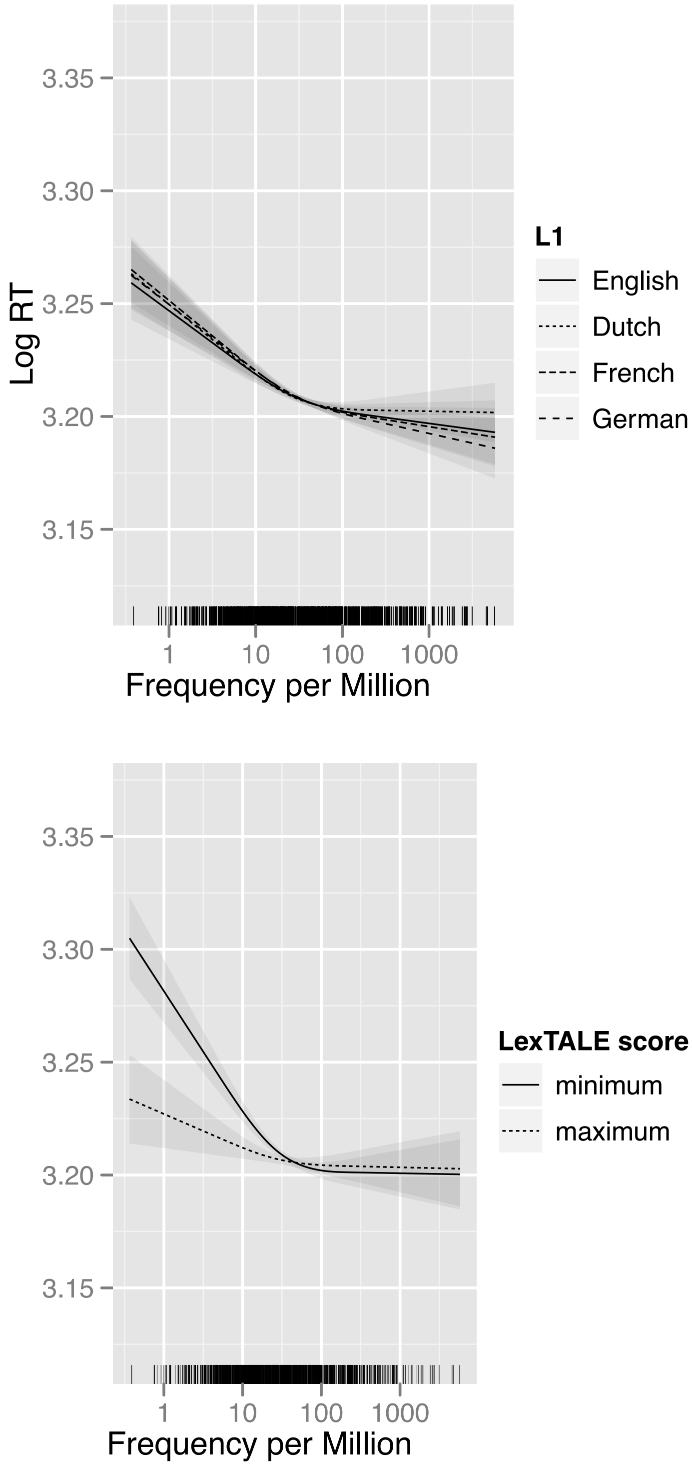

Figure 7. Explaining the *Frequency x L1* interaction through the interaction of frequency with language proficiency. The upper panel shows that the group differences regarding the predicted frequency curve (see Figure 6) disappear once LexTALE scores are introduced as a predictor in the model. The lower panel shows the predicted frequency curves and 95% confidence bands for the maximum and minimum LexTALE scores in the data (i.e.,70% and 100%).

Figure 7 and Table 2 provide the results of the present analysis. Most clearly, Figure 7 shows that by introducing the *Frequency x Proficiency* interaction the previously observed group differences, with steeper frequency curves for bilinguals, disappears. In Table 2 it can be verified that indeed none of the previously significant *Frequency x L1* interaction terms remains significant, whereas the interaction with proficiency is highly significant. The *Frequency x L1* interaction also does not contribute significantly to the overall model fit, as shown by a Log Likelihood Ratio Test comparing the full model (see Table 2) with a reduced version without the six Frequency x L1 terms: $\chi^2(6) = 9.73$, $p = .14$.

Table 2. Summary of the *Frequency x L1 + Frequency x Proficiency* model (see Table 1 for a description of the parameters). LexTALE scores were centered to their mean.

Random effects	σ^2			
intercept|word	0.0005	***correlation***		
intercept|participant	0.0048			
$frequency_{linear}$|participant	< 0.0001	*-0.1520*		
$frequency_{cubic}$|participant	< 0.0001	*0.2410*	*-0.9070*	
residual	0.0063			
Fixed effects	***β***	***SE***	***t(81923)***	***p(>|t|)***
intercept (English)	3.1979	0.0186	171.78	< 0.0001
$frequency_{linear}$ (English)	-0.0283	0.0044	-6.46	< 0.0001
$frequency_{cubic}$ (English)	0.0186	0.0053	3.54	0.0004
Dutch vs. English	-0.0190	0.0264	-0.72	0.4711
French vs. English	0.0113	0.0240	0.47	0.6368
German vs. English	0.0094	0.0274	0.34	0.7316
LexTALE score	-0.0972	0.1136	-0.86	0.3919
$frequency_{linear}$: Dutch vs. English	-0.0028	0.0050	-0.56	0.5758
$frequency_{cubic}$: Dutch vs. English	0.0058	0.0057	1.03	0.3041
$frequency_{linear}$: French vs. English	-0.0031	0.0045	-0.68	0.4951
$frequency_{cubic}$: French vs. English	0.0017	0.0051	0.33	0.7390
$frequency_{linear}$: German vs. English	-0.0014	0.0052	-0.26	0.7924
$frequency_{cubic}$: German vs. English	-0.0017	0.0059	-0.29	0.7747
$frequency_{linear}$: LexTALE score	0.1287	0.0214	6.03	< 0.0001
$frequency_{cubic}$: LexTALE score	-0.1042	0.0243	-4.28	< 0.0001

At this point, it is important to verify that *Proficiency* surpasses *L1* as an explanatory variable, i.e., that *Proficiency* explains more of the between-participants frequency differences than *L1*. Specifically, the factor *L1* is only able to account for group-level differences. *Proficiency*, on the other hand, can additionally account for within-group fluctuations and, importantly, on a numerical basis. The fact that the *Frequency x L1* interaction 'dissolves' into the *Frequency x Proficiency* interaction is a critical observation, but it can be argued that this could happen for any numerical predictor whose group-averages map onto the native-nonnative distinction. To see whether *Proficiency* does more than just explaining group-level differences (like *L1*), we need to verify that the introduction of *Proficiency* into our model also leads to a better account ('fit') of the data. This can be done by comparing model fits following the stepwise introduction of the *Frequency x Proficiency* interaction, i.e., a comparison of our first model (see Table 1) with a model including a simple effect of *Proficiency* and a comparison of the latter model with the model including both the simple effect of *Proficiency* and the interaction with *Frequency* (see Table 2). As shown in Table 3, although the inclusion of LexTALE scores per se does not increase the fit, the interaction with frequency does so significantly. It is thus clear that the individual LexTALE scores surpass the explanatory value of *L1*.

Table 3. Evaluation of model fits following the stepwise introduction of the *Frequency x Proficiency* interaction on the basis of Akaike information criterion (AIC), Bayesian information criterion (BIC), log likelihood (ln L) and log likelihood ratio (χ^2; see e.g., Kutner, Nachtsheim, Neter & Li, 2005, for details). Specifically, the models respectively correspond to the fixed-effects formulas:
Log RT ~ 1 + Frequency + L1 + Frequency x L1 (Table 1),
Log RT ~ 1 + Frequency + L1 + Frequency x L1 + Proficiency, and
Log RT ~ 1 + Frequency + L1 + Frequency x L1 + Proficiency + Frequency x Proficiency (Table 2).

Model	*df*	*AIC*	*BIC*	*ln L*	χ^2	*df*(χ^2)	*p*(> χ 2)
no vocabulary	20	-179232	-179046	89636			
+ vocabulary	21	-179230	-179035	89636	0.2226	1	0.6371
+ vocabulary x frequency	23	-179262	-179048	89654	35.897	2	< 0.0001

Consistency of Language Groups

A further critical component of the entrenchment account is that individual skill differences should not only explain group differences across mono- and bilingual participants, but also differences within these groups and, quantitatively speaking, in exactly the same way. We thus need to test whether the *Frequency x Proficiency* interaction yields similar estimates in the data of all four groups (i.e., L1 = {English, Dutch, French, German}). It remains possible that the bilingual participants dictated the results so far, since these made up about 75% of the data. To investigate this, we tested (a) whether frequency still interacted significantly with individual skill (i.e., LexTALE scores) when only the data of monolinguals are considered and (b) whether estimates would indicate a similar quantitative relation in each of the four groups.

The results indicate that this is indeed the case. The *Frequency x Proficiency* model for the monolingual data showed significant estimates for the two interaction terms and, importantly, the estimates are numerically very close to those in the *Frequency x L1 + Frequency x Proficiency* model for the full data (see Table 2).

Specifically, the estimates for $\text{frequency}_{linear}$: LexTALE and frequency_{cubic} : LexTALE were now 0.1312 (*SE* = 0.053, *t(20478)* = 2.48, *p* = 0.0133) and -0.1079 (*SE* = 0.0717, *t(20478)* = -1. 50, *p* = 0.1325), respectively. In the earlier analysis including all participants, these estimates were 0.1287 and -0.1042 (see Table 2). It is important to realize that, as illustrated in Figure 4, the same numerical relation between frequency and LexTALE scores arises for monolinguals and bilinguals irrespective of the fact that the distribution of the LexTALE scores is almost entirely non-overlapping. Figure 8 further illustrates the result by showing the predicted effects when the *Frequency x Proficiency* interaction was fitted on the data of each group separately. The results are clearly highly consistent: the analysis shows that we can predict the *Frequency x Proficiency* relation in the bilingual data by merely analyzing the monolingual data and vice versa. This provides a particularly strong case for the idea that we do not need to assume cross-language interaction as the source of the observed frequency-effects differences.

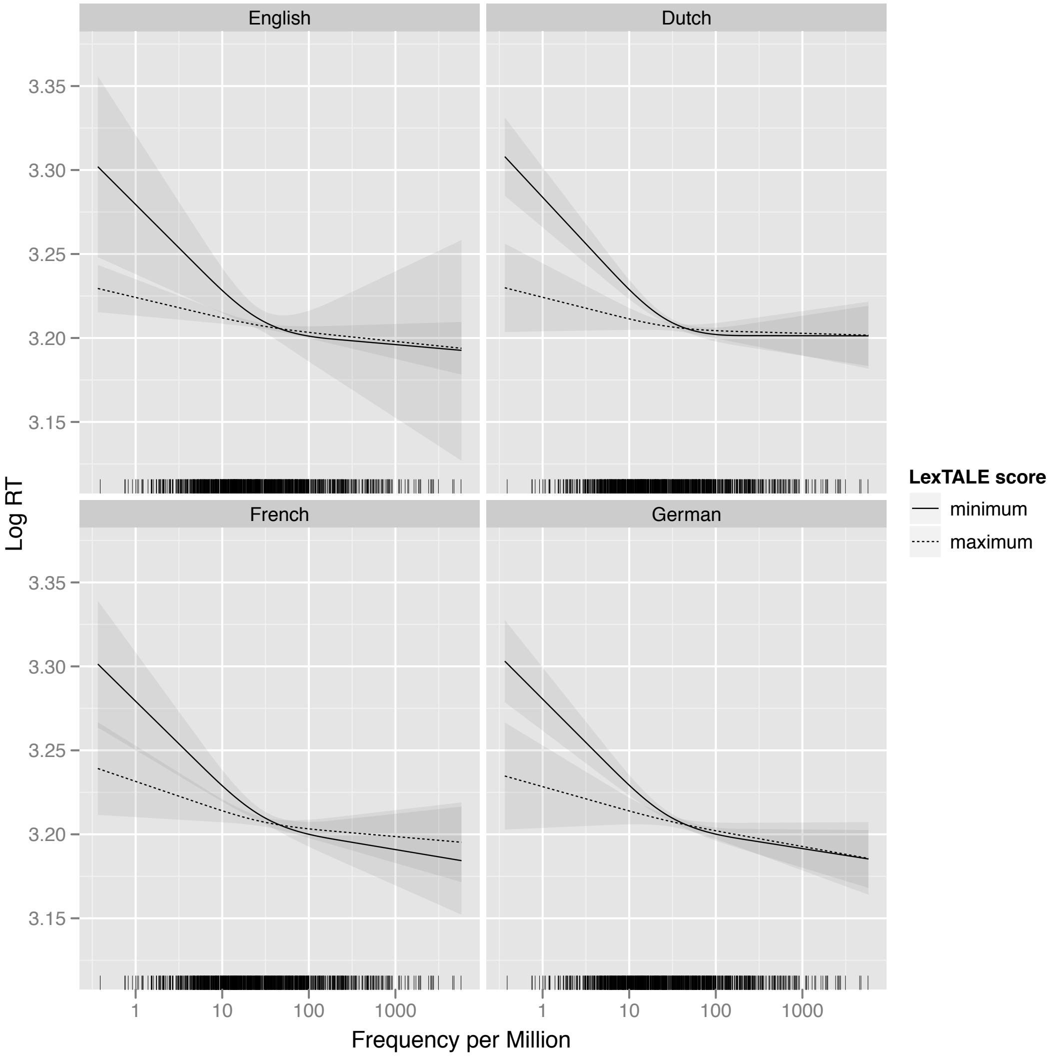

Figure 8. The *Frequency x Proficiency* interaction estimated separately from the data of each group. The lines represent the model predictions for the minimum and maximum LexTALE scores in the full data set (i.e.,70% and 100%). The 95% confidence bands in the monolinguals (L1 = English) analysis are considerably wider for the minimum LexTALE score. The reason is that this minimum falls well outside the range of the monolinguals scores (85-100%). The estimation is nevertheless made for the overall minimum to illustrate the consistent values for the parameter estimates (i.e., β weights).

Discussion

In the present study, we sought to explain the observation that the word frequency effect is larger in L2 than in L1. We investigated two alternative explanations that can be derived from the literature: the ‘language competition’ and the ‘lexical entrenchment’ accounts. As illustrated in Figure 1, competition from L1 representations can in principle cause larger frequency differences in L2 processing. The critical ingredients of this account are (a) that the frequency-effect difference is due to the structural difference between the bilingual and monolingual lexical systems, and (b) that the frequency effect will increase as a function of L1-L2 similarity. It also predicts that (c) the frequency effect will be *larger* for people with a larger lexical space, i.e., a larger vocabulary than for those with a smaller vocabulary.

The alternative account is the lexical entrenchment explanation, which emphasizes the strength/weakness of the lexical memory representations themselves. According to this explanation, extensive practice with words enhances the entrenchment of lexical representations, which implies faster activation and less interference from similar representations, leading to smaller processing differences between high and low frequency words. Entrenchment can be mapped onto various parameters. We have demonstrated reduced frequency effects in the IA framework with higher resting levels (i.e., subjective frequencies) and stronger word-word inhibition. The latter parameter links entrenchment to the concept of lexical ‘precision’ or ‘lexical quality’, which has been proposed for monolingual speakers by Perfetti (1997, 2002), among others, and recently received empirical support from Andrews and colleagues (Andrews & Hersch, 2010; Andrews & Lo, 2012). These authors found stronger orthographic inhibition from masked primes in participants with high scores of reading and writing proficiency and vocabulary size. In contrast to the language competition account, the entrenchment account predicts that (a) language proficiency (as measured by, for instance, vocabulary size) has the same effect in L1 and L2, and that (b) the relation between proficiency and frequency effects does not depend on the similarity between L1 and L2. Furthermore, and in sharp contrast to the first account, this account predicts (c) *smaller* frequency effects for larger vocabulary sizes, since larger vocabularies correlate with better entrenchment.

To distinguish between the two explanations, we ran a mixed-effects analysis of the relationship between English word identification times and proficiency collected by Lemhöfer et al. (2008). Such an analysis was possible because we had the same high-resolution measure of English proficiency (LexTALE; Lemhöfer & Broersma, 2012) for one monolingual English and three bilingual groups with English as L2 and different L1's (Dutch, French, and German). This enabled us to test the following critical questions: (a) Is there a relation between proficiency and the frequency effect, (b) To what degree does the interaction of frequency and proficiency explain the variance associated with group-level frequency differences?, and (c) How similar is the interaction for monolinguals and the different groups of bilinguals? Concerning (b) and (c), only the lexical entrenchment account predicts that the relationship between frequency effects and proficiency should be invariant across several groups of speakers (L1 speakers, or L2 speakers with different L1's).

Our analysis clearly supports the predictions of the entrenchment explanation. We found that the group-level interaction, reflecting steeper frequency slopes for bilingual than monolingual participants, was fully accounted for by the individual proficiency levels (i.e., the participants' LexTALE scores). Proficiency outperformed the explanatory value of the 'nativeness' of the language, and its effect could not simply be explained by a correlation with overall processing speed (which was taken into account in the analysis; see Figure 5 and the related discussion). Most importantly, despite an almost completely non-overlapping distribution of proficiency scores between monolinguals and bilinguals, and despite the differences in L1-L2 similarity among bilinguals, we found exactly the same quantitative relationship between frequency and proficiency for the monolingual and bilingual participants. It is important to realize that this in fact means that on the basis of the monolinguals' LexTALE scores, we are able to predict the size of the frequency effect of any bilingual as soon as we know their LexTALE score.

In our view, the most far-reaching conclusion to be drawn from these results is that basic individual differences in lexical processing – such as in the size of the frequency effect – can be attributed to a single causing factor, namely vocabulary size (or lexical proficiency) in the target language. Importantly, this factor explains not only differences *between* native and non-native speakers in terms of visual word recognition, but also differences *within* seemingly 'homogeneous' groups of speakers (monolinguals, or bilinguals with a particular L1-L2 combination) in exactly the same way. We can therefore conclude that, at least for the purpose of explaining differences in the size of the frequency effect, the assumption of qualitatively different lexical processing mechanisms between native and non-native speakers is unnecessary.

More specifically, our results indicate that interference between known languages is not a critical moderator of frequency effects. Since frequency remains the most important psycholinguistic variable in various tasks, the present study further illustrates that, although they are real, language competition effects in bilinguals should not be overestimated when building models of the bilingual lexicon (see also Lemhöfer et al., 2008; Davis, 2003, 2010). Being an IA model, the well-known BIA+

model of bilingual word recognition (Dijkstra & van Heuven, 1998, 2002) clearly provides an interesting case to study the balance between lexical competition and lexical entrenchment in bilingual word processing. As discussed in the Introduction, the concept of lexical entrenchment can be mapped onto the resting level and/or word-word inhibition parameters in the IA framework (see Figures 2 and 3). In the BIA framework, resting levels are generally lower for L2 words than for L1 words. It is therefore often claimed that the BIA(+) readily captures larger frequency effects in L2. Duyck et al. (2008, p. 853), for instance, say that “In the BIA(+) model (Dijkstra et al., 1998; see also Dijkstra & van Heuven, 2002), L2 words generally have lower resting-level activations than do L1 words of the same corpus frequency. Hence, BIA(+) would predict a larger FE in L2 than in L1, which is consistent with the present findings.”. However, the way lower resting levels for L2 are implemented in BIA(+) does not simply correspond to multiplicative downscaling as in our illustration in Figure 3 (see footnote 1). The minimum resting level is the same for L1 and L2 (i.e., -.92 see, McClelland & Rumelhardt, 1981), but the maximum resting level is lower for L2 words (-.3 instead of 0), which reduces the actual frequency range for L2 words. Correspondingly, when we implement the BIA(+) strategy in our illustrative model (i.e., “L1”: $-.92 \leq \text{RLA} \leq 0$ versus “L2”: $-.92 \leq \text{RLA} \leq -.3$), we obtain a very similar frequency curve and, consistent with the reduced RLA range in L2, the frequency effect is even smaller in L2 (linear regression weights: “L1”: -9.74, $SE = .31$ versus “L2”: -6.51, $SE = .27$). As Dijkstra and van Heuven (1998) noted themselves: “Future analyses of the development of ‘real’ human lexica over time are needed to determine how frequency differences can best be implemented …” (p. 201). It also remains to be seen whether the BIA+ model also captures the absence of a language competition influence on frequency effects. In principle, top-down inhibition by the so-called ‘language nodes’ could cancel any between language inhibition. The question is whether such strong language-selective behavior will still allow simulating benchmark language interference effects.

Our results are clearly also important for other bilingual processing models and different modeling frameworks than the IA framework. As for the BIA+ model, the critical question is whether differential frequency effects in L1 and L2 can be accounted for by language specific lexical characteristics (entrenchment) rather than language interference (competition). Duyck et al. (2008), for instance, discuss that the serial search framework (Murray & Forster, 2004) can account for larger frequency effects in L2 if language-independent lexical activation is taken into account. Under this assumption, word recognition in L1 and L2 would take the form of a frequency ordered serial search through the same pool of L1 and L2 words. Since L2 subjective frequencies are generally lower, larger frequency effects could be expected. The present findings challenge this account because, even when proficiency (affecting the relative search order) is taken into account, knowing more than one language should still result in larger frequency effects. No matter what, bilingual word recognition will always involve a greater search space compared to the monolingual case. This is clearly not supported by our findings, since proficiency fully accounted for the bilingual-monolingual difference.

In the distributed connectionist approach to bilingual word recognition (e.g., French, 1998; Li & Farkas, 2002) it would appear that entrenchment can be mapped onto the concepts of idiosyncrasy, redundancy and locality. As proficiency increases, the (‘hidden’) activity pattern resulting from a word input will grow more distinct from that of other word inputs and thus become more idiosyncratic, redundant and local. If recognition time is modeled as a function of this characteristic, it is readily predicted that frequency differences will become less salient with increased proficiency. However, since current implementations stress the integrated nature of processing different languages, it would seem that language interference effects can be particularly strong in these models. From this perspective, it seems likely that frequency effect differences will persist when proficiency is controlled for. As for all models discussed here, simulation studies are needed to investigate this.

Our results are further in line with those of Whitford and Titone (2012) who analyzed eye-movements during paragraph reading. They found that a higher degree of current L2 exposure leads to a smaller frequency effect when reading in L2, but to a larger frequency effect when reading in L1. The fact that frequency effects in L1 and L2 are a function of L2 exposure is clearly in line with the idea behind the proposed lexical entrenchment account, where higher representational strength leads to smaller frequency effects: the more time spent in an L2 context, the more opportunity to improve L2 lexical memory traces and the less opportunity to improve L1 lexical representations. L2 proficiency should thus lead to a smaller frequency effect in L2 and a larger one in L1. Our study nevertheless provides a critical extension to Whitford and Titone (2012), since their study did not allow them to decide between the two accounts under consideration here. The critical tool to distinguish between the language-competition and lexical representations account is a proficiency measure that allows testing the same interaction with frequency for both monolinguals and bilinguals. Only the lexical entrenchment account predicts the same quantitative relation. A further difference with Whitford and Titone is our nonlinear approach to the frequency curve. Given that in our analysis, the frequency differences concerning multilingualism only appeared to be evident for frequencies below 100 per million, we believe that the linear curves provided in Whitford and Titone (2012) should be handled with caution when formulating quantitative predictions in the future.

An important innovation of our study was the use of the LexTALE test (Lemhöfer & Broersma, 2012) as a covariate for response times. This test provided us with a proficiency measure that had a much higher resolution and also higher validity than the typical questionnaire measures (e.g., Marian et al., 2007), enabling us to differentiate not only among highly proficient bilinguals but also among English monolinguals (Figure 4). Even though the test only took a few minutes to complete, the scores turned out to be a very useful instrument to differentiate between participants. A similar observation was recently made by Khare, Verma, Kar, Srinivasan, and Brysbaert (2012). They started from the observation that the attentional blink effect is larger in bilinguals than monolinguals (Colzato, et al., 2008) and wondered whether the same difference would be found between high and low proficiency bilinguals. Testing a large sample of Hindi-English bilinguals, they found that they could replicate the effect, but only when English proficiency was measured with LexTALE. No correlation was found with the outcome of a language proficiency questionnaire. Therefore, we think inclusion of the LexTALE test should become standard in research on bilingualism.

A clearly interesting direction for future research is to exploit the greater resolution and precision of LexTALE proficiency scores in tasks where frequency effects are known to be more modest, such as naming and reading with eye-movement recording. When frequency plays a less important role, response variability that can be mapped onto proficiency in the data analysis is limited. Standard low-resolution proficiency scores will often lack power to investigate the frequency x proficiency interaction in such a situation, yielding uninformative null-effects. Frequency effects should nevertheless reflect the same principles of lexical entrenchment in different types of language processing. If a sufficiently high resolution is present in the proficiency score, we thus expect the same pattern as the present one: the same relation between proficiency and frequency effects across individuals with different language backgrounds. Of course, the exact quantities that define this relation will depend on the specific role frequency plays in the task at hand.

In summary, we have analyzed into great detail frequency effects in native and non-native word recognition. Our conclusion is that no qualitative differences need to be invoked to explain the commonly observed larger frequency effects in L2 than in L1. English word recognition times show the same quantitative relation to word frequency for natives and non-natives when proficiency is taken into account. The fact that exactly the same interaction of frequency and proficiency arises within natives and different groups of bilinguals provides a strong argument for the lexical entrenchment explanation. This conclusion provides a clear challenge for any computational model of bilingual word recognition. While these need to account for a certain degree of language interference, this interference does not seem to affect the influence of the most important variable in word recognition: exposure frequency.

References

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General*, *139*, 299–318.

Andrews, S., & Lo, S. (2011, August 29). Not all skilled readers have cracked the code: Individual differences in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication.

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

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (Release 2) [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*, 445-459.

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42. http://CRAN.R-project.org/package=lme4.

BNC Consortium. (2001). The British national corpus (Version 2) (BNC World). Oxford, England: Oxford University Computing Services [Distributor]. http://www.natcorp.ox.ac.uk.

Bowers, J. S., Davis, C. J., & Hanley, D. A. (2005). Interfering neighbors: The impact of novel word learning on the identification of visually similar words. *Cognition*, *97*, B45-B54.

Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary: Reading skill differences in young adults. *Journal of Learning Disabilities*, *40*, 226-243.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977-990.

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

Brysbaert, M., Van Dyck, G., & Van de Poel, M. (1999). Visual word recognition in bilinguals: Evidence from masked phonological priming. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 137-148.

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

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

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204-256.

Colzato, L.S., Bajo, M.T., van den Wildenberg, W., Paolieri, D., Nieuwenhuis, S., La Heij, W., & Hommel, B. (2008). How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 302–312.

Cristoffanini, P., Kirsner, K., & Milech, D. (1986). Bilingual lexical representation: The status of Spanish-English cognates. *Quarterly Journal of Experimental Psychology, 38A*(3), 367-393.

Davis, C. J. (2003). Factors underlying masked priming effects in competitive network models of visual word recognition. In S. Kinoshita & S. J. Lupker (Eds.), *Masked priming: The state of the art* (pp. 121-170), Hove, England: Psychology Press.

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review*, *117*, 713-758.

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

Diependaele, K., Ziegler, J., & Grainger, J. (2010). Fast phonology and the bi-modal interactive activation model. *European Journal of Cognitive Psychology*, *22*, 764-778.

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

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

Dijkstra, T., Hilberink-Schulpen, B., & van Heuven, W. J. B. (2010). Repetition and masked form priming within and between languages using word and nonword neighbors. *Bilingualism: Language and Cognition, 13*(3), 341-357.

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

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

Ellis, N. C. 2002. Frequency effects in language acquisition. *Studies in Second Language Acquisition*, *24*, 143-188.

French, R.M. (1998). A simple recurrent network model of bilingual memory. In M.A. Gernsbacher & S.J. Derry (Eds.), *Proceedings of the 20th Annual Conference of the Cognitive Science Society*, (pp. 368-373). Mahwah, NJ: Erlbaum.

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: General*, *140*, 186-209.

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

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, *103*, 518-565.

Grainger, J., & Segui, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception & Psychophysics*, *47*, 191-198.

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

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662-720.

Harrell, F. E., Jr. (2001). *Regression modeling strategies*. Berlin: Springer.

Harrell, F., E., Jr. and with contributions from many other users. (2011). Hmisc: Harrell miscellaneous. R package version 3.9-0. http://CRAN.R-project.org/package=Hmisc

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology*, *1*.

Khare, V., Verma, A., Kar, B., Srinivasan, N., & Brysbaert, M. (2012). Bilingualism and the increased attentional blink effect: Evidence that the difference between bilinguals and monolinguals generalizes to different levels of second language proficiency. *Manuscript submitted for publication*.

Kim, J., & Davis, C. (2003). Task effects in masked cross-script translation and phonological priming. *Journal of Memory and Language, 49*(4), 484-499.

Kuperman, V. & Van Dyke, J. A. (2012). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance*. In press.

Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2005). *Applied linear statistical models, 5th edition*, Boston, MA: McGraw-Hill/Irwin.

Lemhöfer, K. & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*, 325-343.

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

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwisterlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *34*, 12-31.

Li, P., & Farkas, I. (2002). A Self-Organizing Connectionist Model of Bilingual Processing. In: R. Heredia & J. Altarriba (Eds.), *Bilingual Sentence Processing*. North Holland: Elsevier Science Publisher.

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. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375-407.

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: Erlbaum.

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

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Amsterdam: John Benjamins.

Perry, C., Ziegler, J., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *114*, 273-315.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Schepens, J., Dijkstra, T., & Grootjen, F. (2012). Distributions of cognates in Europe as based on Levenshtein distance. *Bilingualism: Language and Cognition*, *15*, 157-166.

Sears, C. R., Siakaluk, P. D., Chow. V., & Buchanan, L. (2008). Is there an effect of print exposure on the word frequency effect and the neighborhood size effect? *Journal of Psycholinguistic Research*, *37*, 269-291.

Segui, J., & Grainger,J. (1990). Priming word recognition with orthographic neighbors: Effects of relative prime-target frequency. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 65-76.

Spielberger, C. D., & Denny, J. P. (1963). Visual recognition thresholds as a function of verbal ability and word frequency. *Journal of Experimental Psychology*, *65*, 597-602.

Van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 616-627.

Verhoeven, L., & van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, *22*, 407-423.

Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review*, *19*, 73-80.

Wickham, H. (2009). ggplot2: Elegant graphics for data analysis. New York: Springer.