

Interplay of bigram frequency and orthographic neighborhood statistics in language membership decision*

YULIA OGANIAN

Department of Education and Psychology, Freie Universitaet Berlin, Germany

Bernstein Center for Computational Neuroscience, Humboldt-Universität zu Berlin, Germany

MARKUS CONRAD

Department of Education and Psychology, Freie Universitaet Berlin, Germany

Universidad de La Laguna, Tenerife, Spain

ARASH ARYANI

Department of Education and Psychology, Freie Universitaet Berlin, Germany

HAUKE R. HEEKEREN

Department of Education and Psychology, Freie Universitaet Berlin, Germany

Bernstein Center for Computational Neuroscience, Humboldt-Universität zu Berlin, Germany

KATHARINA SPALEK

Institut für deutsche Sprache und Linguistik, Humboldt Universitaet zu Berlin, Germany

(Received: October 23, 2014; final revision received: April 29, 2015; accepted: April 30, 2015; first published online 2 July 2015)

Language-specific orthography (i.e., letters or bigrams that exist in only one language) is known to facilitate language membership recognition. Yet the contribution of continuous sublexical and lexical statistics to language membership decisions during visual word processing is unknown. Here, we used pseudo-words to investigate whether continuous sublexical and lexical statistics bias explicit language decisions (Experiment 1) and language attribution during naming (Experiment 2). We also asked whether continuous statistics would have an effect in the presence of orthographic markers. Language attribution in both experiments was influenced by lexical neighborhood size differences between languages, even in presence of orthographic markers. Sublexical frequencies of occurrence affected reaction times only for unmarked pseudo-words in both experiments, with greater effects in naming. Our results indicate that bilinguals rely on continuous language-specific statistics at sublexical and lexical levels to infer language membership. Implications are discussed with respect to models of bilingual visual word recognition.

Keywords: bilingualism, language membership decision, naming, sublexical units, lexical neighborhood, statistical learning, orthographic marker

Introduction

The ultimate aim of language comprehension is the extraction of a coherent and meaningful percept from a stream of noisy and often ambiguous input. In the bilingual case, the input can potentially stem from two different languages. This adds a level of complexity to language comprehension, since the two languages rely on different, and often contradicting, linguistic

structures. Thus, early recognition of the language of a word is potentially advantageous for language comprehension in several ways. First, it allows narrowing down possible interpretations of preceding and following inputs to those that fit the language of the current input at the single word as well as at the sentential level (Libben & Titone, 2009). Second, knowing the language of an input facilitates its recognition itself by narrowing lexical search to one language only (Casaponsa, Carreiras & Duñabeitia, 2014; Schulpen, Dijkstra, Schriefers & Hasper, 2003). The extent to which this is possible is of high importance for models of bilingual word recognition, which assume that lexical access is fundamentally language-unselective, meaning

* We thank Frederike Albers and Ulrike Schlickeiser for assistance in data collection and analysis of the voice files, Carsten Schlieve for technical assistance, and Assaf Breska for fruitful discussions of the manuscript. This research was supported by the Deutsche Forschungsgemeinschaft (GRK1589/1, doctoral scholarship to YO).

Address for correspondence:

Yulia Oganian, Freie Universitaet Berlin, Department of Education and Psychology, Habelschwerter Allee 45, JK25/215, 14195 Berlin, Germany
yulia.oganian@fu-berlin.de

Supplementary material can be found online at <http://dx.doi.org/10.1017/S1366728915000292>

that any visual input activates orthographic, lexical, and phonological representations in both languages (Conrad, Alvarez, Afonso & Jacobs, 2014; De Groot, Delmaar & Lupker, 2000; Dijkstra, Hilberink-Schulpen & van Heuven, 2010, but see also Costa, La Heij & Navarrete, 2006).

In natural settings, language membership information can be inferred from a variety of cues external to the word itself, such as sentential context or prior information about the speaker. Even in the absence of such external cues, however, bilinguals are able to reliably identify the language of a single word (Casaponsa et al., 2014; Grainger & Dijkstra, 1992; Vaid & Frenck-Mestre, 2002; van Kesteren, Dijkstra & de Smedt, 2012). For example, a trivial source for language membership information is the lexical word form itself, which – unless a cognate – is unambiguously associated with a certain language (Grainger & Dijkstra, 1992). Differences between orthographic scripts (i.e., English–Chinese), or unique graphemes (i.e., the letter Ø in Norwegian vs. English, van Kesteren et al., 2012) can also cue language membership. Even in the case of a shared orthographic system, two-letter combinations (*bigrams*) that are orthographically legal in only one of two languages (i.e., *TX* in Basque vs. Spanish, Casaponsa et al., 2014, *OE* in French vs. English, Vaid & Frenck-Mestre, 2002), termed *orthographic markers*, can speed language categorization and lexical access. A recent extension of the most prominent cognitive model of visual word recognition, the bilingual interactive activation model plus (BIA+, van Kesteren et al., 2012), incorporates the effects of orthographic markers on language membership decisions through inclusion of separate sublexical and lexical language nodes, which represent the language membership of language-unique units, i.e., orthographic markers and lexical word-forms, respectively. Importantly, in this model, the language membership of a letter string can be identified based on sublexical information alone. This is in line with the suggestion that orthographic markers allow for language identification without processing of the complete letter string or access to lexical representations (Vaid & Frenck-Mestre, 2002). However, the amount of lexical activation during language decisions has not been explicitly tested.

The evidence reviewed so far focuses exclusively on dichotomic language membership cues, such as language membership of lexical word forms and orthographic markers, which exist in one language only. However, ample evidence suggests that the visual word processing system is sensitive to continuous statistical patterns in the native language. For instance, lexical orthographic neighborhood size, a measure of lexical co-activation during lexical search (Andrews, 1997), is negatively correlated with RT on a range of tasks, such as lexical decision and naming (Carreiras, Perea & Grainger, 1997;

for a review see Andrews, 1997). Similarly, a large body of findings shows that the frequency of sublexical units modulates word processing in different languages (syllables: Carreiras, Alvarez & de Vega, 1993; Conrad, Carreiras, Tamm & Jacobs, 2009; Conrad, Grainger & Jacobs, 2007; Conrad & Jacobs, 2004; morphemes: Deutsch, Frost, Pollatsek & Rainer, 2000; Frost, Kugler, Deutsch & Forster, 2005; bigrams: Westbury & Buchanan, 2002). Furthermore, Bailey and Hahn (2001) concurrently manipulated word-likeness at sublexical and lexical levels and demonstrated a unique contribution of each of the two levels to word-likeness ratings of visually and auditory presented pseudo-words. They found a correlation between word-likeness ratings and lexical neighborhood sizes as well as with sublexical bigram frequencies.

In bilinguals, differential effects of within- and between-language orthographic neighborhood sizes on word processing were found (De Groot, Borgwaldt, Bos & van den Eijnden, 2002; Lemhöfer, Dijkstra, Schriefers, Baayen, Grainger & Zwitserlood, 2008). For example, Lemhöfer and colleagues (2008) showed that in a progressive demasking task within-language neighbors have stronger effects on word recognition than between-language neighbors. Moreover, Conrad et al. (2014) recently found that bilingual visual word recognition is sensitive to the frequencies of syllabic units in both languages. In particular, processing of sublexical units in first and second language (L1 and L2, respectively) seems to benefit from sublexical frequencies in the respective non-presented language.

Given monolinguals' ability to use statistical linguistic information and bilinguals' sensitivity to linguistic structure of both of their languages, the question arises whether bilinguals can use statistical information for language membership identification. Namely, differences in frequencies of occurrence between languages could be used as cues to determine language membership. This is only possible if bilinguals are able to assess sublexical (i.e., bigram frequencies) as well as lexical (i.e., orthographic neighborhood sizes) statistics separately for L1 and L2. The alternative would be that they represent statistical information in a way that summarizes across languages, in which case statistical information could not be used in language membership decisions. For example, in the former case a German–English bilingual would correctly realize that *forn* is more English-like than German-like although it is orthotactically legal in both languages. By contrast, the latter case would implicate that the same bilingual would only be able to estimate how word-like *forn* is without differentiating between languages. The effect of fine-grained sublexical and lexical statistical information in language membership decisions has so far not been investigated systematically (though note that both alternatives are possible within

models of non-selective lexical access). Bridging this gap is important for any comprehensive model of bilingual word recognition that describes language membership representations and accounts for their involvement in visual word recognition, such as the BIA+ model (Dijkstra & van Heuven, 2002; van Kesteren et al., 2012).

The present study

We designed the present study to shed light on the contribution of continuous sublexical statistics, measured through bigram frequencies, and lexical statistics, measured through orthographic neighborhood sizes, to language membership attribution. Moreover, we investigated whether continuous language membership information is ignored if a decision can be made based on an orthographic marker, as suggested by Vaid and Frenck-Mestre (2002). Finally, we asked whether the contribution of sublexical and lexical variables to language decisions depends on the output modality – comparing performance across decision and naming tasks.

First, we performed a corpus analysis to investigate the extent to which English and German words differ in their bigram frequencies and orthographic neighborhood sizes in English and German. We reasoned that only variables that are differently distributed in the two languages are potentially relevant for language membership decisions. For example, for bigram frequency to be informative about language membership, bigrams must exist that are more frequent in German than in English, and vice versa. At the level of orthographic neighborhoods, only if German words have more orthographic neighbors in German than in English, and vice versa, can the orthographic neighborhood of a word be diagnostic of its language membership.

In the second step (Experiment 1), we designed an experiment that probed the effect of differences in orthographic neighborhood sizes and bigram frequency measures between languages on language decision behavior of German–English bilinguals. To isolate the effects of differences in language-similarity statistics we employed pseudo-words (PWs). This allowed us to eliminate the influence of additional factors such as word frequency and semantics and rendered the exact manipulation of frequency variables more feasible. We built on results from the corpus analysis to identify the relevant range for each variable. Additionally, we contrasted the effects of continuously varying variables with the effect of orthographic markers (such as *pf*, as in *Pfanne* (*pan*), for German or *gh*, as in *laugh*, for English). We expected to find no effects of continuous differences in sublexical and lexical similarity to the two languages if bilinguals are sensitive to orthographic markers only. However, if bilinguals are sensitive to fine-grained statistical differences between languages,

continuous variation of sublexical and lexical statistics should affect bilinguals' language membership decisions.

In the third step (Experiment 2), we investigated the extent to which the effects of probabilistic cues depend on the output modality by contrasting the 2-alternative language decision task of Experiment 1 with a naming task. The language decision task requires an explicit decision between the two languages, based on sublexical and lexical cues. Contrary to this, naming requires a mapping from orthographic, sublexical, representations to language-specific phonology, which for pseudo-words does not necessarily require involvement of lexical representations (Coltheart, Rastle, Perry, Langdon & Ziegler, 2001). Thus we expected sublexical representations to play a larger role in the naming task than in the language decision task, and orthographic neighborhoods to play a larger role in the language decision task than in the naming task. A comparison of naming and language decision tasks will thus show whether language attribution in naming would preferentially be resolved based on sublexical orthographic and phonological information with decreasing involvement of lexical representations.

Corpus analysis

Methods

The corpus analysis was based on all spelling-corrected words of the German and the full English Subtlex databases (Brysbaert, Buchmeier, Conrad, Jacobs, Boelte & Boehl, 2011). We analyzed Levenshtein distance (OLD20; Yarkoni, Balota & Yap, 2008) as a measure of orthographic neighborhood size, and mean and maximal bigram frequencies as measures of sublexical frequency. These variables were calculated for each word of both corpora, separately for each language. This made it possible to examine the extent to which the statistical properties of a given variable differ between languages and whether words of one language are probable in the other language with respect to each variable. Since the focus of this research was on characterizing properties of letter strings that are unique for a certain language, identical interlingual homographs were excluded from this analysis.

Orthographic neighborhood size

The Levenshtein orthographic neighborhood distance of a letter string (Yarkoni et al., 2008) is its average distance to its 20 closest Levenshtein orthographic neighbors in a lexicon. The Levenshtein distance between two letter strings is computed as the minimal number of letter deletions, insertions, and changes that is needed to transform their orthographic word forms into each other. We computed OLD20 using the “vwr” library in

the statistical package R (Keuleers, 2013). OLD20 is a variable with a strong dependency on word length, such that if a specific word length is more common in a language, the orthographic neighborhoods of words of that length are denser than for other word lengths. To control for the difference in average word length between German and English and make OLD20 values in German and English comparable, OLD20 was mean-normalized for each word length within each language. All language comparisons were based on normalized scores. For each word in both corpora we computed the difference between its average Levenshtein distances to the 20 closest neighbors in the German Subtlex (OLD_G) and in the English Subtlex (OLD_E), and the difference between these variables ($diffOLD$). $diffOLD$ was coded such that positive values reflect a larger orthographic neighborhood in German than in English.

Bigram frequencies

Positional log10 token bigram frequencies were computed based on word form frequencies from the Subtlex databases and normalized to frequency per million bigrams. This allows better comparability across languages than normalization per million words, as German words are longer on average. Bigram frequencies were also mean-normalized within each language. For each word in the two corpora we computed the difference between the mean German bigram frequency (mBG_G) and the mean English bigram frequency (mBG_E), $diffBF$, of its constituent bigrams.

Furthermore, we hypothesized that language decisions might be guided not only by the average difference in bigram frequency between languages, but also by single sublexical units with a strong difference in language similarity. For each word we thus also identified its constituent bigram with the largest difference between its frequency in German and English. We denote this maximal bigram frequency difference as $maxdiffBF$, with positive values for high German-typicality and negative values for higher English-typicality.

Results and discussion

The descriptive statistics for the three difference variables are presented in Table 1 and their distributions are plotted in Figure 1.

Of the three difference variables, $diffOLD$ shows the largest difference between its distributions in English and German (Figure 1, upper panel), with more than 90% of words of each language having more orthographic neighbors in their language than in the respectively other language ($d' = 2.7$). While the distributions of $maxdiffBF$ are also similarly different between languages ($d' = 1.4$), there is a large overlap between the distributions of $diffBF$ in German and English ($d' = 0.24$).

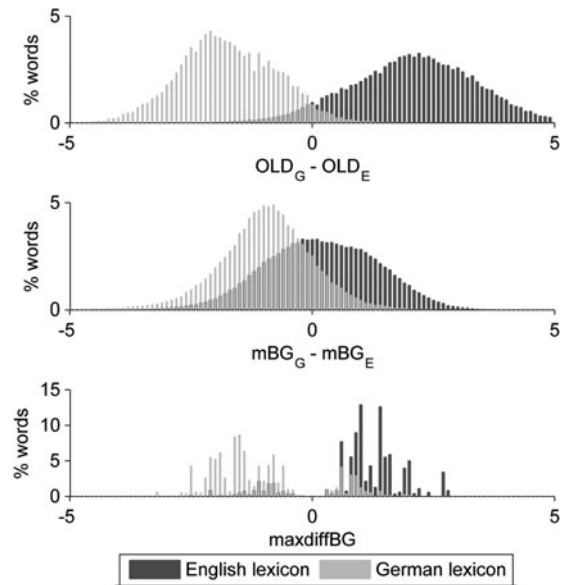


Figure 1. Distribution of differences in orthographic neighborhood size ($diffOLD$, upper panel), mean bigram frequency differences ($diffBF$, middle panel), and maximal bigram frequency differences ($maxdiffBF$, bottom panel) in German and English SUBTLEX lexica.

The corpus analysis, thus, reveals that lexical neighborhoods can be very informative for the language membership of a word, in line with previous findings showing larger same-language than cross-language neighborhoods (Marian, Bartolotti, Chabal & Shook, 2012). It also provides $maxdiffBF$ as a potential source of language membership information, whereas mean bigram frequency differences were more similar across languages. In the following two experiments we investigate the effects of all three variables on language attribution.

Experiment 1: Language Decision Task

Methods & materials

Participants

This study was conducted with highly proficient German–English bilinguals ($n = 25$, 5 male, ages 21–35, mean age 27 years). All were students at the Freie Universitaet in Berlin and had studied English as their first foreign language in high school. All had spent at least 9 consecutive months living in an English-speaking foreign country (GB, USA, English-speaking Canada, Australia, or New Zealand). Participants were right-handed, had normal or corrected-to-normal vision, and did not suffer from a reading disability or other learning disorders. All participants completed an online language history questionnaire (adapted from Li, Sepanski & Zhao, 2006) prior to participation and were only admitted to the

Table 1. Descriptive statistics of the difference variables, their pair-wise correlation in the English and German SUBTLEX lexica, and Cohen's *d* for the comparison between English and German distributions of each variable. Orthographic neighborhood size: $\text{diffOLD} = \text{OLD}_E - \text{OLD}_G$; mean bigram frequency difference: $\text{diffBF} = \text{BF}_G - \text{BF}_E$; maximal bigram frequency difference: maxdiffBF is the maximal bigram frequency difference across all bigrams of a word.

	Cohen's <i>d</i>	SUBTLEX Lexicon							
		English				German			
		mean	min	max	SD	mean	min	max	SD
diffOLD	2.7	-1.7	-5.4	3.3	1.0	2.1	-3.2	7.5	1.3
diffBF	0.2	-0.9	-5.9	3.4	0.9	0.2	-5.5	4.4	1.1
maxdiffBF	1.4	-1.0	-3.1	2.8	1.0	1.0	-3.1	2.8	1.0
Correlations									
		diffBF		maxdiffBF		diffBF		maxdiffBF	
diffOLD		0.22		0.27		0.34		0.38	
diffBF				0.3				0.28	

Table 2. Profiles of participants in Experiment 1 and 2. There were no significant differences between participants of Experiment 1 and 2.

		Experiment 1				Experiment 2			
		L1 (German)		L2(English)		L1 (German)		L2(English)	
		mean	SD	mean	SD	mean	SD	mean	SD
reading rate w/min	Lextale	90.8	6.9	80.5*	8.7	90.4	5.0	77.0*	13.0
	<i>real words</i>	131.4	10.2	124 *	9.0	130.0	14.0	123.2 *	9.5
	<i>pseudo-words</i>	83.5	17.5	81.3	13.5	83.5	20.0	78.6	10.6
Self report	<i>Age of acquisition (years)</i>	-	-	9.7	1.64	-	-	9.3	2.0
	<i>Proficiency¹</i>	-	-	6.0	0.5	-	-	5.9	0.4
	<i>Accent²</i>	-	-	2.4	1.0	-	-	2.7	0.8

Notes.

1. On a scale of 1 (basic) – 7 (native).

2. On a scale of 1 (no accent) – 7 (very strong accent).

* $p < .05$ for comparison between L1 and L2.

study if they fulfilled the above criteria. Self-reports of L2 proficiency were made on a 1-7 Likert scale, separately for reading, writing, speaking and listening abilities. The averaged self-estimated English proficiency was 5.9 (range 4.5–7). Additionally, participants' general proficiency and reading abilities were assessed after the experiment using the LEXTALE tests of German and English proficiency (Lemhöfer & Broersma, 2011). The tests consist of short lexical decision tasks which include words of varying frequency and pseudo-words. The final score is the average percentage of correct responses to words and pseudo-words. Reading abilities were assessed using the reading and phonological decoding subtests of the TOWRE (Torgesen & Rashotte, 1999) for English and the word and pseudo-word reading subtests of the SLRT-II test (Moll & Landerl, 2010) for German. Both tests assess

reading rate (words/min) in single-item reading of words and pseudo-words (PW). Participants' language profile is summarized in Table 2. Participants were recruited through advertisements on campus and in mailing lists for experiment participation. All participants completed an informed consent form prior to beginning the experiment. They were reimbursed either monetary or with course credit.

Stimuli & Design

The stimulus set contained 192 marked pseudo-words, half of which were orthographically legal in German only, while the other half was orthographically legal in English only, and 192 neutral pseudo-words which were orthographically legal in both languages. Within the set of marked pseudo-words diffOLD and diffBF were

Table 3. Properties of neutral pseudo-words stimuli of Experiments 1 and 2. Neutral pseudo-words contained only bigrams that were legal in both languages. Orthographic neighborhood size: $\text{diffOLD} = \text{OLD}_E - \text{OLD}_G$; mean bigram frequency difference: $\text{diffBF} = \text{BF}_G - \text{BF}_E$; maximal bigram frequency difference: maxdiffBF is the maximal bigram frequency difference across all bigrams of a word.

	distribution			correlations	
	mean	min	max	diffBF	maxdiffBF
diffOLD	-0.16	-2.60	2.40	0.24	0.38
diffBF	0.10	-2.70	2.30		0.60
maxdiffBF	0.05	-1.00	1.10		

varied parametrically. Within the set of neutral pseudo-words, diffOLD, diffBF, and maxdiffBF were varied parametrically. Pseudo-words were selected from the English lexicon project (Balota, Yap, Cortese, Hutchison, Kessler, Loftis, Neely, Nelson, Simpson & Treiman, 2007), a German pseudo-word database provided by authors MC and AA, as well as a pseudo-word set created specifically for this study. Note that all marked pseudo-words were composed of letters that exist in both languages (i.e., excluding the German letters “ä, ö, ü, ß”), such that decisions based on low-level visual pop-out effects would not be possible.

Neutral pseudo-words

Neutral PWs were orthographically legal and pronounceable, with the same number of phonemes and syllables when pronounced in either language. Selection of neutral pseudo-words was guided by three considerations. First, we ensured that for all three variables – diffBF, maxdiffBF, and diffOLD – the range of overlap between their English and German distributions was covered (Figure 1 for the distributions in the corpus and Table 3 for properties of the stimulus set). Second, for each of the variables half of the stimuli had positive values and half had negative values, such that an approximately equal number of neutral PW was German-like or English-like respectively in each of the three variables. Third, the pseudo-words were chosen such as to reduce the pair-wise correlations between the three variables as well as their correlations with pseudo-word length (see Table 3). Note that the correlation between diffBF and maxdiffBF in neutral pseudo-words is high, as the two variables rely on the same information source.

Marked pseudo-words

Marked PWs contained a letter combination (1–3 letters) that violated orthographic rules of one of the languages

Table 4. Properties of marked pseudo-words. Marked PWs contained at least one bigram that was legal in one language (i.e. the marker language) and had a frequency of 0 in monomorphemic words of the other language. Orthographic neighborhood size: $\text{diffOLD} = \text{OLD}_E - \text{OLD}_G$; mean bigram frequency difference: $\text{diffBF} = \text{BF}_G - \text{BF}_E$; maximal bigram frequency difference: maxdiffBF is the maximal bigram frequency difference across all bigrams of a word.

	English-marked PW			German-marked PW		
	mean	min	max	mean	min	max
diffOLD	-0.90	-4.00	1.00	0.70	-2.00	2.70
diffBF	-0.70	-3.60	1.60	0.50	-2.40	3.10
maxdiffBF	-1.80	-2.70	-1.20	1.60	1.10	2.70

(such as th, ght, or word-final y for English, or pf, hl or word-middle sch for German markers). The bigram frequency of marker bigrams is not necessarily 0 in the respective other language if computed across the whole SUBTLEX, because these letter combinations could occur in rare cases at the boundaries between (otherwise free) morphemes, or in loan words and proper names. For example, the German marker *pf* occurs in English words such as *cuPFul* or *campFire* but never as a grapheme, within a morpheme, or syllable. Similarly, the English marker *th* occurs in German names (e.g., *FuerTH*), Greek loan words (e.g., *THEologie (theology)*), and as a bigram unifying two normally free morphemes (e.g., *achTHundert, (eight hundred)*), but not as a grapheme in other etymologically German words. Since all our PWs were mono-morphemic, we recomputed the frequencies for all orthographic markers based on mono-morphemic words (that is consisting of only one root morpheme plus a simple ending) of the relevant word lengths. In the resulting set of words marker frequencies were 0 in the respectively other language. As orthographic markers coincide with bigrams that define maxdiffBF values, marked pseudo-words were chosen to differ parametrically in their diffBF and diffOLD values only, whereas English-marked and German-marked pseudo-words were balanced with respect to maxdiffBF values (correlation between diffOLD and diffBF: $r_E = -.19$, $r_G = .42$, see Table 4 for descriptive statistics). Length and syllable number were matched across marker type conditions and did not correlate with experimental variables.

To further ensure that marked pseudo-words were indeed pronounceable and orthographically legal in one language only and neutral pseudo-words in both, all pseudo-words were rated for their pronounceability and orthographic legality in German and English by 3 native

speakers of German and English respectively. Only pseudo-words that were rated as pronounceable and orthographically legal in both languages by all referees were included in the study as neutral pseudo-words. Similarly, marked pseudo-words were only included if all referees rated them as pronounceable in the marking language and illegal in the other language.

Procedure

Participants performed a speeded language decision task, in which they were required to categorize whether a pseudo-word was more likely to be a German or an English word. Stimuli were presented on a 18" computer screen (Arial, font size 40) using the Psychtoolbox (Brainard, 1997; Kleiner, Brainard, Pelli, Ingling, Murray & Broussard, 2007) for MATLAB (version 7.10.0, Mathworks Inc., Natick, Massachusetts). A trial consisted of a fixation cross presented for 800 ms, followed by presentation of the stimulus until the participant responded, for at most 2 sec. Responses were given by button press on a custom-made response box with the two index fingers. Buttons were assigned to a language by German and British flags in the respective corners of the screen below the stimulus. Button assignment remained constant within but switched between blocks. On each trial response latency (RT) and language decision were recorded.

The task began with instructions and an example trial, followed by the 384 experimental trials, presented in 8 blocks of equal length¹. Pseudo-randomization ensured that not more than three equally marked items would appear consecutively. Participants could have self-paced breaks after each block. If a participant failed to respond before the time-out for more than three times he or she was reminded to respond faster. Subsequent to completion of the language decision task, participants completed the proficiency tests described above.

Statistical analysis

All statistical analyses were performed in the open source statistical programming environment R (R core team, 2013). RTs were analysed with linear mixed-effects models and language decisions were analysed using logistic mixed-effects models (Baayen, Davidson & Bates, 2008; Jaeger, 2008), as implemented in the lme4 package for R (Bates, Maechler, Bolker & Walker, 2014), and included subjects and items as random factors. Due to the complex random factor structure of our data, the implementation of maximal random factor structure was not practicable. Instead, we modelled the random factor structure with the intercept and the maximal order

interaction of fixed factors, as recently suggested (Barr, 2013). Significance of fixed effects was assessed through type-III Wald-tests of single parameter estimates. We follow the convention of the lme4 package by reporting the z-test statistic for logistic mixed-effects models and the χ^2 -test statistic for linear models (note though that both test statistics are equivalent).

Note that it is not possible to define correct and wrong responses for neutral pseudo-words, since most of them contain conflicting language membership information and no exclusive language cues. Thus, language decision analyses for neutral pseudo-words modelled the probability of English responses. However, for marked pseudo-words the marking creates a strong preference for the marker language. Thus, for marked pseudo-words we conducted analyses on the % marker-incongruent responses, to which we refer as error responses.

First, to examine whether orthographic markers biased language decisions, we analysed the effects of marker type (German (G) vs. English (E) vs. neutral (N))². Then the effects of continuous variables on language decision and response latencies were analysed separately for neutral and marked pseudo-words.

Outlier exclusion was performed for each participant and separately for marked and neutral PWs based on separate multiple regression models for each participant, containing the same factors that were used for RT analyses. The expected RT for each trial was determined and trials with a difference of more than 2.5 residuals' SD between expected and actual RT were labelled as outliers and discarded from further analyses.

Results

Data of all participants were included in the analyses. Outlier analysis led to a removal of 2.5% of the trials. For illustration purposes only, continuous variables were subdivided in three equally large subsets, of which the two extreme ones are depicted in figures, with the label "German" for the subset with most positive, i.e., most German-like values, and the label "English" for the most negative, i.e., most English-like values.

Effect of marker presence on language decisions

We analyzed the effect of marker language (G vs. E vs. N) on the probability of English responses with a logistic mixed-effects model, which included marker language as a fixed factor, intercepts for subject and items as random factors, and random slopes for marker language

¹ The first 2 blocks contained neutral pseudo-words only. Block type did not affect the variables of interest, and hence, results are reported collapsed across block types.

² Due to the high number of marker-congruent responses for marked PWs, there were not enough error trials (< 15 for most participants) to analyze RTs across marked and neutral PWs including response as factor. Thus we analyze RTs separately for marked and neutral PWs.

Table 5. Mean and standard deviations (SD) for RTs to marked and neutral PW in Experiment 1 and 2.

		RT					
		Response: German		Response: English		% Response English	
	Exp. 1	mean	SD	mean	SD	mean	SD
Markedness	English	853	315	746	204	90	29
	German	776	220	859	316	16	36
	Neutral	871	279	905	292	45	50
Exp. 2		mean	SD	mean	SD	mean	SD
Markedness	English	758	244	778	311	79	41
	German	730	215	792	307	17	38
	Neutral	712	211	768	318	42	49

within subject. Marker language was contrast-coded in the model, with the two orthogonal contrasts English-marked and neutral versus German-marked, and English-marked vs. neutral PWs. Marker language influenced the attribution of pseudo-words towards that language (E: 90%, G: 17%, N: 45% response English, see Table 5), $\chi^2(1) = 462.9, p < .001$. Planned comparisons, which were directly encoded in the model, showed that English-marked and neutral PWs were more often classified as English than German-marked PWs, $b = 2.08, SD = 0.11, z = 18.7, p < .001$, and that the same held for English-marked as compared to neutral pseudo-words, $b = 1.4, SD = 0.08, z = 18.0, p < .001$.

Effects of continuous variables in the absence of markers: Language decision

The analysis of language decisions on neutral PWs involved the continuous variables diffBF, maxdiffBF, and diffOLD, as well as all possible interactions between them as fixed effects. Random factor structure of the model included intercepts for subjects and items, as well as random slopes for the interaction of diffBF, maxdiffBF, and diffOLD. As expected, neutral pseudo-words were more often categorized as similar to the language in which their orthographic neighborhood was larger, $b = -0.28, SD = 0.07, z = 3.9, p < .001$ (Figure 2). No significant effects were obtained for maxdiffBF and diffBF or for any interactions.

Effects of continuous variables in the absence of markers: Response times

The analysis of response times to neutral PWs (Table 6) contained the fixed factors response language (G vs. E, dummy coded) and the continuous variables diffBF, maxdiffBF, and diffOLD. Random factor structure of the model included intercepts for subjects and items, random slopes for the interaction of response, diffBF, maxdiffBF, and diffOLD within subjects and random slopes for

response within items. Responses “German” were overall faster than responses “English” (see Table 5 for mean RTs). While there was no main effect for diffOLD, its interaction with response was significant (Figure 3b): increases in diffOLD (i.e., increasingly larger German compared to English neighborhood) resulted in faster reaction times ($b = -3.54$) for “German” responses, while slowing “English” responses ($b = -3.54 + 22.51 \approx 19$). A similar, marginally significant, pattern was found for maxdiffBF (Figure 3c; $b_{\text{response German}} \approx -12, b_{\text{response English}} \approx 14$). “German” responses were faster if maxdiffBF was positive (i.e., German-typical), whereas “English” responses were faster for negative maxdiffBF values (i.e., English-typical). The significant interaction of diffOLD and maxdiffBF (Figure 3a) resulted from faster responses for items with consistent diffOLD and maxdiffBF values, and slower responses when either positive diffOLD was accompanied by negative maxdiffBF values, or vice versa. In other words, responses were fast if both variables supported the choice of the same language and slowed if they pointed to different languages (e.g., when a PW had more orthographic neighbors in English than in German, but a more German-typical maxdiffBF value), supporting the importance of both variables for the decision process. All other main effects and interactions were not significant.

Interactions of markers and continuous variables³: Language decision

Analysis of % errors for marked PWs contained the fixed factors marker language (G. vs. E), and the continuous variables diffBF and diffOLD, as well as all of their interactions. Random factor structure of

³ In both analyses marker language was dummy-coded with English as baseline, such that main effects of continuous variables reflected their effect for English-marked PWs, and their interaction with marker language the addition in beta for German-marked PWs as compared to English-marked PWs.

Table 6. Summary of linear mixed-effect regression for reaction times to neutral PW in Experiment 1. Orthographic neighborhood size: $diffOLD = OLD_E - OLD_G$; mean bigram frequency difference: $diffBF = BF_G - BF_E$; maximal bigram frequency difference: $maxdiffBF$ is the maximal bigram frequency difference across all bigrams of a word.

Predictor	beta-estimate	SD	t-value	χ^2	p-value	Significance
Intercept	877.74	32.48	27.02	730.3	< .001	**
diffOLD	-3.54	6.00	-0.59	0.35	.56	
diffBF	-16.73	8.99	-1.86	3.46	.06	+
maxdiffBF	-11.58	9.57	-1.21	1.47	.23	
Response ¹	47.49	8.93	5.32	28.27	< .001	**
diffOLD : diffBF	12.59	9.31	1.35	1.83	.18	
diffOLD : maxdiffBF	-18.25	8.90	-2.05	4.21	.04	*
diffBF : maxdiffBF	-4.85	10.28	-0.47	0.22	.64	
diffOLD : Response	22.51	8.82	2.55	6.52	.01	*
diffBF : Response	1.65	12.13	0.14	0.02	.89	
maxdiffBF : Response	25.54	13.91	1.84	3.37	.07	+
diffOLD : diffBF : maxdiffBF	11.89	10.41	1.14	1.31	.25	
diffOLD : diffBF : Response	-15.24	13.09	-1.16	1.35	.24	
diffOLD : maxdiffBF : Response	9.62	13.00	0.74	0.55	.46	
diffBF : maxdiffBF : Response	-22.60	14.09	-1.6	2.57	.11	
diffOLD : diffBF : maxdiffBF : Response	-14.78	14.50	-1.02	1.04	.31	

Note. Significance of p-values: + $p < .1$; * $p < .05$; ** $p < .001$
 1. Responses are dummy-coded as 0 – German, and 1 – English, such that Response German is the reference label.

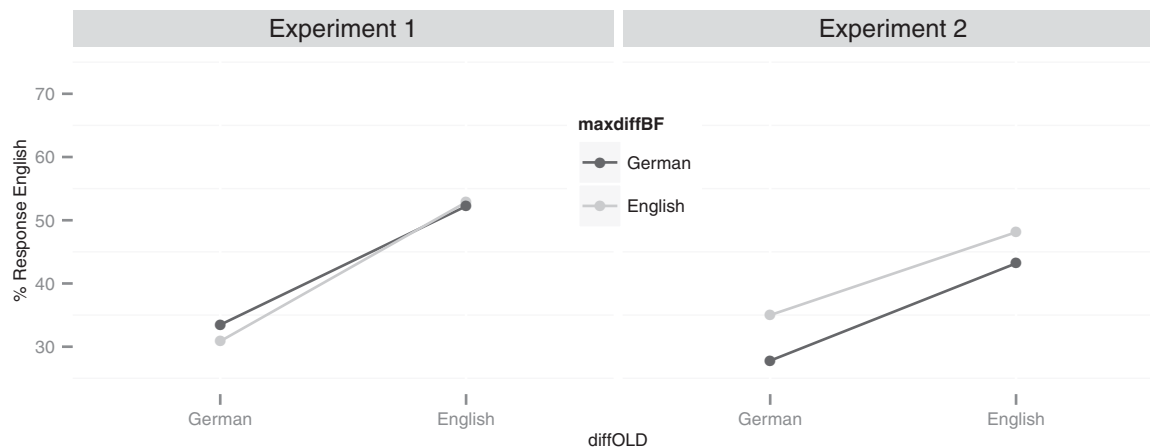


Figure 2. Visualization of the effects of differences in orthographic neighborhood size (diffOLD) and maximal bigram frequency difference (maxdiffBF) on language decisions for neutral pseudo-words in Experiments 1 and 2.

the model included intercepts for subjects and items, and random slopes for the interaction of response and diffOLD within subjects. Overall, above 80% of responses to marked pseudo-words were in line with their marker, although more marker-incongruent responses were made for German-marked than for English-marked PWs, $b = 0.6$, $SD = 0.22$, $z = 2.8$, $p = .006$ (G: 17%, E: 10% incongruent responses; see Table 5). Crucially, the effect of marker language interacted with diffOLD, $b = -0.6$, $SD = 0.2$, $z = -3.3$, $p < .001$ (Figure 4a). Planned comparisons showed an increase in marker-

incongruent responses for German-marked PWs with a decrease in diffOLD (i.e., more English neighbors than German neighbors), $b = -0.4$, $SD = 0.1$, $z = -3.7$, $p < .001$, whereas diffOLD had no effect on responses for English-marked PWs ($p > .1$).

Interactions of markers and continuous variables: Response times

Analysis of RTs to marked PWs involved the fixed factors marker language (G vs. E), diffBF, and diffOLD

Table 7. Summary of linear mixed-effects regression for reaction times to marked PW in Experiment 1. Orthographic neighborhood size: $diffOLD = OLD_E - OLD_G$; mean bigram frequency difference: $diffBF = BF_G - BF_E$; maximal bigram frequency difference: $maxdiffBF$ is the maximal bigram frequency difference across all bigrams of a word.

Predictor	beta-estimate	SD	t-value	X ²	p-value	Significance
(Intercept)	745.56	24.69	30.20	911.95	< .001	**
diffOLD	7.36	6.94	1.06	1.13	.29	
diffBF	-12.26	8.42	-1.46	2.12	.15	
Marker Language ¹	47.67	13.93	3.42	11.70	< .001	**
diffOLD : diffBF	-3.23	6.41	-0.50	0.25	.61	
diffOLD : Marker Language	-25.39	9.65	-2.63	6.93	.008	*
diffBF : Marker Language	3.79	10.93	0.35	0.12	.73	
diffOLD : diffBF : Marker Language	15.80	8.50	1.86	3.46	.06	+

Note. Significance of p-values: + $p < .1$; * $p < .05$; ** $p < .001$
 1. Marker language was dummy-coded with English as baseline.

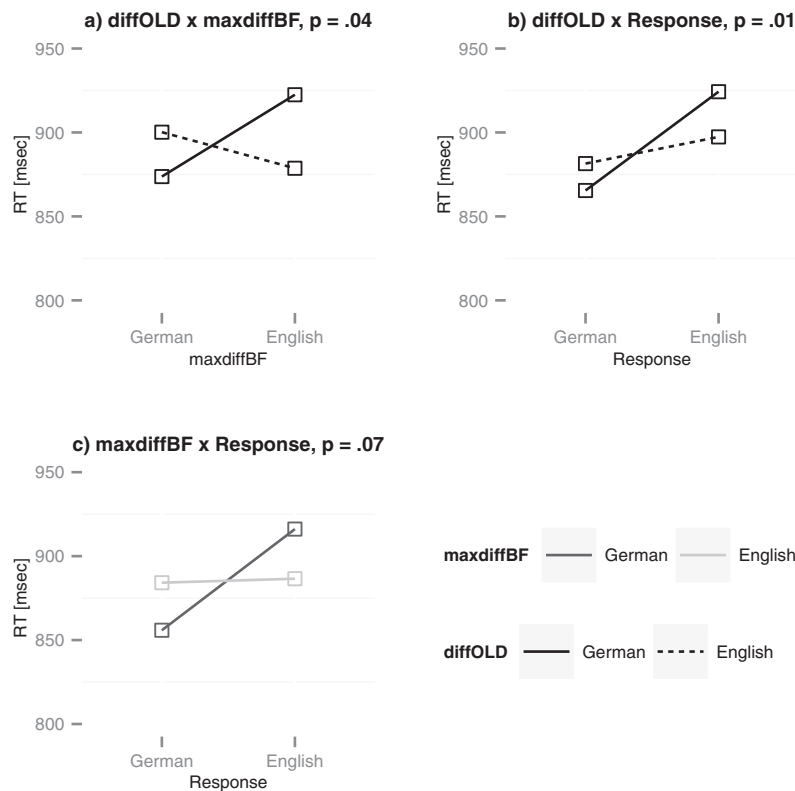


Figure 3. Visualization of the effects of differences in orthographic neighborhood size (diffOLD), maximal bigram frequency difference (maxdiffBF), and language decisions (response) on RTs to neutral pseudo-words in Experiment 1. a) Interaction effect of diffOLD and maxdiffBF; b) Interaction effect of diffOLD and response; c) marginally significant interaction effect of maxdiffBF and response.

(Table 7). Random factor structure of the model included intercepts for subjects and items, and random slopes for the interaction of marker language, diffBF, and diffOLD. Participants made 12 errors on average (English-marked: 2–24 error trials, German-marked: 6–33 error

trials), which was not sufficient for an RT analysis on error trials. Thus RTs were analyzed for correct responses only. The results largely mirrored the pattern of language decisions on marked PWs. Responses to German PWs were slower than responses to English PWs

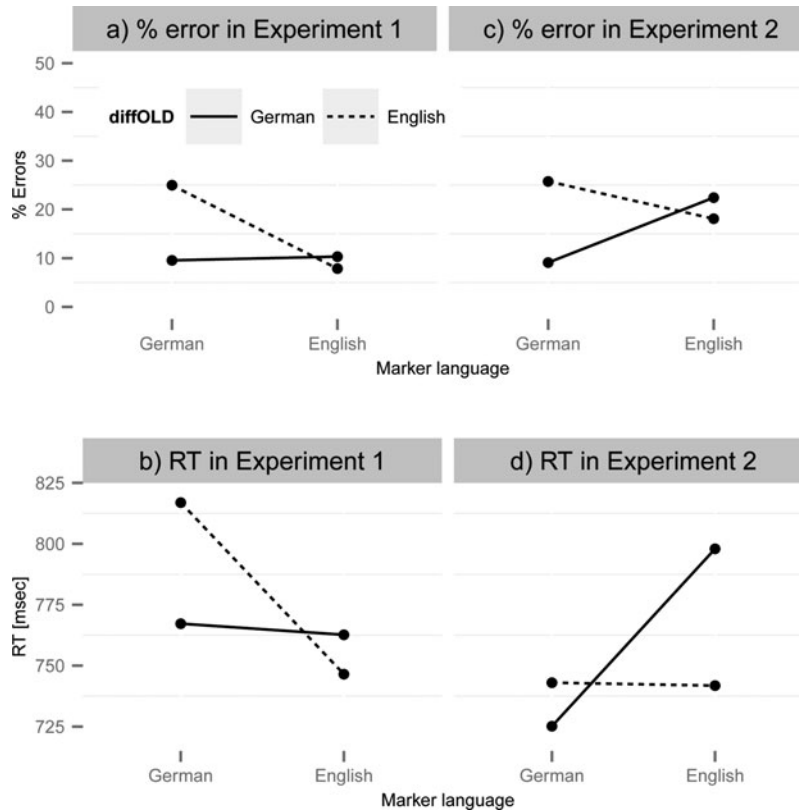


Figure 4. Visualization of interaction effects of differences in orthographic neighborhood size (diffOLD) and marker language on the % errors and RTs to marked pseudo-words in Experiments 1 and 2.

($b = 47.7$, see Table 7). Moreover, there was a significant 2-way interaction of marker language and diffOLD ($b = -25.4$, Figure 4b) and a marginally significant 3-way interaction between diffOLD, diffBF, and marker language ($b = 15.8$). A separate model for English-marked PWs showed no effects of the continuous variables for English-marked PWs ($p > .1$). However, the separate model for German-marked PWs showed that responses for German-marked PWs were slowed for larger English than German neighborhoods (i.e., decrease in diffOLD), $b = -18.9$, $SD = 6.5$, $t = -2.80$, $\chi^2(1) = 7.8$, $p = .005$. Additionally, this effect was attenuated for more positive diffBF values, $b = 12.0$, $SD = 5.3$, $t = 2.27$, $\chi^2(1) = 5.1$, $p = .02$. In other words, differences in orthographic neighborhood size mattered less if the mean bigram frequency difference was in line with marker language for German-marked PWs.

Discussion

Our data replicate previous studies (Casaponsa et al., 2014; Thomas & Allport, 2000; Vaid & Frenck-Mestre, 2002; van Kesteren et al., 2012) that showed that sublexical orthographic markers provide strong language

decision cues, as orthographically marked pseudo-words were mostly categorized in line with their marker.

Our data further extends previous findings in several ways. First, we show that continuous differences in lexical neighborhood size and maximal bigram frequency differences influence language decisions in unmarked pseudo-words. This was apparent in language attribution as well as in response latencies to neutral pseudo-words. In particular, response latencies increased when sublexical vs. lexical levels provided conflicting language membership information, suggesting that language membership information from both levels is integrated towards a language decision. Second, we also find that orthographic neighborhoods bias language decisions even for marked letter strings – as reflected by more marker-incongruent responses, and slowed marker-congruent responses for L1-marked PWs with larger orthographic neighborhoods in L2 than in L1. This effect was absent for L2-marked PWs, probably because our German-dominant participants could discard these pseudo-words as non-German based on their orthographic markers – representing violations of L1 orthographic patterns – alone. This also explains why correct categorizations of English marked PWs were faster than respective correct German categorizations (for a similar argument

see Casaponsa et al., 2014, as well as Vaid & Frenck-Mestre, 2002).

Next we asked whether the effects of continuous differences in language similarity would be preserved in a more natural task, where language decisions are necessary but are not made explicitly. To achieve this, we applied the design of Experiment 1 to a naming task, during which language categorization happens by choice of language-specific articulation (which differed between languages in our stimuli, see Table S1 in supplementary online material, Supplementary Material).

Experiment 2: Naming Task

Methods & materials

Participants

Twenty-four (6 male, aged 18 – 29, mean age 24) late German–English bilinguals from the same population as the participants of Experiment 1 participated in Experiment 2 (Table 2). None of the participants of Experiment 2 participated in Experiment 1.

Procedure

In the naming task participants were required to name (read out) pseudo-words. They were told that although the PWs were unknown to them, they could be read according to the pronunciation rules of either German or English. Participants were instructed to name the pseudo-words as fast as possible. Importantly, they were instructed not to choose a pronunciation language prior to naming but rather to read them out spontaneously and intuitively. An algorithm based on zero-crossings and power estimation was employed to register response onsets online. Based on the results of this algorithm the end of a trial was determined and participants were provided with visual feedback indicating that their response was registered. Responses were recorded with a headset microphone (Sennheiser PC 131 headset).

Results

The recordings were analyzed offline to determine naming latencies and response language by an independent highly-proficient German–English bilingual referee with the CheckVocal software (v. 2.2.2, Protopapas, 2007). Another referee reanalyzed a randomly chosen subset consisting of 50 neutral and 50 marked pseudo-words for each participant. Across referee reliability was above 95% for naming latencies and above 85% for response language. Stimuli, design, and presentation details were identical to Experiment 1. Naming latencies (RT) and response language were analyzed with the same procedures as in Experiment 1. In particular, we used the same mixed-effects modeling approach (including

random and fixed effect structures) as in Experiment 1. Outlier analyses lead to the exclusion of 3.5% of all trials.

Effect of marker presence on response language

As in Experiment 1, marker type (G vs. E vs. N) influenced the attribution of PWs towards a language (E: 79%, G: 17%, N: 42% English pronunciations, Table 5), $\chi^2(2) = 245.78$, $p < .001$. Planned comparisons showed that English-marked and neutral PWs were more often pronounced as English than German pseudo-words, $b = 1.8$, $SD = 0.14$, $z = 12.1$, $p < .001$, and that English-marked pseudo-words were more often pronounced as English than neutral pseudo-words, $b = 1.1$, $SD = 0.08$, $z = 12.7$, $p < .001$.

Effect of continuous variables in the absence of markers

Data of three participants who named less than 10 neutral PWs in English were excluded from this analysis, thus naming of neutral PWs was analyzed based on data from 21 participants.

Effect of continuous variables in the absence of markers: Response language

Similar to the results of Experiment 1, the probability for the choice of English pronunciations increased with the difference between English and German orthographic neighbourhood sizes, $b = -0.25$, $SD = 0.08$, $z = -3.1$, $p = .002$. Additionally and different from Experiment 1, the probability for the choice of German pronunciations increased with the German-typicality of maxdiffBF, $b = -0.23$, $SD = 0.1$, $z = -1.9$, $p = .06$, see Figure 2. All other main effects and interactions were not significant.

Effect of continuous variables in the absence of markers: Naming latencies

The linear mixed-effects model for naming latencies to neutral pseudo-words is summarized in Table 8. Naming was faster for PWs with more positive diffBF, i.e., high German-typicality at the sublexical level (Figure 5b) independently of response language ($\beta_{\text{German}} \sim -24$, $\beta_{\text{English}} \sim -33$). Naming latencies in both languages were marginally slower when diffOLD and maxdiffBF provided conflicting language information (Figure 5a). To elucidate the 3-way interaction of diffBF, maxdiffBF, and response (Figure 5c) we conducted LME models for each response language separately. There was no interaction of diffBF and maxdiffBF for “German” responses, suggesting that the 3-way interaction in the omnibus LME model was driven by trials with “English” responses. Indeed, for “English” responses, the interaction of diffBF and maxdiffBF, $b = 34.4$, $SD = 15.3$, $t = 2.27$, $\chi^2(1) = 5.2$, $p = .02$, reflected that the general speed-up of responses with increases in diffBF (i.e.,

Table 8. Summary of linear mixed-effects regression for reaction times to neutral PW in Experiment 2. Orthographic neighborhood size: $\text{diffOLD} = \text{OLD}_E - \text{OLD}_G$; mean bigram frequency difference: $\text{diffBF} = \text{BF}_G - \text{BF}_E$; maximal bigram frequency difference: maxdiffBF is the maximal bigram frequency difference across all bigrams of a word.

Predictor	beta-estimate	SD	t-value	χ^2	p-value	Significance
Intercept	746.39	33.27	22.44	503.3	< .001	**
diffOLD	-1.60	8.85	-0.18	0.03	.86	
diffBF	-24.07	10.99	-2.19	4.80	.03	*
maxdiffBF	19.02	14.01	1.36	1.84	.17	
Response	6.00	5.79	1.04	1.07	.30	
diffOLD : diffBF	-9.46	11.51	-0.82	0.68	.41	
diffOLD : maxdiffBF	23.68	12.98	1.82	3.33	.07	+
diffBF : maxdiffBF	19.08	11.12	1.72	2.94	.09	+
diffOLD : Response	-2.17	5.59	-0.39	0.15	.70	
diffBF : Response	-8.61	6.90	-1.25	1.56	.21	
maxdiffBF : Response	10.11	8.90	1.14	1.29	.26	
diffOLD : diffBF : maxdiffBF	12.49	12.99	0.96	0.93	.34	
diffOLD : diffBF : Response	-12.52	7.27	-1.72	2.96	.09	+
diffOLD : maxdiffBF : Response	13.61	8.25	1.65	2.72	.10	
diffBF : maxdiffBF : Response	17.26	7.05	2.45	6.00	.01	*
diffOLD : diffBF : maxdiffBF : Response	15.98	9.92	1.61	2.59	.11	

Note. Significance of p-values: + $p < .1$; * $p < .05$; ** $p < .001$

Responses are dummy-coded as 0 – German, and 1 – English, such that Response German is the reference label.

more German-typical values) was reduced for positive maxdiffBF (i.e., German-typical) values – presumably because particularly strong “German” evidence at one bigram position reduced the effect of the remaining bigrams.

Interactions of markers and continuous variables: Naming language

Overall, marking provided a very strong cue to the naming language, resulting in on average 80% of marker-congruent pronunciations (Table 5). The interaction of diffOLD and marker language, $b = .84$, $SD = 0.2$, $z = 3.9$, $p < .001$, was due to differential effects of diffOLD on % errors for German-marked and English-marked PWs (Figure 4c). For German marked PWs, increasingly more English than German neighbors lead to an increase in the number of marker-incongruent responses, $b = -0.6$, $SD = 0.15$, $z = -4.0$, $p < .001$, whereas there was no such effect for English-marked PWs ($p > .1$). There were no other main effects or interactions.

Interactions of markers and continuous variables: Naming latencies

Results of the analysis of RTs to marked PWs are summarized in Table 9. The main effect of diffBF was

marginally significant ($b = -26.87$), suggesting that naming was faster if mean bigrams frequencies were more positive (i.e., more German-typical), similar to the effect for neutral PWs. While there were no main effects of diffOLD and marker language, their interaction was significant ($b = -29.9$) – as expected from Experiment 1 – involving opposite effects of diffOLD for German- vs. English-marked pseudo-words. Namely, naming of German-marked PWs got faster with more German than English neighbors, whereas naming of English-marked PWs tended to be slowed, accordingly. When tested separately within each marker type, the effect of diffOLD was marginally significant for German-marked PWs, $b = -13.3$, $SD = 7.6$, $t = 1.74$, $\chi^2(1) = 3.04$, $p = .08$, but not significant for English-marked PWs, the latter probably due to larger variance in participants’ naming latencies to English-marked PWs. In summary, the pattern of effects on naming latencies for correctly named marked PWs was similar to the pattern for naming language choices, as well as the RT pattern in Experiment 1, as can be seen in Figure 4d.

Discussion

In Experiment 2, German–English bilinguals named the same PWs that were presented for language decision in Experiment 1. In general, naming in L1 was not faster

Table 9. Summary of linear mixed-effects regression for reaction times to marked PW in Experiment 2. Orthographic neighborhood size: $diffOLD = OLD_E - OLD_G$; mean bigram frequency difference: $diffBF = BF_G - BF_E$; maximal bigram frequency difference: $maxdiffBF$ is the maximal bigram frequency difference across all bigrams of a word.

Predictor	beta-estimate	SD	t-value	X ²	p-value	Significance
Intercept	774.75	38.66	20.04	401.60	< .001	**
diffOLD	15.83	12.30	1.29	1.66	.20	
diffBF	-26.87	15.11	-1.78	3.16	.08	+
Marker Language ¹	-25.21	24.37	-1.03	1.07	.30	
diffOLD : diffBF	-14.37	12.81	-1.12	1.26	.26	
diffOLD : Marker Language	-29.90	16.84	-1.78	3.15	.035	¹
diffBF : Marker Language	19.72	19.11	1.03	1.07	.30	
diffOLD : diffBF : Marker Language	20.56	16.33	1.26	1.59	.21	

Note. Significance of p-values: + $p < .1$; * $p < .05$; ** $p < .001$; ¹ one-tailed $p < .05$.
 1. Marker language was dummy-coded with English as baseline.

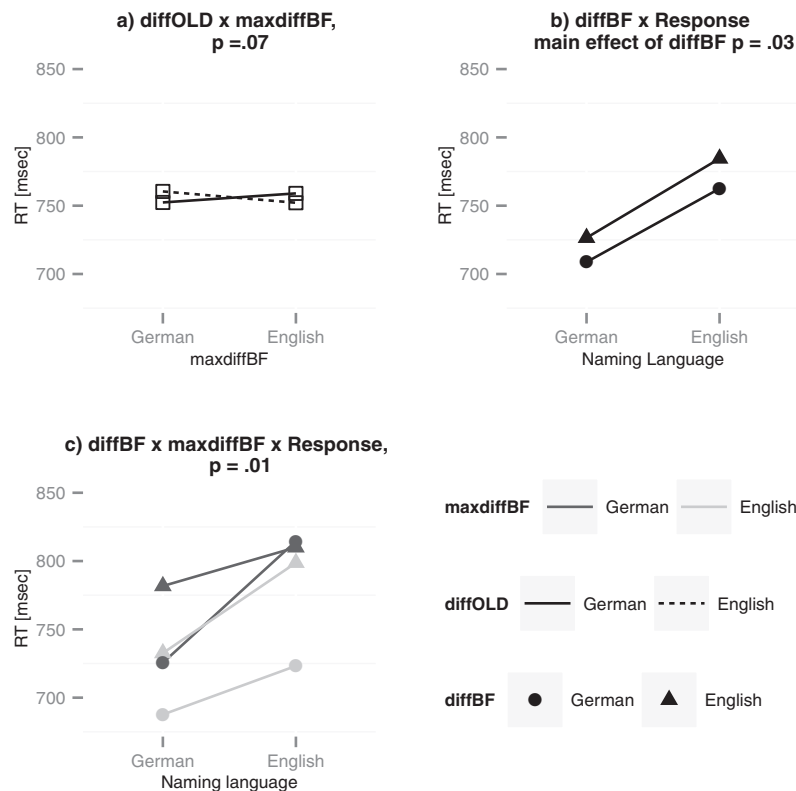


Figure 5. Visualization of the effects of differences in orthographic neighborhood size (diffOLD), maximal bigram frequency difference (maxdiffBF), mean bigram frequency difference (diffBF), and response language on RT to neutral pseudo-words in Experiment 2. a) Interaction effect of diffOLD and diffBF; b) Effect of diffBF, which was independent of the response language; c) Three-way interaction effect of diffBF, maxdiffBF, and response language.

than naming in L2, reflecting the relatively high L2 proficiency of our participants, even though descriptive statistics showed a tendency towards faster response times when naming in German than in English. As expected, continuous differences in orthographic neighbourhood

sizes as well as the language-typicality of single bigrams guided language attribution of neutral pseudo-words. We also found that – in both languages – naming latencies for neutral pseudo-words were reduced when mean bigram frequency was more L1-typical.

Large orthographic neighbourhoods in the non-marker language lead to more errors for L1-marked PWs only. As already discussed in Experiment 1, this is probably due to especially high sensitivity to violation of L1 orthographic patterns; such that respective items are immediately assigned to L2 before orthographic neighbourhoods can affect this decision.

Overall, the results of Experiment 2 further support the results of Experiment 1, suggesting that language membership information from all processing levels is integrated during naming, even in cases where categorical sublexical evidence in form of orthographic markers is available. Additionally they demonstrate that in a production task processing of letter strings with L1-typical sublexical structure is facilitated.

General discussion

The present study investigated the contribution of fine-grained sublexical and lexical statistical differences between languages to explicit language decisions in a language decision task (Experiment 1) and implicit language decisions in a naming task (Experiment 2), and contrasted them with the effects of orthographic markers. To create a task context most sensitive to small variations in language similarity we used pseudo-words, which have ambiguous language membership and no semantic meaning.

Our results show that subtle differences in language similarity at sublexical and lexical levels affect language attribution and naming of language-ambiguous letter strings, corroborating bilinguals' sensitivity to language-specific frequency statistics, despite generally language non-selective processing at all stages (Costa et al., 2006; Jared & Kroll, 2001; Kaushanskaya & Marian, 2007; Lemhöfer & Dijkstra, 2004).

Moreover, we extend previous studies on orthographic markers (Casaponsa et al., 2014; Vaid & Frenck-Mestre, 2002; van Kesteren et al., 2012) by directly showing that in the presence of orthographic markers of L1, but not L2, lexical information is assessed, as continuous differences in orthographic neighborhood sizes influenced processing of L1-marked pseudo-words. Our results provide a direct comparison of the effects of cross-language lexical neighborhood information for L1- and L2-marked PWs.

Effects of differences in sublexical frequencies

We manipulated two different types of continuous sublexical information, namely maximal (maxdiffBF) and mean (diffBF) bigram frequency difference.

The main effect of maxdiffBF for neutral PWs persisted beyond the effects of differences in lexical neighborhood sizes, suggesting that maxdiffBF captures a distinct, sublexical, source of language membership information.

In fact, for marked letter strings, maxdiffBF captures the frequency difference of the marker bigram, opening up the question whether in fact orthographic markers constitute the extremes of a continuous statistic, rather than being a different dichotomous variable.

Distributions of mean bigram frequencies in German and English differed only minimally in the corpus analysis. This is not surprising, as it is expected that two languages with similar morphological structure, such as German and English, will be similar in sublexical structure (cf. Marian et al., 2012 for similar findings on English and Spanish). As expected, given this high similarity between mean bigram frequency distributions in German and English, this variable did not cue language decisions. However, our data show that continuous differences in mean bigram frequencies affect response latencies. This effect was most prominent for neutral PWs in the naming task, but also marginally present for marked PWs in the naming task and for neutral PWs in the language decision task. We interpret this in terms of greater reliance on the sublexical reading route in the naming task than in the language decision task, as overt pronunciation of non-lexical items is required in the former but not in the latter task (Coltheart et al., 2001). The efficiency of the sublexical reading route depends on the typicality, and thus frequency, of involved sublexical representations. The fact that higher L1-similarity at the sublexical level led to shorter naming latencies, suggests that mapping of L1-typical bigrams onto phonology is especially efficient (Gollan & Goldrick, 2012). Alternatively, and without assuming a phonological source of this effect – that likely requires activation of language-specific phonological units – the present effect may also arise when participants activate (language-unspecific) those orthographic units more quickly that are more familiar to them, because they encounter them more often in their (constantly used) first than in their second language. We cannot distinguish between these options based on our data, as in both cases greater effects in the naming task for neutral PWs could be due to increased reliance on sublexical representations. The presence of a marginal effect for marked PWs in the naming task only is likely due to the fact that a single bigram (the marker) can suffice for language decision, whereas naming requires the complete mapping of all graphemes to phonemes (allowing for fine grained differences concerning other than the marker bigrams to influence naming latencies).

Note that differences between mean and maximal bigram frequencies were apparent in both tasks. While maxdiffBF affected language decisions and reaction times in the naming task, diffBF only had an effect on reaction times in both tasks. The effect maxdiffBF on reaction times was modulated by participants' response. Specifically conflict between language membership information stemming from maxdiffBF and diffOLD

slowed processing and agreement lead to faster responses. This pattern is most in line with its role as a source for language membership information. In contrast to this, the effects of diffBF are best interpreted in terms of an advantage for L1-similar letter strings, as high diffBF values speeded responses independently of the response language.

The different roles of diffBF and maxdiffBF in our study might be due to the specific languages at hand. Future research should investigate whether a comparison of languages with more distinct orthotactic structures will yield more pronounced effects of diffBF on language attribution.

Effects of differences in lexical neighborhood sizes

Language membership categorization in both experiments was guided by differences in orthographic neighborhood sizes. More specifically, neutral pseudo-words were preferentially categorized to the language that was predominant in their orthographic neighborhood. This converges with our corpus analysis, where strong differences between the number of within and cross-language neighbors for English and German words were apparent (c.f. Shook & Marian (2013) for similar findings on English and Spanish). It is also in line with other studies reporting effects of within-language and cross-language neighbors (Dijkstra et al., 2010; Midgley, Holcomb, van Heuven & Grainger, 2008; see also Jared, 2001).

We also find that lexical neighborhood statistics play a central role in language attribution of L1-marked but not L2-marked PWs. More specifically, for L1-marked PWs with more cross-language than within-language neighbors, erroneous categorizations were more frequent and correct categorizations were slowed. Interestingly, this was not the case for L2-marked PWs, the processing of which appeared unaffected by the number of orthographic neighbors from the two languages. Vaid & Frenck-Mestre (2002) already suggested that L2-specific orthography (violating L1 orthographic rules) allows for a rather perceptual (i.e., non-lexical) language attribution strategy. Our manipulation of diffOLD in marked PWs now allows directly estimating the extent of lexical search in the two languages during processing of marked PW. Results show that, indeed, perception of L2 markers was sufficient to trigger language decisions, while for L1-marked PWs mandatory activation of lexical neighbors seems to influence the decision process. In other words, violation of overlearned L1 orthographic patterns offers sufficient cues for language attribution, whereas the same does not hold for less well-represented orthographic patterns from L2.

Overall, our findings provide novel evidence for the activation of cross-language lexical neighbors for language-ambiguous as well as L1-marked pseudo-

words, demarcating a difference from the (lack of) effects of cross-language lexical neighbors reported for sentential monolingual contexts (Schwartz & Kroll, 2006). Furthermore, they offer an interesting insight into how sublexical orthographic cues might modulate the – otherwise consistently reported – activation of L1 lexical representations when reading L2. Namely, such cross-language activation of words from L1 might be restricted to cases of sublexical ambiguity whereas activation of lexical representations might focus more exclusively on the presented L2 language when orthographic patterns would be illegal in L1.

Comparison of Experiments 1 and 2

Overall, patterns driving language attribution were comparable across both tasks. But data from the two tasks also comprise interesting differences: First, while RTs to neutral PWs in Experiment 1 were shorter for “German” than “English” responses, this difference was not significant in Experiment 2. Presumably, for neutral PWs our German–English bilinguals tended to respond “German” by default, but respective effects seem to have decayed until phonological motor output could be produced. For marked PWs, however, RTs to English-marked PWs were significantly faster than to German-marked PWs in Experiment 1, while the opposite tendency was found in Experiment 2. This important difference aligns perfectly with specific task demands: In Experiment 1, lexical neighbourhoods could be blended out for English-marked but not for German-marked PWs – as participants seem to have been using a sublexical strategy responding especially quickly to violations of L1 orthographic patterns in the case of L2-marked PWs, which was faster than the integration of lexical and sublexical information for L1-marked PWs. In the naming task, on the other hand, already the need to produce less familiar L2 overt pronunciation may have cancelled out the processing advantage for L2 marked PWs of Experiment 1.

Second, in Experiment 2, there were more and greater effects of bigram frequency differences for neutral PWs than in Experiment 1, suggesting that language decisions in naming rely more heavily on language membership information from sublexical orthographic and phonological representation levels. Third, responses to sublexically German-typical PWs in the naming task were faster independently of the response language, suggesting that mapping to phonology is more easily accomplished for more L1-typical items, to which our participants have been more extensively exposed (Gollan & Goldrick, 2012). Fourth, effects of diffOLD on RTs for neutral and marked PWs were stronger in the language decision task than in the naming task.

This distinctive pattern of effects across the two tasks is well in line with the use of a sublexical grapheme-to-phoneme mapping route (Coltheart et al., 2001), which is independent of lexical activation, and becomes more relevant when phonological output has to be produced for previously not encountered letter strings.

In general, the comparison between the language decision and the naming task shows that language membership information is present at all levels of visual word processing, including the phonological level. The specific contribution of different sources of language membership information to language attribution appears to depend on task strategies and output modalities. In particular, lexical language similarity has a large impact in a language decision task, but its effects appeared reduced for a naming task, where responses may be produced with less activation of lexical representations (see e.g., Carreiras & Perea, 2004; Conrad, Stenneken & Jacobs, 2006). Accordingly, we suggest that, during naming, conflict in language membership information is propagated to the phonological level, where sublexical and lexical similarity statistics are integrated in an implicit language decision, made through the choice of the most active phonological units.

Integration in current models of bilingual visual word recognition

Our data provide ample evidence in favor of language membership representations at the sublexical level, not only for orthographic markers but also for shared bigrams with different frequencies in the two languages. Moreover, we find that language membership information is not only available for language-specific word-forms, as shown in previous studies (Casaponsa et al., 2014; Vaid & Frenck-Mestre, 2002; van Kesteren et al., 2012), but that continuous language similarity can be inferred for non-lexical items from the activation of orthographic neighborhoods. These findings challenge models of bilingual visual word recognition to allow for 1) continuous language membership information in addition to dichotomic language membership cues and 2) language membership information originating from sublexical representations. In particular, our data support the recent extension of the bilingual interaction activation model (BIA+, van Kesteren et al., 2012) that includes sublexical language nodes. Two features of this model appear especially relevant with regard to the present data.

First, dichotomic language membership cues, which previous studies mainly focused on, are usually implemented in terms of all-or-none activations of language nodes by language-unique units at lexical and sublexical levels. Such dichotomic cues could also be implemented as language tags ‘attached’ to single language-unique representations (for early

evidence against this concept see Grainger & Dijkstra, 1992). However, language tags are not compatible with continuous language membership information: at the sublexical level language-shared bigrams cannot be tagged unambiguously, whereas at the lexical level, language tags could only have an effect after identification of a word-form, which would contradict the orthographic neighborhood effects for pseudo-words in our study. Thus overall, our data provide further support for the concept of language nodes, as introduced by Grainger & Dijkstra (1992) and implemented in the BIA model and its recent extension to the BIA+ model (van Kesteren et al., 2012). Although language nodes in the BIA+ were conceptualized for dichotomic language membership information, they can be extended to include the effects of probabilistic language membership information. Namely, the strength of connections between sublexical and lexical units and language nodes could reflect the frequency of these units in the respective language – meeting computational principles of connection weights or frequency dependent resting levels, as suggested by e.g., Grainger & Jacobs (1996).

Second, van Kesteren and colleagues (2012) proposed to extend the BIA+ model with an additional set of language nodes accumulating language membership information from sublexical representations only (“sublexical language nodes”). Alternatively, one unique set of language nodes might accumulate lexical and sublexical language membership information in parallel. In principle, our data, as well as the data of van Kesteren et al., can be accommodated with either version of language membership nodes. However, a set of language membership nodes activated by language information from both levels of processing – lexical and sublexical – appears a more parsimonious solution, which would incorporate the general principles of interactive activation spreading over different representation levels. Future studies – including simulation studies and neuroimaging – are required to further test respective hypotheses.

Conclusions

Language membership information appears to be available at all levels of the visual word processing system. It may be delivered via definitive cues, such as orthographic markers, but as well via probabilistic cues, such as sublexical and lexical statistics. The present study provides ample evidence for the processing and interaction of language-specific sublexical and lexical statistics in the bilingual brain. It shows that despite bilinguals’ generally assumed language-independent way of processing, if required, probabilistic information can be retrieved for each language separately and can be used to guide the processing of linguistic input. Future studies

may specify the dependence of such findings on language proficiency and extend them to real word material.

Supplementary material

For supplementary material accompanying this paper, visit <http://dx.doi.org/10.1017/S1366728915000292>

References

- Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, 4(4), 439–461.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. doi:10.1016/j.jml.2007.12.005
- Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language*, 44(4), 568–591. doi:10.1006/jmla.2000.2756
- 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(3), 445–59. doi:10.3758/BF03193014
- Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, 4, 328. doi:10.3389/fpsyg.2013.00328
- Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. Retrieved from <http://cran.r-project.org/package=lme4>
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436. doi:10.1163/156856897X00357
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect. *Experimental Psychology*, 58(5), 412–24. doi:10.1027/1618-3169/a000123
- Carreiras, M., Alvarez, C., & de Vega, M. (1993). Syllable frequency and visual word recognition in Spanish. *Journal of Memory and Language*, 32(6), 766–780. doi:10.1006/jmla.1993.1038
- Carreiras, M., & Perea, M. (2004). Naming pseudowords in Spanish: effects of syllable frequency. *Brain and Language*, 90(1-3), 393–400. doi:10.1016/j.bandl.2003.12.003
- Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(4), 857–71. doi:10.1037/0278-7393.23.4.857
- Casaponsa, A., Carreiras, M., & Duñabeitia, J. A. (2014). Discriminating languages in bilingual contexts: the impact of orthographic markedness. *Frontiers in Psychology*, 5(May), 424. doi:10.3389/fpsyg.2014.00424
- 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(1), 204–56.
- Conrad, M., Alvarez, C., Afonso, O., & Jacobs, A. M. (2014). Sublexical modulation of simultaneous language activation in bilingual visual word recognition: The role of syllabic units. *Bilingualism: Language and Cognition*, in press. doi:10.1017/S1366728914000443
- Conrad, M., Carreiras, M., Tamm, S., & Jacobs, A. M. (2009). Syllables and bigrams: orthographic redundancy and syllabic units affect visual word recognition at different processing levels. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 461–79. doi:10.1037/a0013480
- Conrad, M., Grainger, J., & Jacobs, A. M. (2007). Phonology as the source of syllable frequency effects in visual word recognition: evidence from French. *Memory & Cognition*, 35(5), 974–83. doi:10.3758/BF03193470
- Conrad, M., & Jacobs, A. M. (2004). Replicating syllable frequency effects in Spanish in German: One more challenge to computational models of visual word recognition. *Language and Cognitive Processes*, 19(3), 369–390. doi:10.1080/01690960344000224
- Conrad, M., Stenneken, P., & Jacobs, A. M. (2006). Associated or dissociated effects of syllable frequency in lexical decision and naming. *Psychonomic Bulletin & Review*, 13(2), 339–345. doi:10.3758/BF03193854
- Costa, A., La Heij, W., & Navarrete, E. (2006). The dynamics of bilingual lexical access. *Bilingualism: Language and Cognition*, 9(2), 137–151. doi:10.1017/S1366728906002495
- De Groot, A. M. B., Borgwaldt, S., Bos, M., & van den Eijnden, E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language*, 47(1), 91–124. doi:10.1006/jmla.2001.2840
- De Groot, A. M. B., Delmaar, P., & Lupker, S. J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: Support for non-selective access to bilingual memory. *The Quarterly Journal of Experimental Psychology*, 53(2), 397–428. doi:10.1080/713755891
- Deutsch, A., Frost, R., Pollatsek, A., & Rayner, K. (2000). Early morphological effects in word recognition in Hebrew: Evidence from parafoveal preview benefit. *Language and Cognitive Processes*, 15(4/5), 487–506. doi:10.1080/01690960050119670
- 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(03), 341–357. doi:10.1017/S1366728909990575
- 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(03), 175–197. doi:10.1017/S1366728902003012
- Frost, R., Kugler, T., Deutsch, A., & Forster, K. I. (2005). Orthographic structure versus morphological structure: principles of lexical organization in a given language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6), 1293–1326. doi:10.1037/0278-7393.31.6.1293
- Gollan, T. H., & Goldrick, M. (2012). Does bilingualism twist your tongue? *Cognition*, 125(3), 491–7. doi:10.1016/j.cognition.2012.08.002

- Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. In R. J. Harris (Ed.), *Cognitive Processing in Bilinguals* (Vol. 83, pp. 207–220).
- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychological Review*, *103*(3), 518–565.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, *59*(4), 434–446. doi:10.1016/j.jml.2007.11.007
- Jared, D. (2001). Do Bilinguals Activate Phonological Representations in One or Both of Their Languages When Naming Words? *Journal of Memory and Language*, *44*(1), 2–31. doi:10.1006/jmla.2000.2747
- Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language*, *44*(1), 2–31. doi:10.1006/jmla.2000.2747
- Kaushanskaya, M., & Marian, V. (2007). Bilingual language processing and interference in bilinguals: Evidence from eye tracking and picture naming. *Language Learning*, *51*(1), 119–163. doi:10.1111/j.1467-9922.2007.00401.x
- Keuleers, E. (2013). vwr: Useful functions for visual word recognition research. R package version 0.3.0. Retrieved from <http://cran.r-project.org/package=vwr>
- Kleiner, M., Brainard, D. H., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1–1.
- Lemhöfer, K., & Broersma, M. (2011). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. doi:10.3758/s13428-011-0146-0
- Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition*, *32*(4), 533–50. doi:10.3758/BF03195845
- Lemhöfer, K., Dijkstra, T., Schriefers, H. J., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: a megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 12–31. doi:10.1037/0278-7393.34.1.12
- Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods*, *38*(2), 202–10. doi:10.3758/BF03192770
- Libben, M., & Titone, D. (2009). Bilingual lexical access in context: evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 381–390. doi:10.1037/a0014875
- Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS One*, *7*(8), e43230. doi:10.1371/journal.pone.0043230
- Midgley, K. J., Holcomb, P. J., van Heuven, W. J. B., & Grainger, J. (2008). An electrophysiological investigation of cross-language effects of orthographic neighborhood. *Brain Research*, *1246*, 123–35. doi:10.1016/j.brainres.2008.09.078
- Moll, K., & Landerl, K. (2010). *SLRT-II-Verfahren zur Differentialdiagnose von Störungen der Teilkomponenten des Lesens und Schreibens*. Bern: Hans Huber.
- Protopapas, A. (2007). CheckVocal: a program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, *39*(4), 859–62. doi:10.3758/BF03192979
- Schulpen, B., Dijkstra, T., Schriefers, H. J., & Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(6), 1155–78. doi:10.1037/0096-1523.29.6.1155
- Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, *55*(2), 197–212. doi:10.1016/j.jml.2006.03.004
- Shook, A., & Marian, V. (2013). The Bilingual language interaction network for comprehension of speech. *Bilingualism: Language and Cognition*, *16*(2), 304–324. doi:10.1017/S1366728912000466
- Thomas, M. S. C., & Allport, A. (2000). Language Switching Costs in Bilingual Visual Word Recognition. *Journal of Memory and Language*, *43*(1), 44–66. doi:10.1006/jmla.1999.2700
- Torgesen, J. K., & Rashotte, C. A. (1999). *TOWRE-2 Test of Word Reading Efficiency*.
- Vaid, J., & Frenck-Mestre, C. (2002). Do orthographic cues aid language recognition? A laterality study with French-English bilinguals. *Brain and Language*, *82*(1), 47–53. doi:10.1016/S0093-934X(02)00008-1
- Van Kesteren, R., Dijkstra, T., & de Smedt, K. (2012). Markedness effects in Norwegian – English bilinguals: Task-dependent use of language-specific letters and bigrams. *The Quarterly Journal of Experimental Psychology*, *65*(11), 2129–54. doi:10.1080/17470218.2012.679946
- Westbury, C., & Buchanan, L. (2002). The Probability of the Least Likely Non-Length-Controlled Bigram Affects Lexical Decision Reaction Times. *Brain and Language*, *81*(1-3), 66–78. doi:10.1006/brln.2001.2507
- Yarkoni, T., Balota, D. A., & Yap, M. J. (2008). Moving beyond Coltheart's N: a new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–9. doi:10.3758/PBR.15.5.971