

# The frequency effect in second-language visual word recognition

WOUTER DUYCK

*Ghent University, Ghent, Belgium*

DIETER VANDERELST

*University of Antwerp, Antwerp, Belgium*

AND

TIMOTHY DESMET AND ROBERT J. HARTSUIKER

*Ghent University, Ghent, Belgium*

A lexical decision experiment with Dutch–English bilinguals compared the effect of word frequency on visual word recognition in the first language with that in the second language. Bilinguals showed a considerably larger frequency effect in their second language, even though corpus frequency was matched across languages. Experiment 2 tested monolingual, native speakers of English on the English materials from Experiment 1. This yielded a frequency effect comparable to that of the bilinguals in Dutch (their L1). These results constrain the way in which existing models of word recognition can be extended to unbalanced bilingualism. In particular, the results are compatible with a theory by which the frequency effect originates from implicit learning. They are also compatible with models that attribute frequency effects to serial search in frequency-ordered bins (Murray & Forster, 2004), if these models are extended with the assumption that scanning speed is language dependent, or that bins are not language specific.

Probably the most controlled variable in the literature on word recognition and production is word frequency. Participants respond faster to high-frequency (HF) words than to low-frequency (LF) words in almost any lexical processing task, including lexical decision, reading aloud, semantic categorization, and picture naming (e.g., Carroll & White, 1973; Forster & Chambers, 1973; Whaley, 1978).

Although there is little doubt that the word frequency effect (FE) reflects an important property of the organization of the mental lexicon, there is debate about the specific locus of this effect. Some models consider the FE a result of implicit learning. In this view, lexical representations are strengthened by repeated exposure—for example, by lowering a recognition threshold (e.g., the logogen model of Morton, 1970). HF words are recognized faster than LF words are, because their lexical representations reach and surpass the threshold faster. A similar mechanism is also responsible for FEs in localized and distributed connectionist models of visual word recognition (e.g., McClelland & Rumelhart, 1981). In contrast, the influential “rank hypothesis” proposed by Murray and Forster (2004) assumes that the lexicon is organized into frequency-ordered bins that are searched serially during visual word recognition. According to that account, HF words are recognized faster than LF words are, because the search process considers

them earlier. Contrasting with the learning account, FEs should thus stay the same as long as the relative frequency does not change, even if the absolute number of exposures increases.

Surprisingly, no study has ever explicitly focused on FEs in visual word recognition in a second language (L2). This offers a complex but interesting case for these accounts of the FE, because words in L2 are encountered less often than L1 words, even if these words have the same “objective” frequency—namely, the frequency determined from “monolingual” language corpora—and even though the relative frequency ranking within L2 is probably very similar to that in L1. In this view, it is a challenging open question whether the L2 FE is different from the L1 FE. Therefore, it is the aim of the present study to directly compare L1 and L2 FEs in word recognition by bilinguals, and to explore the resulting constraints for future modeling of (bilingual) word recognition.

Interestingly, in the production domain, two studies have recently adopted the same approach. Gollan, Montoya, Cera, and Sandoval (2008) found that English-dominant Spanish–English bilinguals showed a larger FE in picture naming in the nondominant language than they did in the dominant language. Gollan et al. (2008) interpreted these results within a version of the rank model adopted for word

---

W. Duyck, wouter.duyck@ugent.be

---

production. By contrast, comparing Catalan–Spanish with Spanish–Catalan bilinguals in Spanish picture naming, Ivanova and Costa (2008) found no larger FE when bilinguals were naming in their nondominant language than in their dominant language. Ivanova and Costa argued that the larger L2 FE of Gollan et al. (2008) could be attributed to the fact that the nondominant language of Gollan et al.'s (2008) bilinguals was still the first-acquired, yielding larger age-of-acquisition (AoA) effects.

These mixed findings from the production domain do not allow clear predictions about the relative size of L1 and L2 FEs in word recognition by bilinguals. At present, the only available indirect evidence for word recognition comes from a modeling study by Dijkstra, Van Heuven, and Grainger (1998) with their influential bilingual interactive activation model [BIA(+); see also Dijkstra & Van Heuven, 2002]. They argued that BIA simulations fitted the observed bilingual word recognition effects better if “a reduced frequency range for L2 words” (p. 193) was used.<sup>1</sup> More specifically, frequency was transformed onto resting-level activations that had a lower maximum for L2 words than for L1 (Walter van Heuven, personal communication); L2 words, therefore, generally had lower resting-level activations than did L1 words of the same corpus frequency. Therefore, BIA would predict a different FE for L2 than for L1 words. However, BIA's L2 frequency assumptions and resulting predictions are based on theoretical considerations and intuition, but not on empirical evidence. In this view, the present study contributes not only to monolingual accounts of the FE, but also to future modeling developments in bilingual word recognition.

In Experiment 1, we explicitly compared FEs in L1 and L2 lexical decision of unbalanced Dutch–English bilinguals, using words that are matched on objective corpus frequency. Experiment 2 replicated the English condition with monolingual speakers of English (similar to Gollan et al., 2008), to ensure that any differential effects in Experiment 1 for L1 and L2 could not be confounded with some inherent property of the English stimuli.

## EXPERIMENT 1

### Method

**Participants.** Eighteen first-year Ghent University students participated for course credit. All reported Dutch and English as, respectively, their L1 and L2. They received formal instruction in L2 for at least 5 years, and were regularly exposed to L2 through popular media (music, television, etc.), textbooks, and so on.

**Materials.** The stimuli (see the Appendix) were 50 Dutch words and 50 English words, selected from the CELEX Lexical Database (Baayen, Piepenbrock, & van Rijn, 1993), using the WordGen stimulus generation software (Duyck, Desmet, Verbeke, & Brysbaert, 2004) and meeting the restrictions below. For each language, there were 25 LF and 25 HF words. As can be seen in Table 1, log frequency per million occurrences differed significantly between conditions within languages ( $ps < .001$ ). Across languages, frequency was matched ( $ps > .78$ ). Also, L1 and L2 LF and HF words were matched for three additional variables influencing lexical decision times: number of letters, neighborhood size, and summed bigram frequency, all calculated using WordGen. Interlingual homographs and identical cognates were excluded from the stimuli.

**Table 1**  
Properties of the Target Stimuli  
As a Function of Language and Frequency

Language	Frequency	LogFreqMln <sup>a</sup>	Length <sup>b</sup>	BigrFreq <sup>c</sup>	<i>N</i> <sup>d</sup>
Dutch	Low	0.82	6.24	12,841	1.52
	High	2.22	6.24	12,782	1.72
English	Low	0.78	6.28	12,942	1.20
	High	2.22	6.24	12,645	1.36

<sup>a</sup>Mean log frequency per million words, according to the CELEX lemma database (Baayen, Piepenbrock, & van Rijn, 1993). <sup>b</sup>Mean word length (number of letters). <sup>c</sup>Mean summated bigram frequency, normalized for corpus size (calculated using WordGen; Duyck, Desmet, Verbeke, & Brysbaert, 2004). <sup>d</sup>Mean neighborhood size (the number of words in each language that differ only by a single letter; WordGen, Duyck et al., 2004).

Using WordGen (Dutch) and the ARC nonword database (English) (Rastle, Harrington, & Coltheart, 2002), we also constructed 50 nonwords for each language. These were as wordlike as possible, to avoid lexical decision responses based on low-level orthographic features of the letter strings. Nonwords did not contain illegal bigrams, respected the phonotactic rules of the target language, and had summed bigram frequencies equal to those of word targets.

**Procedure and Apparatus.** Participants received written instructions to perform a language-specific Dutch (L1) and English (L2) lexical decision task. Order of tasks was counterbalanced across participants. Each task consisted of 100 trials, preceded by 20 practice trials. After the presentation of a fixation cross (1,000 msec), a random HF/LF word or nonword target was presented and remained on the screen until a response was made. Participants were instructed to respond as accurately and as quickly as possible. In order to prevent any language switching effects, there was a break between the two tasks.

### Results

Only reaction times (RTs) of correct word responses were included in the RT analyses. RTs that deviated more than 2.5 standard deviations from the respective participants' mean RTs were excluded from analyses (2.99% of datapoints). Mean RTs and accuracy across participants as a function of language (L1 vs. L2) and frequency (high vs. low) are displayed in Table 2.

RTs and accuracy (% correct) were analyzed using ANOVAs with participants ( $F_1$ ) and items ( $F_2$ ) as random variables. Language and frequency were the independent (within-participants and between-items) variables. As expected, responses were faster for L1 words than for L2 words [ $F_1(1,17) = 61.92, p < .001$ ;  $F_2(1,96) = 64.66, p < .001$ ]. Also, participants responded more quickly to HF words than to LF words [ $F_1(1,17) = 36.42, p < .001$ ;  $F_2(1,96) = 37.77, p < .001$ ]. Planned comparisons

**Table 2**  
Mean Reaction Times (RTs), Standard Errors, and Accuracy (%) in the Lexical Decision Task As a Function of Language and Frequency (Experiment 1: Dutch–English Bilinguals)

Word Frequency	Dutch (L1)				English (L2)			
	RT		Acc.		RT		Acc.	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Low	573	12.6	94	1.4	702	25.4	86	1.9
High	527	12.9	98	0.7	599	13.2	98	0.9
Effect	46		4		103		12	

showed that this FE was significant for both L1 (Dutch) [ $F_1(1,17) = 24.26, p < .001; F_2(1,96) = 7.55, p < .01$ ] and L2 (English) [ $F_1(1,17) = 26.10, p < .001; F_2(1,96) = 35.33, p < .001$ ]. Crucially, there was an interaction between language and frequency: The FE was significantly larger for L2 (103 msec) than for L1 (46 msec) [ $F_1(1,17) = 8.53, p < .01; F_2(1,96) = 5.11, p < .05$ ].<sup>2</sup>

Participants made fewer errors in L1 than in L2 [ $F_1(1,17) = 14.19, p < .01; F_2(1,96) = 4.60, p < .05$ ]. Also, there was an FE (fewer errors for HF words) [ $F_1(1,17) = 43.70, p < .001; F_2(1,96) = 21.29, p < .001$ ] that was significant for L1 (only by participants) [ $F_1(1,17) = 12.69, p < .01; F_2(1,96) = 3.38, p = .07$ ] and L2 [ $F_1(1,17) = 50.65, p < .001; F_2(1,96) = 21.97, p < .001$ ]. Similar to RTs, the accuracy FE was larger for L2 (12%) than for L1 (4%) [ $F_1(1,17) = 19.06, p < .001; F_2(1,96) = 4.06, p < .05$ ].

## EXPERIMENT 2

### Method

**Participants.** Twenty students from Michigan State University participated in exchange for course credit. They were monolingual, native speakers of English.

**Materials, Procedure, and Apparatus.** These were identical to Experiment 1, but only the English (L1) lexical decision task was administered.

### Results

The same outlier criterion was used as in Experiment 1 (3.20% of datapoints). Mean RTs and accuracy across participants as a function of frequency (high vs. low) are displayed in Table 3.

Again, RTs and accuracy were analyzed using ANOVAs with participants ( $F_1$ ) and items ( $F_2$ ) as random variables. Frequency was the independent (within-participants and between-items) variable. There was a 52-msec FE on RTs [ $F_1(1,19) = 44.48, p < .001; F_2(1,48) = 15.84, p < .001$ ] and a 5% effect on accuracy [ $F_1(1,19) = 25.86, p < .001; F_2(1,48) = 6.82, p < .05$ ].

### Comparison of Experiments 1 and 2

To further compare the FEs in L2 and L1, we analyzed the data for the (same) English word stimuli when English was the L2 of bilinguals (Experiment 1) and when English was the L1 of monolinguals (Experiment 2), with group as an additional independent variable. The monolinguals responded more quickly than the bilinguals did [ $F_1(1,36) = 12.80, p < .01; F_2(1,48) = 120.62, p < .001$ ].

Importantly, the FE was significantly larger for the bilinguals (103 msec) than for the monolinguals (52 msec) [ $F_1(1,36) = 5.91, p < .05; F_2(1,48) = 7.77, p < .01$ ] (see note 2). The error analyses also showed an effect of group [ $F_1(1,36) = 9.33, p < .01; F_2(1,48) = 10.72, p < .01$ ] and a group  $\times$  frequency interaction [ $F_1(1,36) = 9.09, p < .01; F_2(1,48) = 4.79, p < .05$ ].

Finally, we compared the English data of English monolinguals (Experiment 2) with the Dutch data of Dutch–English bilinguals (Experiment 1). There was no group difference between experiments ( $F_s < 1$ ). More importantly, the FE in the Dutch (L1) lexical decision task of Experiment 1 did not differ from that of English (L1) in Experiment 2 ( $F_s < 1$ ). The same holds for the accuracy data (all  $F_s < 1$ ).

## DISCUSSION

This is the first study that has directly compared the size of L1 and L2 FEs in visual word recognition by bilinguals, using words that were matched on objective corpus frequency. We observed that the FE in L2 reading by bilinguals was about twice as large as that in L1 reading, in the specific frequency range that we tested. A second experiment showed that monolingual English participants displayed an FE in English comparable to that of the bilingual participants in Dutch (their L1), confirming that the crucial language  $\times$  frequency interaction could not be due to some confounding variable inherent to the English stimuli.

These findings support models of lexical access that incorporate the FE as the result of an asymptotic learning process. In such an account (e.g., McCusker, 1977), with each word occurrence, progressively smaller learning takes place, resulting in a nonlinear, logarithmic relation between frequency and RT. Because Dutch–English bilinguals encounter English (L2) words much less often than their Dutch (L1) counterparts do, recognition of L2 words will be like the recognition of L1 words of lower objective frequency. And, more importantly, because the slope of the logarithmic learning function is steeper at the lower end of the scale (where L2 words are overrepresented), a corpus frequency manipulation in L2 will yield a larger effect on RTs than will a similar corpus frequency manipulation in L1, which is consistent with finding a larger L2 FE.

The present findings also offer some constraints for possible extensions of Murray and Forster's (2004) rank model to bilingual lexicons. If one assumes that frequency-ranked bins contain lexical entries only from a single language, it would be reasonable to assume that the lexical entries in the L1 and L2 bins are ranked in approximately the same order. For instance, even though a Dutch–English bilingual has encountered the HF L2 word *dog* on fewer occasions than its L1 equivalent *hond*, *dog* will occupy a rank in the set of L2 words similar to that occupied by *hond* in the set of L1 words. Because frequency ranks of translation equivalents' lexical entries are similar across languages, FEs, originating from rank differences during lexical search, should also be similar;

Table 3

Mean Reaction Times (RTs), Standard Errors, and Accuracy (%) in the Lexical Decision Task As a Function of Language and Frequency (Experiment 1: Dutch–English Bilinguals)

Word Frequency	English (L1)			
	RT		Acc.	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Low	584	18.5	94	1.3
High	532	20.0	99	0.4
Effect	52		5	

that is, the difference in lexical search time through the L1 bin between HF and LF words should be comparable to the difference in search time through the L2 bin between HF and LF words of similar corpus frequencies. A model with language-selective bins would therefore predict equally large L1 and L2 FEs. As such, the rank model with language-specific bins would be able to account for a larger FE in L2 only if the additional but plausible assumption is made that search speed is different within L1 and L2 bins.

Another possible extension of the rank model could assume that bins might contain lexical entries from both languages, consistent with the evidence against language-specific lexical access in bilinguals (e.g., Dijkstra & Van Heuven, 2002; Duyck, 2005). These shared bins could then be organized following the same principle: serially ranked according to the relative frequency of occurrence of the lexical entries. Because L2 words are encountered less often than L1 words are, they would then generally occupy relatively lower ranks than would L1 words with the same objective frequencies in corpora. Because the number of words in a given frequency band is larger as frequency decreases,<sup>3</sup> the processor would have to consider more words before arriving at the correct position in the bin, and search time would increase nonlinearly. Hence, the same corpus frequency manipulation would have more impact on lexical search RTs in L2 than in L1, consistent with the present findings. This interpretation is consistent with the observation that lexical access is especially delayed for the LF L2 words (Table 2). Interestingly, one could also assume shared bins for same-alphabet languages but language-selective lexical access, and thus language-selective bins, for different-alphabet orthographies (e.g., Gollan, Forster, & Frost, 1997). Within such a model, one would expect a larger L2 FE in Dutch–English bilinguals, but not in Chinese–English bilinguals, for example, unless one also assumes different scanning speeds, as in the interpretation above. Finally, note that all rank model extensions above would probably predict equally large FEs in balanced bilinguals, because then both scanning speed and absolute frequency ranks should be similar for L1 and L2, assuming similar exposure to both languages.

As noted in the introduction, the present study also provides valuable input for computational models of bilingual visual word recognition. 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. Hence, the present study provides empirical support for the theoretical and intuitive considerations on which this BIA(+) assumption is based. The present findings also have methodological implications for bilingual studies. In general, studies on bilingual word processing that match stimuli across languages use monolingual frequency corpora to do so. The present study demonstrates that this practice may lead to a selection of L2 stimuli that is consistently too low-frequent in comparison with the L1 stimuli. In this view, it is very important to dissociate the fundamental

differences between L1 and L2 processing and effects of confounded subjective frequency.

This recognition study supplements interesting recent developments in word production. As in the present study, Gollan et al. (2008) found that English-dominant Spanish–English bilinguals showed a larger FE in picture naming in the nondominant language than in the dominant language. These convergent findings from recognition and production would be consistent with the notion of amodal lexical representations, as proposed, for instance, in the production model of Dell (1990). In this view, future bilingual research may contribute to theoretical debates in the monolingual literature. Of course, the present study cannot rule out the alternative possibility that the larger L2 FEs for both recognition and production still originate from modality-specific lexical representations that respond to frequency-driven learning (and therefore, also to experimental frequency manipulations) in the same way. Because asymptotic learning, one possible explanation of the FE and its interaction with language, has been reported in many domains (e.g., motor learning), it might also have the same effects in both recognition and production. Another complication missing from these interpretations is the observation that Ivanova and Costa (2008), comparing Spanish picture naming between Spanish–Catalan and Catalan–Spanish bilinguals, did not observe a larger FE in L2 than in L1, unlike Gollan et al. (2008). According to Ivanova and Costa, Gollan et al.'s (2008) larger L2 FE could be due to the fact that the nondominant language of Gollan et al.'s (2008) participants was still the first-acquired language. Ivanova and Costa argued that the first-acquired language is typically learned during a longer time interval than is the second-acquired language (in which most words are often learned during a timespan of a few years), and should therefore yield larger AoA effects. As such, a larger FE in an L2 that was first-acquired may partly be an AoA effect. This criticism does not apply to the present study. As in Ivanova and Costa, our bilinguals acquired L2 later, but we still obtained a larger FE in L2 than in L1.<sup>4</sup>

Finally, it is also interesting to note that our study differs somewhat from these production studies when comparing FEs in L1 between monolinguals and bilinguals. Our English monolinguals showed an L1 FE (in English) similar to that for our Dutch–English bilinguals (in Dutch).<sup>5</sup> This contrasts with Gollan et al. (2008), who found that bilinguals showed a larger frequency effect in their dominant language than monolinguals did. Similarly, Ivanova and Costa (2008) reported that Spanish–Catalan bilinguals yielded an L1 FE 16 msec larger than that of Spanish monolinguals. However, note that this FE difference between monolinguals and bilinguals was only 7 msec (Ivanova & Costa, 2008, Table 3) and was probably not significant if only a single stimulus presentation was considered, as in the present procedure. Because this is currently the only available recognition study on this issue, future research will be needed to disentangle this complex pattern of findings across production and recognition studies.

In conclusion, our results clearly show that the word FE in lexical decision by bilinguals is larger in L2 than in L1.

These findings offer important input for future developments in (bilingual) visual word recognition modeling.

#### AUTHOR NOTE

This research was supported by the Research Foundation Flanders, where W.D. is a postdoctoral research fellow. We thank Fernanda Ferreira for allowing us to run Experiment 2 in her lab. We are also grateful to Ton Dijkstra, Tamar Gollan, and Walter van Heuven for discussions on this article, and clarifications regarding BIA's implementation. Correspondence concerning this article should be addressed to W. Duyck, Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium (e-mail: wouter.duyck@ugent.be).

#### REFERENCES

- BAAYEN, R. H., PIEPENBROCK, R., & VAN RIJN, H. (1993). The CELEX Lexical Database (Release 1) [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.
- CARROLL, J. B., & WHITE, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, *25*, 85-95.
- DELL, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language & Cognitive Processes*, *5*, 313-349.
- DIJKSTRA, T., & VAN HEUVEN, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language & Cognition*, *5*, 175-197.
- DIJKSTRA, T., VAN HEUVEN, W. J. B., & GRAINGER, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica*, *38*, 177-196.
- DUYCK, W. (2005). Translation and associative priming with crosslingual pseudohomophones: Evidence for nonselective phonological activation in bilinguals. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31*, 1340-1359.
- DUYCK, W., DESMET, T., VERBEKE, L. P. C., & BRYLSBAERT, M. (2004). WordGen: A tool of word selection and nonword generation in Dutch, English, German, and French. *Behavior Research Methods, Instruments, & Computers*, *36*, 488-499.
- FORSTER, K. I., & CHAMBERS, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning & Verbal Behavior*, *12*, 627-635.
- GOLLAN, T. H., FORSTER, K. I., & FROST, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *23*, 1122-1139.
- GOLLAN, T. H., MONTOYA, R. I., CERA, C., & SANDOVAL, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory & Language*, *58*, 787-814.
- IVANOVA, I., & COSTA, A. (2008). Does bilingualism hamper lexical access in speech production? *Acta Psychologica*, *127*, 277-288.
- 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.
- MCCUSKER, L. M. (1977, April). *Some determinants of word recognition: Frequency*. Paper presented at the 24th Annual Convention of the Southwestern Psychological Association, Fort Worth, TX.
- MORTON, J. (1970). A functional model of human memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 203-254). New York: Academic Press.
- MURRAY, W. S., & FORSTER, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, *111*, 721-756.
- RASTLE, K., HARRINGTON, J., & COLTHEART, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology*, *55A*, 1339-1362.
- WHALEY, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning & Verbal Behavior*, *17*, 143-154.

#### NOTES

1. Because BIA(+) is basically a bilingual extension of McClelland and Rumelhart's (1981) interactive activation model, it models the FE as asymptotic learning that changes threshold levels or resting activations, consistent with the learning account.
2. Note that this interaction effect between language and frequency remained significant when RTs were transformed to z scores, equating overall performance for L1 and L2.
3. For instance, the CELEX corpus contains 4,924 English lemmata with log frequency per million between 0.5 and 1.0, and only 3,176 entries between 1.0 and 1.5.
4. Note that word-level AoA (as a lexical variable) was not manipulated independently from frequency in the present study, similar to Gollan et al. (2008) and Ivanova and Costa (2008). According to the same rationale of Ivanova and Costa, the L2 of our Dutch-English bilinguals was the second-acquired language, and should therefore yield smaller word-level AoA effects. Hence, finding a larger frequency effect in L2 than in L1 is not likely to be a confounded effect of word-level AoA.
5. Note that this comparison of the FE in L1 between monolinguals and bilinguals concerned stimuli in different languages, which were nevertheless matched on a number of lexical variables across languages.

**APPENDIX**  
**Low-Frequency (LF) and High-Frequency (HF) English and Dutch Stimuli**

English		Dutch	
LF	HF	LF	HF
balcony	afternoon	bedrog [deceit]	begrip [understanding]
birthrate	council	bladzijde [page]	blik [look]
catacomb	couple	bliksem [lightning]	buurt [neighborhood]
ceiling	court	dauw [dew]	duur [duration]
ceremony	daughter	etmaal [day]	functie [function]
crystal	dress	fanfare [brass band]	gebruik [use]
detour	floor	forel [trout]	gedrag [behavior]
deviation	freedom	gasfles [gas cylinder]	geval [case]
diver	garden	gevloek [swearing]	gevolg [consequence]
enquirer	health	gewei [antlers]	grond [land]
fairness	income	illusie [illusion]	hoofd [head]
flame	industry	knop [button]	hoofdstuk [chapter]
font	member	kreeft [lobster]	indruk [impression]
graph	morning	kwadraat [square]	inzicht [insight]
lawn	process	monnik [monk]	koffie [coffee]
liberty	promise	monoloog [monologue]	meisje [girl]
melon	result	ontbijt [breakfast]	mevrouw [madam]
peanut	river	oppas [babysitter]	militair [soldier]
pirate	secretary	rijkdom [wealth]	oorlog [war]
portion	sign	speld [pin]	oorzaak [cause]
puma	teacher	tapijt [carpet]	opzicht [respect]
salary	town	voetstap [footstep]	praktijk [practice]
seaman	truth	weetje [fact]	soldaat [soldier]
stair	union	zwaai [swing]	vrijheid [freedom]
waist	worker	zwijn [pig]	vuur [fire]

Note—English translations of Dutch stimuli are included between brackets.

(Manuscript received October 5, 2007;  
revision accepted for publication March 14, 2008.)