Journal of Memory and Language 93 (2017) 1–21

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading

Padraic Monaghan [a,b,*], Ya-Ning Chang [a], Stephen Welbourne [c], Marc Brysbaert [d]

[a] Lancaster University, UK
[b] Max Planck Institute for Psycholinguistics, Netherlands
[c] University of Manchester, UK
[d] Ghent University, Belgium

ARTICLE INFO

Article history:
Received 14 December 2015
revision received 24 August 2016
Available online 12 September 2016

Keywords:
Reading
Frequency effects
Computational modelling
Individual differences
Bilingualism
Lifespan development

ABSTRACT

Individuals show differences in the extent to which psycholinguistic variables predict their responses for lexical processing tasks. A key variable accounting for much variance in lexical processing is frequency, but the size of the frequency effect has been demonstrated to reduce as a consequence of the individual's vocabulary size. Using a connectionist computational implementation of the triangle model on a large set of English words, where orthographic, phonological, and semantic representations interact during processing, we show that the model demonstrates a reduced frequency effect as a consequence of amount of exposure to the language, a variable that was also a cause of greater vocabulary size in the model. The model was also trained to learn a second language, Dutch, and replicated behavioural observations that increased proficiency in a second language resulted in reduced frequency effects for that language but increased frequency effects in the first language. The model provides a first step to demonstrating causal relations between psycholinguistic variables in a model of individual differences in lexical processing, and the effect of bilingualism on interacting variables within the language processing system.

Introduction

Word frequency is a key variable in predicting differences in word processing efficiency: High frequency words are recognized faster and more accurately than low frequency words (Forster & Chambers, 1973). Measured against a range of other psycholinguistic properties, frequency accounts for a far larger amount of variance in response times and accuracies than other variables. For instance, in one of the earlier "mega-studies" of word processing, Balota, Cortese, Sergent-Marshall, Spieler, and Yap (2004) found that frequency exceeded neighbourhood size and consistency in explaining variance of response times for word naming, and matched the size of the effect of word length. For lexical decision, they found that the standardized regression coefficient for frequency was at least four times as great as that of any other psycholinguistic variable (for other regression analyses demonstrating a similarly greater effect of frequency, see Brysbaert, Stevens, Mandera, & Keuleers, 2016; Brysbaert et al., 2011; Cortese & Khanna, 2007; Keuleers, Stevens, Mandera, & Brysbaert, 2015; Spieler & Balota, 1997; Yap & Balota, 2009). Frequency is taken to indicate greater efficiency of access, more salient representation of the lexical item, and greater availability of the representation within the individual's vocabulary (Adelman, Brown, & Quesada, 2006).

* Corresponding author at: Department of Psychology, Lancaster University, Lancaster LA1 4YF, UK. Fax: +44 1524 593744.
E-mail address: p.monaghan@lancaster.ac.uk (P. Monaghan).

http://dx.doi.org/10.1016/j.jml.2016.08.003
0749-596X/

The frequency effect is typically treated in analyses as a random effect as if variance across participants is random. Hence, until very recently, frequency effects have tended to have been related to mean group responses to individual words, rather than appraised in terms of individuals responding to individual words. However, in the first study on the phenomenon it was already reported that the frequency effect differed between participants who had small and large vocabularies. In a largely overlooked paper, Preston (1935) was the first to examine the word frequency effect. She measured the 'speed of word perception' for familiar and unfamiliar words of the same length. The stimulus words consisted of 50 familiar and 50 unfamiliar six-letter two-syllable words chosen on the basis of Thorndike's (1931) 20,000 Word List. The familiar words were selected from the 1500 highest words of the list (i.e., those used most frequently in printed matter). The unfamiliar words were selected from the 19th and the 20th thousand lowest words. Speed of word perception was "measured by the time between the exposing of a stimulus word and the verbal reading of it" (nowadays called a word naming task). Eighty-one members of elementary psychology classes at the University of Minnesota served as participants. Their average "perception time" for the familiar words was 578 ms; that for the unfamiliar words 691 ms.

A second purpose of Preston's study was "the study of the relation of various measures of reading ability to speed of word perception."[1] The reading ability of the participants was determined by the administration of the Vocabulary Test of the Minnesota Reading Examination, the Chapman Cook Speed of Reading Test, and Test V of the Iowa Silent Reading tests. The first test contained 100 words with five possible definitions from which examinees had to select the correct definition. In the Chapman Cook Speed of Reading Test participants were presented with 25 short paragraphs in which one word spoiled (sic) the paragraphs. Participants had to find as many intruder words as possible in 2.5 min and cross out these words. Test V of the Iowa Silent Reading tests was a paragraph comprehension test, in which 12 paragraphs had to be read and 3 questions answered per paragraph. Preston observed significant negative correlations between the language proficiency test scores and the word perception response times, with the highest correlation between vocabulary size and word perception response times, and the lowest correlation between text comprehension and word perception response times. The correlation was higher for the unfamiliar words than the familiar words (e.g., the correlation between vocabulary size and word perception response time was −.508 for the unfamiliar words, and −.412 for the familiar words). In other words, the relation between vocabulary size and response times was greater for low- than high-frequency words, suggesting that individual differences in reading responses may reduce as a consequence of exposure.

Preston's (1935) paper was not mentioned in Howes and Solomon's (1951) article examining the relationship between word frequency and visual duration thresholds in a word identification task. This publication is (erroneously) considered to be the start of word frequency research by many researchers. In two experiments, Howes and Solomon presented evidence that the visual duration threshold in word identification decreased as a function of the logarithm of word frequency (also based on Thorndike's counts). Importantly, and unfortunately, no individual differences were examined and the word frequency effect was presented as a group effect, assumed to be observed to the same degree in all participants. Howes and Solomon's view has dominated the literature, even though occasionally differences in the frequency effect between groups have been investigated (e.g., Chateau & Jared, 2000; Lewellen, Goldinger, Pisoni, & Greene, 1993; Sears, Siakaluk, Chow, & Buchanan, 2008).

Our own interest in individual differences in the word frequency effect arose from a series of experiments published by Yap, Balota, Tse, and Besner (2008).[2] In this article the authors presented data from three different universities on the same lexical decision task. Table 1 gives a summary of the finding that caught our attention. As in Preston's (1935) study, students with a smaller vocabulary size had longer reaction times and, more importantly, showed a larger frequency effect.

The influence of vocabulary size on the frequency effect was later replicated in a large-scale analysis of individual differences in the English Lexicon Project (Yap, Balota, Sibley, & Ratcliff, 2012).

At first sight, it seems surprising that people with a larger vocabulary are more efficient at activating the correct representation than those with a smaller vocabulary, given that they have to select among more candidates in the vocabulary (Lewellen et al., 1993). Still, there are at least four mechanisms that may contribute to the effect. The first is that a larger frequency effect may be a side-effect of longer reaction times (RTs; Faust, Balota, Spieler, & Ferraro, 1999): Comparing the data from Yap et al. (2008) shown in Table 1, 678 ms is 11% longer than 612, and 844 is 15% longer than 732 ms. If we assume that part of the RT to words is not due to word processing but to constant durations such as those involved in stimulus transmission and action planning and performance, it could even be possible that the proportional increase between low and high frequency words is the same across the groups. For the example at hand, this would be the case when the constant time period for stimulus transmission and action is around 438 ms, as then for the lowest vocabulary group the stimulus processing time would be 240 ms [678–438], and 174 ms for the highest vocabulary group, which is 38% different. For the high frequency words, the differences between the highest and lowest vocabulary group would be 406 ms and 294 ms, which is again 38% more. Thus, it is feasible that vocabulary size affects word processing speed generally, rather than affecting the variance associated with the frequency effect.

[1] There was also a third purpose: To determine the test-retest reliability of the speed of word perception measure by asking participants to name the words twice with six days or more in-between. The reliability was .93.

[2] Just like many other researchers, we were until recently unaware of the Preston (1935) paper. We thank Andy Ellis for pointing it out to us.