The word frequency effect in word processing: A review update

Marc Brysbaert [1]   Paweł Mandera [1]   Emmanuel Keuleers [2]

[1] Ghent University, Belgium

[2] Tilburg University, the Netherlands

Keywords: word recognition, word frequency, learning, SUBTLEX

To be published in *Current Directions in Psychological Science*

Address: Marc Brysbaert
Department of Experimental Psychology
Ghent University
Henri Dunantlaan 2
B-9000 Gent
Belgium
Tel. +32 9 264 94 25
Fax. +32 9 264 64 96
E-mal: marc.brysbaert@ugent.be

Abstract

The word frequency effect refers to the observation that high-frequency words are processed more efficiently than low-frequency words. Although the effect was first described over 80 years ago, in recent years it has been investigated in more detail. It has become clear that considerable quality differences exist between frequency estimates and that we need a new standardized frequency measure which does not mislead users. Research also points to consistent individual differences in the word frequency effect, meaning that the effect will be present at different word frequency ranges for people with different degrees of language exposure. Finally, a few ongoing developments are mentioned, which point to the importance of semantic diversity rather than mere differences in the number of times words have been encountered, and to the importance of taking into account word prevalence in addition to word frequency.

When word recognition is analyzed, frequency of occurrence is one of the strongest predictors of processing efficiency. High-frequency words are known to more people and are processed faster than low-frequency words (Monsell, Doyle, & Haggard, 1989). This is true for tasks such as word naming, lexical decision (does the letter string refer to an existing word or not?), and semantic decision (e.g., does the word refer to an animal?). Word frequency is also of importance for memory performance. In this research, participants first study a list of words and are later required to recall the stimuli or to discriminate them from lures (new items). Interestingly, here the pattern of results depends on the task: Low-frequency words in general are more difficult to recall but lead to better performance in a recognition task (Yonelinas, 2002).

Word frequency typically explains some 30-40% of the variance in word recognition tasks (Brysbaert, Stevens, Mandera, & Keuleers, 2016). The effect was first reported by Preston (1935) and has received renewed attention in recent years due to the collection of new, improved word frequency norms and word processing data for thousands of words. The present article is an update of a review written a few years ago (Brysbaert, Buchmeier, Conrad, Jacobs, Bölte, & Böhl, 2011).

Not all word frequency measures are equal

For a long time, researchers did not have much choice about which word frequency measure to use. Because counting words in printed books and newspapers was time-intensive, only one or two lists existed per language (if at all). The situation changed dramatically when texts became available digitally. Then it became much easier to gather a sample of texts (called a corpus) and count the words in them.

Surprisingly, psychologists were not eager to turn to the new word frequency lists. They preferred to continue working with the established and familiar lists (such as Kucera and Francis, 1967, for English), arguably because they did not trust the validity of the new word counts.[1] This situation did not change until frequency lists could be validated against word processing times for thousands of words (collected in so-called megastudies).

The outcome of the validation studies showed that the best word frequency norms are based on language the participants are likely to have been exposed to. This may sound like a truism, but before the validation studies researchers typically used word frequencies based mainly on non-fiction texts, such as newspapers, magazines, and scientific books. When fiction materials were included, they consisted of a limited number of novels and stories. The new frequency measures explain more than 10% extra variance in word recognition performance than the Kucera and Francis (1967) measure.

For undergraduate students (the most commonly used participants in psychology studies), the best word frequency measures turned out to be based on corpora of television subtitles (Brysbaert & New, 2009), social media (Herdağdelen & Marelli, 2017; Gimenes & New, 2016), and blogs (Gimenes & New, 2016). In general, a combination of these sources gives better results than each source alone. For older participants, traditional word frequency measures based on books are sometimes better (Brysbaert & Ellis, 2016). Unpublished research by Johns et al. suggests that further gains may be possible by compiling frequency lists tailored to the participants of a study, depending on their learning histories (e.g., how much television they watched, which book authors they read, how active they have been on social media, which school books they used, etc.).

[1] This is still largely true for memory researchers.

Good frequency lists are based on a corpus that is large enough (not smaller than 20 million words). Such lists also include information about the syntactic roles played by the words (nouns, verbs, …), so that this information can be used in research as well. In addition, good frequency lists should have proven their metal in validation tests based on megastudy data.

A good standardized measure of word frequency

Because frequency counts depend on the size of corpus, researchers typically work with a standardized measure, so that the various counts can be compared. Thus far, the main standardized measure was frequency per million words (fpm). Low-frequency words were typically defined as words with frequencies lower than 5 fpm (e.g., gloom, frenzy, objection); high-frequency words were words with frequencies higher than 100 fpm (such as: energy, market, area).

There are two problems with the fpm measure. First, in corpora with tens of millions of words most words will have a frequency lower than 1 fpm. For instance, in the SUBTLEX-US corpus (Brysbaert & New, 2009), which is based on a corpus of 50 million words, three quarters of the words occur with a frequency below 1 fpm (i.e., 56 thousand of the 74 thousand word types). The percentage becomes even higher for word frequency lists based on larger corpora. Since many of these words are well-known, more than half of the word frequency effect is situated below the intuitive start of the scale (1 fpm).

A second problem with the fpm measure is that the frequency effect is a compressed effect, typically represented by a logarithmic curve. That is, the difference in frequency between 1 fpm and 2 fpm has more or less the same effect on processing times as the difference between 10 fpm and 20 fpm, between 100 fpm and 200 fpm, and between 1000 fpm and 2000 fpm. The compressed nature is even more of a problem for word frequencies below 1 fpm, because it means that the difference in frequency between 1 fpm and 2 fpm has the same effect as the difference between .1 fpm (1 per 10 million words) and .2 fpm (2 per 10 million words) and even between .01 fpm (1 per 100 million words) and .02 fpm if these words are known.

Because the fpm scale provides users with the wrong intuitions (1 fpm is the start of the scale, differences lower than 5 fpm are negligible), van Heuven, Mandera, Keuleers, and Brysbaert (2014) proposed the Zipf scale as an alternative. This scale is logarithmic (like the decibel scale for loudness or the Richter scale for earthquakes) and is calculated as follows: Zipf = log10 (frequency per billion words). In practice, the scale runs from 1 (1 per 100 million words) to 6 (1000 per million words). The lower half of the scale (1-3) represents the low-frequency words, the upper half (4-6) the high-frequency words. There are few words with frequencies higher than 6 Zipf and they are nearly all function words (the, you, but, with, …). Similarly, in a corpus larger than 100 million words, there are words with frequencies below 1 Zipf, but few of these are known.

Figure 1 shows the distribution of Zipf values for the 29,902 words known to more than 75% of the participants in the lexical decision task of the English Lexicon Project, a megastudy with naming and lexical decision times for 40 thousand English words (Balota, Yap, Hutchison, Cortese, Kessler, Loftis, ... & Treiman, 2007).

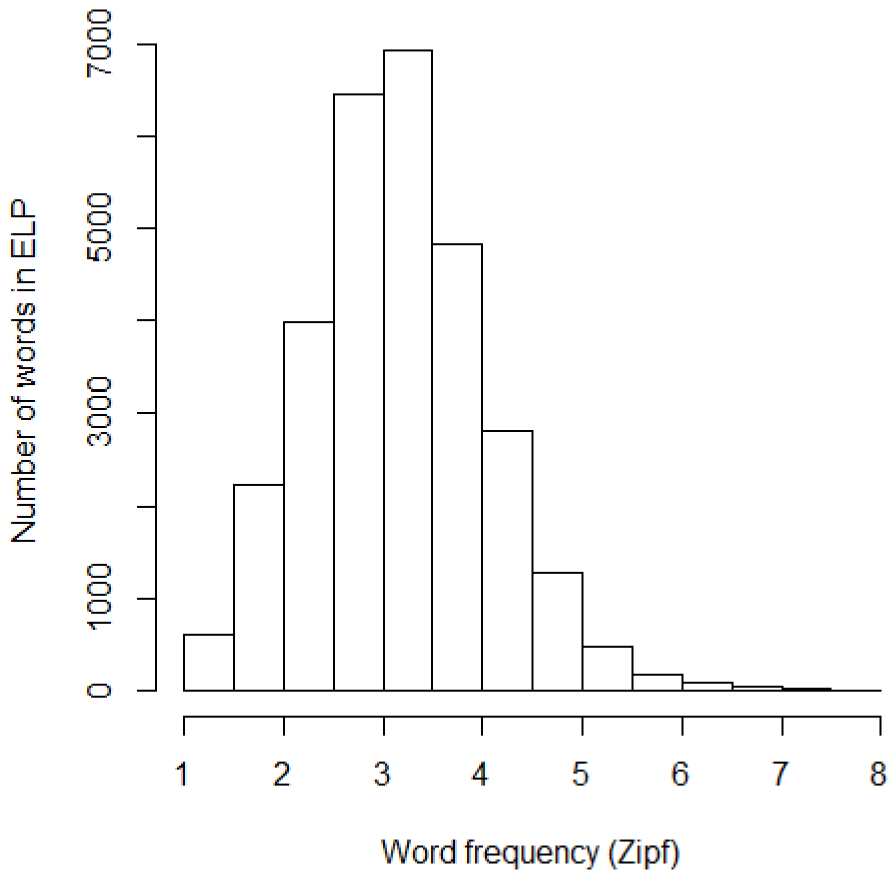

Figure 1: Distribution of English words on the Zipf scale.
This figure shows the number of words with various frequency values from the English Lexicon Project. Data are limited to the words responded to "yes" by more than 75% of the participants taking part in the lexical decision task. Frequencies based on the average of the frequency measures collected by Brysbaert and New (2009) and Gimenes and New (2016), which accounted for more variance in the lexical decision times than the individual measures.

Individual differences in the word frequency effect

Preston (1935) already reported that the frequency effect is larger for university students with a small vocabulary than for students with a large vocabulary. This observation has largely been lost in the time since, but has regained momentum in recent years (Davies, Arnell, Birchenough, Grimmond, Houlson, in press; Kuperman & Van Dyke, 2013; Mandera, 2016), partly after it was discovered that the larger frequency effect is also observed in second language (L2) speakers (Cop, Keuleers, Drieghe, & Duyck, 2015). This suggests that the difference in frequency effect is not related to differences in intelligence (L2 speakers on average are not less intelligent than native speakers), but to differences in language exposure, which can be measured with a vocabulary test (Brysbaert, Lagrou, & Stevens, 2017).

Monaghan, Chang, Welbourne, and Brysbaert (2017) reported that the decrease of the word frequency effect as a function of word exposure can be simulated with a connectionist network. As the network gets more practice with input words, the word frequency effect diminishes. At the same time, simulations showed that a network needs some exposure to words before a word frequency effect emerges (which is understandable, as words need to be encountered a few times before they can show an effect). The net result is that a learning network (and person) at first will show a small frequency effect, which initially grows and then again decreases.

The findings of Monaghan et al. (2017) help us to more deeply understand the individual differences observed in the word frequency effect. At each point in time, individuals have a range of word frequencies to which they show a strong frequency effect, lower-frequency words for which they have a small frequency effect (because they hardly know these words), and high-frequency words for which they have a smaller frequency effect as well (because these words are overlearned).

Figure 2 presents the word frequency effects that can be observed for people with different exposure levels (estimated via vocabulary size), both for accuracy and response times.

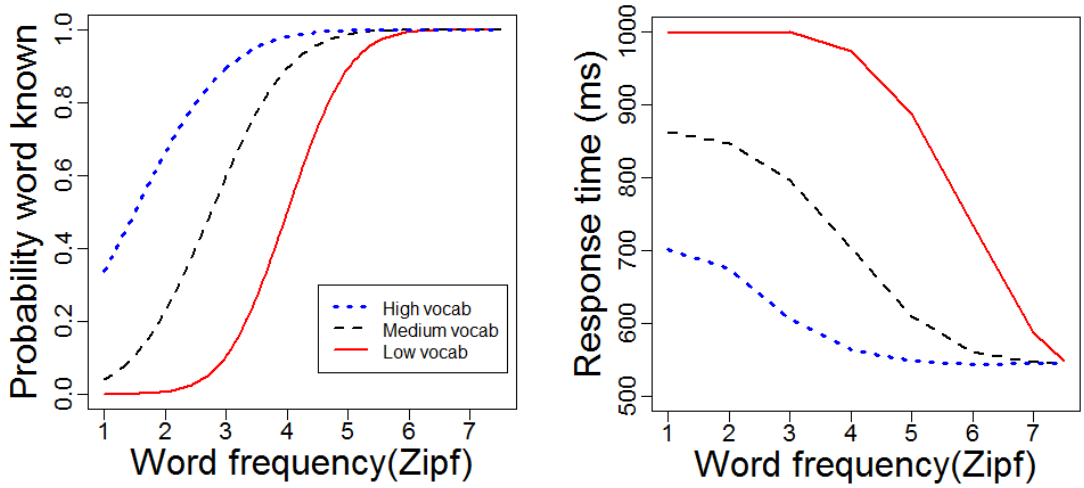

Figure 2: Word frequency curves as a function of word exposure (estimated via vocabulary size)

Left panel: Probability of a word known for a person with a low vocabulary size (a few thousand words of the ELP words shown in Figure 1), for a person with a medium vocabulary size (slightly more than half of the ELP words), and a person with a high vocabulary size (most of the ELP words). Right panel: Estimated response times in a lexical decision task for participants with low, medium, and large vocabulary sizes. Individuals with a low vocabulary size only show a word frequency effect for high-frequency words (Zipf > 4); the other words are not known to them. People with a medium vocabulary size show the largest frequency effect between 3 < Zipf < 5. Finally, persons with a high vocabulary size, show the clearest frequency effect for low frequency words (2 < Zipf < 4). All participants respond equally fast to the very-high-frequency words. Participants with less exposure respond more slowly (and less accurately) to low-frequency words. For the empirical data on which the curves are based, see Keuleers, Diependaele, and Brysbaert (2010) and Brysbaert et al. (2017).

Interpretations of the word frequency effect

The standard interpretation of the word frequency effect is that it is a learning effect. Indeed, the compressed frequency effect has much overlap with the decelerating learning curve observed in repeated tasks.[2] In computational models of word processing, the learning is captured by adapting the activation levels of word representations as a function of their

[2] The compressed nature of the frequency effect can already be observed in the first few times participants read new words (Elgort, Brysbaert, Stevens, & Van Assche, in press)

frequency (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) or by having stronger connections between frequently co-activated representations. The latter follows naturally from the way in which connectionist models learn associations between input patterns and output patterns (Harm & Seidenberg, 2004).

Not everyone agrees with the word frequency effect as a simple learning effect, however. For a start, word frequency is highly correlated with a number of other word features: word length, age at which the word was acquired, similarity to other words. So, in principle the word frequency effect could be confounded with any of these variables (or alternatively, the effect of any of these variables could be a word frequency effect in disguise).

Analyses of the reaction times obtained in megastudies suggest that all variables just listed have an independent effect on word processing (e.g., Brysbaert et al., 2016). For instance, even when the effects of all other variables are partialed out, there still is a robust word frequency effect (although its impact is diminished to some 5-10% of the variance explained). In addition, word frequency interacts with these variables. Typically, the effect of a variable is stronger for low-frequency words than for high-frequency words.

Another possibility is that not so much the number of encounters matter, but the diversity of situations in which the words are encountered. According to this view, words found in many different settings are responded to more efficiently than words present in a small range of settings only. The idea was brought to the forefront by Adelman, Brown, and Quesada (2006), who argued that contextual diversity was a better predictor of word processing efficiency than word frequency. Contextual diversity is defined as the number of texts in which a word appears (rather than the total number of times the word is encountered). A similar idea was defended by Jones, Johns, and Recchia (2012) and Johns, Dye, and Jones (2016). They showed that when novel words are encountered across distinct discourse contexts, people are both faster and more accurate at recognizing them than when the words are seen in redundant contexts. It will be interesting to find out to what extent these alternative measures provide a better account of the frequency effect as outlined in Figure 2.

Finally, a challenge for the word frequency effect is that there are many words in the low-frequency range that are responded to rapidly and accurately. The curves of Figure 2 show the main effects, but not the scatter observed at the low-frequency end. Indeed, some low-frequency words are recognized equally fast and accurately as high-frequency words, making that frequency explains only 30-40% of variance in accuracy and speed data, whereas more than 80% of the variance in megastudies is systematic (and, hence, can be accounted for; see Brysbaert et al., 2016). Several factors are involved.

First, many low-frequency words are related to high-frequency words through inflection, derivation and compounding (e.g. distinctively, microbiologist, reusable, unsweetened, screenshot). Such words can be recognized by decomposing them into their components. Second, some utensils are rarely talked about, even though people are familiar with them (ladle, hinge, sanitizer, …). The frequency of other words may also be misjudged because of the language register tapped into by the corpus (subtitles, texts, social media). Finally, the word frequency effect is built on the idea that each encounter with a word has the same weight. This need not be the case. Some words are much easier to learn than others. Indeed, some words seem to be remembered for the rest of our life upon their first encounter (e.g., a film or a book about a unicorn or a gnome), whereas other words tend to be forgotten easily (kestrel, hangar, cinch). So, the number of encounters itself may not be the best measure of word knowledge.

To counter the shortcomings of word frequency norms, Brysbaert et al. (2016) introduced the variable of word prevalence, defined as the percentage of people knowing the word. This variable explains some 7% of response times extra in lexical decision megastudies in addition to all the known variables, particularly at the low end of the word frequency range. Thus far, the variable has only been collected for the Dutch language. The English language will follow soon.

Recommended reading

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). (See References). A paper on individual differences in the word frequency effect.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). (See References). A review of the variables influencing word processing efficiency and an introduction of the word prevalence variable.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General, 118(1)*, 43-71. Up to a few years ago, the main review of the word frequency effect.

Preston, K. A. (1935). (See References). A seminal study that has been overlooked for decades but that is surprisingly present-day, given its focus on individual differences.

References

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17(9)*, 814-823.

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

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

Brysbaert, M. & Ellis, A.W. (2016). Aphasia and age-of-acquisition: Are early-learned words more resilient? *Aphasiology, 30*, 1240-1263.

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition, 20(3),* 530-548.

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

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 441-458.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. Psychological Review, 108(1), 204-256.

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review, 22(5)*, 1216-1234.

Davies, R. A., Arnell, R., Birchenough, J. M., Grimmond, D., & Houlson, S. (In press). Reading through the life span: Individual differences in psycholinguistic effects*. Journal of Experimental Psychology. Learning, Memory, and Cognition*. DOI: 10.1037/xlm0000366

Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (in press). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition*. DOI: 10.1017/S0272263117000109

Gimenes, M., & New, B. (2016). Worldlex: Twitter and blog word frequencies for 66 languages. *Behavior Research Methods, 48(3)*, 963-972.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. Psychological Review, 111(3), 662-720.

Herdağdelen, A., & Marelli, M. (2017). Social media and language processing: How facebook and twitter provide the best frequency estimates for studying word recognition. *Cognitive Science, 41,* 976-995. DOI: 10.1111/cogs.12392.

Johns, B. T., Dye, M., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review, 23(4),* 1214-1220.

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 66(2)*, 115-124.

Keuleers, E., Diependaele, K. & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology 1:174*. doi: 10.3389/fpsyg.2010.00174.

Kucera, H., & Francis, W. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance, 39(3)*, 802-823.

Mandera, P. (2016). *Psycholinguistics on a large scale: Combining text corpora, megastudies, and distributional semantics to investigate human language processing*. Ghent University: Doctoral dissertation. Available at http://crr.ugent.be/papers/pmandera-disseration-2016.pdf.

Monaghan, P., Chang, Y. N., Welbourne, S., & Brysbaert, M. (2017). Exploring the relations between word frequency, language exposure, and bilingualism in a computational model of reading. *Journal of Memory and Language, 93*, 1-21.

Preston, K. A. (1935). The speed of word perception and its relation to reading ability. *The Journal of General Psychology, 13(1)*, 199-203.

Van Heuven, W.J.B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*, 1176-1190.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory & Language, 46*, 441-517.