# The word frequency effect on second language vocabulary learning

Cesar Koirala[1]

**Abstract**. This study examines several linguistic factors as possible contributors to perceived word difficulty in second language learners in an experimental setting. The investigated factors include: (1) frequency of word usage in the first language, (2) word length, (3) number of syllables in a word, and (4) number of consonant clusters in a word. Word frequency is often treated as the quantifiable correlate of word familiarity, and word length and number of syllables measure structural complexity of a word. Consonant clusters were introduced as the measure of phonetic complexity. A total of 217 native speakers of Spanish and Portuguese were given a vocabulary identification task in which they had to determine whether the words were 1) *Easy to learn,* 2) *Difficult to learn*, or 3) *Unknown*. The findings showed that there is a correlation between English word frequency and perceived word difficulty of the ESL learners. In contrast, there were no clear effects of the other factors on perceived difficulty when the words were controlled for frequency.

**Keywords**: word difficulty, word frequency, word length, syllables, consonant clusters.

## 1. Introduction

Research has shown that some words are relatively harder than others for second language learners to acquire. For instance, it has been reported that English speakers find it difficult to learn Russian words with non-English sound combinations compared to the words with sound combinations that are present in English words (Rodgers, 1969). Hence, estimating the difficulty level of an individual word is important for effective language instruction. In order to do so, it becomes necessary to identify the factors that make words difficult.

1. Voxy, New York, NY, USA; cesar@voxy.com

**How to cite this article**: Koirala, C. (2015). The word frequency effect on second language vocabulary learning. In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds), *Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (pp. 318-323). Dublin: Research-publishing.net. http://dx.doi.org/10.14705/rpnet.2015.000352

First language research has identified several factors that contribute to perceived word difficulty. One such factor is *word frequency*. For instance, the word 'phone' is less difficult than the word 'floccinaucinihilipilification' because we hear 'phone' more frequently than 'floccinaucinihilipilification'. However, frequency is not the only factor that contributes to word difficulty. 'Phone' is the easier of two words also because it is shorter in length. In fact, several variables collectively contribute to L1 word difficulty. In the same vein, research in second language learning has shown the effects of several variables on L2 vocabulary learning. Words can be difficult because of factors like frequency (c.f. Chen & Truscott, 2010), length (Culligan, 2008), abstractness (Higa, 1965), and many others. Most studies have investigated the effect of individual factors on *L2* word difficulty, and there have been very few studies that examined word difficulty in the context of more than one variable at one time (c.f. Alsaif & Milton, 2012).

The contributions of the study undertaken are twofold. First, it is shown that frequency is also a predictor of L2 word difficulty in the case of ESL Spanish and Portuguese speakers. Second, it is shown that within the same frequency band, the other factors have minimal effect on perceived word difficulty. Hence, we examine the relative contribution of the factors rather than individual contribution.

We investigate four variables: (1) *Frequency* of word usage (2) *Word length* in number of characters (3) *Number of syllables* in a word (4) *Number of consonant clusters* in a word. Word frequency is often treated as the quantifiable correlate of *word familiarity*, and word length and number of syllables measure *structural complexity* of a word. In the current study, we introduce consonant clusters as the measure of *phonetic complexity*. Phonetic complexity is a dimension of word difficulty that concerns perception and oral production of the word. Some languages have no (or very few) words with consonant clusters. As a result, speakers of those languages have difficulty perceiving and producing foreign words with consonant clusters. For example, Japanese prohibits consonant clusters, and as a result Japanese speakers report hearing a vowel [u] in words like [ebzo] in between [b] and [z] (Dupoux et al., 1999).

## 2. Experiment design and procedure

A set of 140 words, chosen randomly from a corpus of public domain books from Project Gutenberg (https://www.gutenberg.org/), was divided into four subgroups: words with varying frequencies, words with varying word lengths, words with varying counts of syllables, and words with varying numbers of consonant clusters (Table 1). The words in each subgroup were controlled for other variables, with

equal number of words per condition within each subgroup. The subgroups and conditions are explained in more detail below.

Table 1. Survey subgroups

| **Subgroups** | **Conditions** |
|---|---|
| 1. Varying frequency bands | 1-5, 5-50, 50-500, 500-5000 |
| 2. Varying word length | 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 |
| 3. Varying number of syllables | 1, 2, 3, 4 |
| 4. Varying number of consonant clusters | 0, 1, 2, 3 |

Subgroup 1 consisted of 48 words belonging to four different frequency ranges – 1 to 5, 5 to 50, 50 to 500, and 500 to 5000. There were 12 words in each frequency range, and all the words were of length 5. Subgroup 2 consisted of 36 words of length 3 to 14. There were 3 words in each length condition. All the words were in the frequency range 50-500. Subgroup 3 consisted of 32 words with syllable counts 1 to 4. Like subgroup 2, all 32 words belonged to the frequency range 50-500. Subgroup 4 consisted of 24 words divided equally among the four consonant cluster conditions - 0 clusters, 1 cluster, 2 clusters, and 3 clusters. All the words belonged to the frequency range 50-500. The survey consisting of these 140 words was sent to 217 Spanish and Portuguese ESL learners. Their task was to decide whether a word was 1) *Easy to learn,* 2) *Difficult to learn*, or 3) *Unknown word*.

We used a three-point scale (easy, difficult, and unknown) instead of two (easy and difficult) because we wanted to differentiate words that learners find difficult from the ones that they aren't familiar with. This distinction is especially relevant for subgroup 1 (words with varying frequencies). As mentioned above, word frequency is treated as the quantifiable correlate of word familiarity, and it does not make sense to measure familiarity of unknown words. However, for other measures of complexity (structural and phonetic) we treat unknown words as difficult words and report combined results.

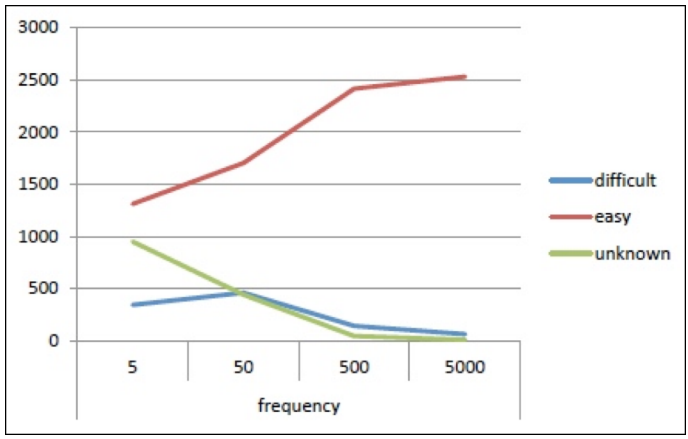
## 3. Results

### 3.1. Frequency

The results showed a negative correlation between word difficulty and word frequency; as frequency increased, difficulty decreased (Figure 1). This is similar

to the relationship between word difficulty and word frequency in the first language. The correlation between word frequency and unknown words is also worth noticing. More words in lower frequency ranges were marked as unknowns than the words in higher frequency ranges.

Figure 1. Effect of frequency




### 3.2. Word length and number of syllables

Unlike the frequency effect, the results did not show a clear trend for varying word length and varying counts of syllables. Most words in these two subgroups were rated as easy by most participants as shown in Figure 2 and Figure 3, respectively.

Figure 2. Effect of word length

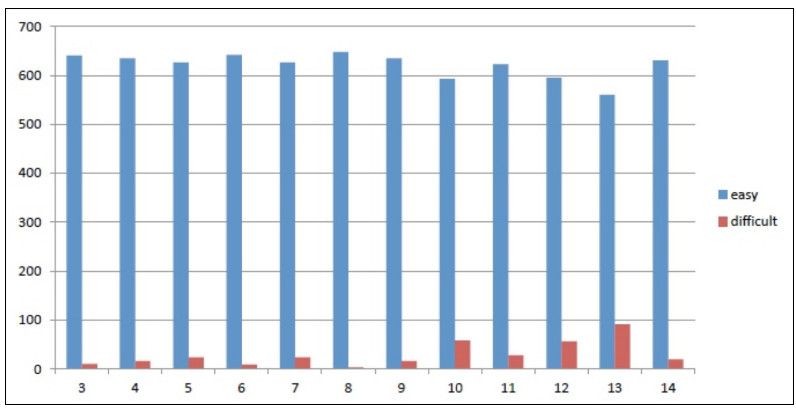
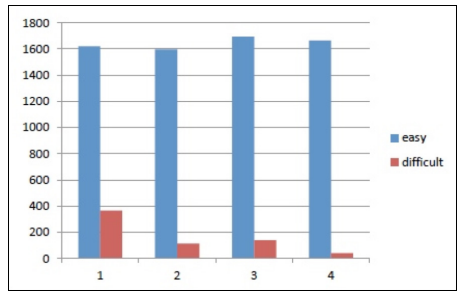


Figure 3. Effect of number of syllables

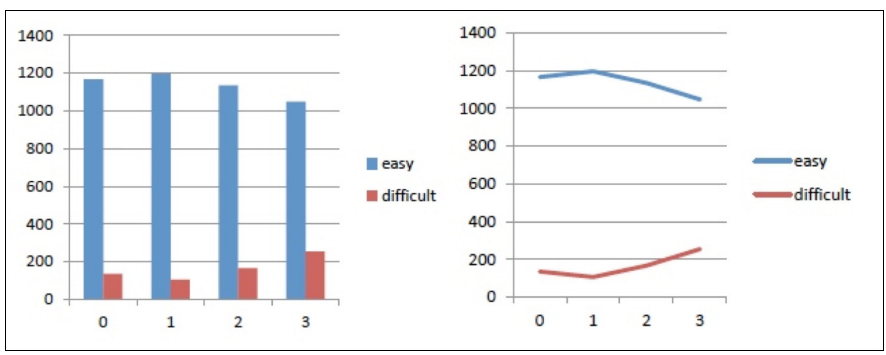



Note that the words in these subgroups were controlled for frequency – they fall in the same frequency range. So, one reason for this result could be that frequency is a better predictor for word difficulty, and as these words fall in the same frequency band, they were rated to be almost equally difficult.

### 3.3. Consonant clusters

Again, word-frequency seems to dominate participants' responses. As the words were in the same frequency band, they were rated similarly irrespective of varying number of consonant clusters. The second graph in Figure 4 shows that difficulty increases with the increase in number of clusters, but the result is not significant.

Figure 4. Consonant clusters




## 4. Discussion and conclusions

Results show a correlation between English word frequency and perceived word difficulty of Spanish and Portuguese speakers. Most participants rated low frequency words to be either difficult to learn or unknown words.

There were no clear results for other factors besides word frequency. It was found that within the same frequency band, the other factors have minimal effect on perceived word difficulty. Most words in the other subgroups were categorized as easy to learn irrespective of their structural or phonetic complexities.

Hence, while examining the relative contribution of the factors on perceived difficulty, word frequency seems to overshadow the effects of other factors.

In order to examine the aforementioned hypothesis, a follow up experiment shall be conducted. In the follow up experiment, the words in subgroups 2, 3 and 4 will be replaced by (1) words in higher frequency range (500-5000), and (2) words in lower frequency range (1-5). Our hypothesis will be supported if most words in 1 are judged easy and most words in 2 are judged difficult irrespective of their structural and phonetic complexities.

These preliminary results could be suggestive to technology-based language instruction platforms. Estimating word frequency as well as controlling for frequency seems essential for effective second language vocabulary learning.

## References

Alsaif, A., & Milton, J. (2012). Vocabulary input from school textbooks as a potential contributor to the small vocabulary uptake gained by English as a foreign language learners in Saudi Arabia. *The Language Learning Journal*, *40*(1), 21-33. doi:10.1080/09571736.2012.658221

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics, 31*(5), 693-713. doi:10.1093/applin/amq031

Culligan, B. (2008). *Estimating word difficulty using yes/no tests in an IRT framework and its application for pedagogical objectives*. Doctoral dissertation. Temple University, Japan.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1568-1578. doi:10.1037/0096-1523.25.6.1568

Higa, M. (1965). The psycholinguistic concept of "difficulty" and the teaching of foreign language vocabulary. *Language Learning*, *15*(3-4), 167-179. doi:10.1111/j.1467-1770.1965.tb00799.x

Rodgers, S. (1969). Measuring vocabulary difficulty: an analysis of item variables in learning Russian-English vocabulary pairs. *International Review of Applied Linguistics in Language Teaching*, *7*, 327-343. doi:10.1515/iral.1969.7.4.327

Published by Research-publishing.net, not-for-profit association
Dublin, Ireland; info@research-publishing.net



Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy
Edited by Francesca Helm, Linda Bradley, Marta Guarda, and Sylvie Thouësny





Typeset by Research-publishing.net
Fonts used are licensed under a SIL Open Font License

ISBN13: 978-1-908416-28-5 (Paperback - Print on demand, black and white)
Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

ISBN13: 978-1-908416-29-2 (Ebook, PDF, colour)
ISBN13: 978-1-908416-30-8 (Ebook, EPUB, colour)

Legal deposit, Ireland: The National Library of Ireland, The Library of Trinity College, The Library of the University of Limerick, The Library of Dublin City University, The Library of NUI Cork, The Library of NUI Maynooth, The Library of University College Dublin, The Library of NUI Galway.

Legal deposit, United Kingdom: The British Library.
British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: décembre 2015.