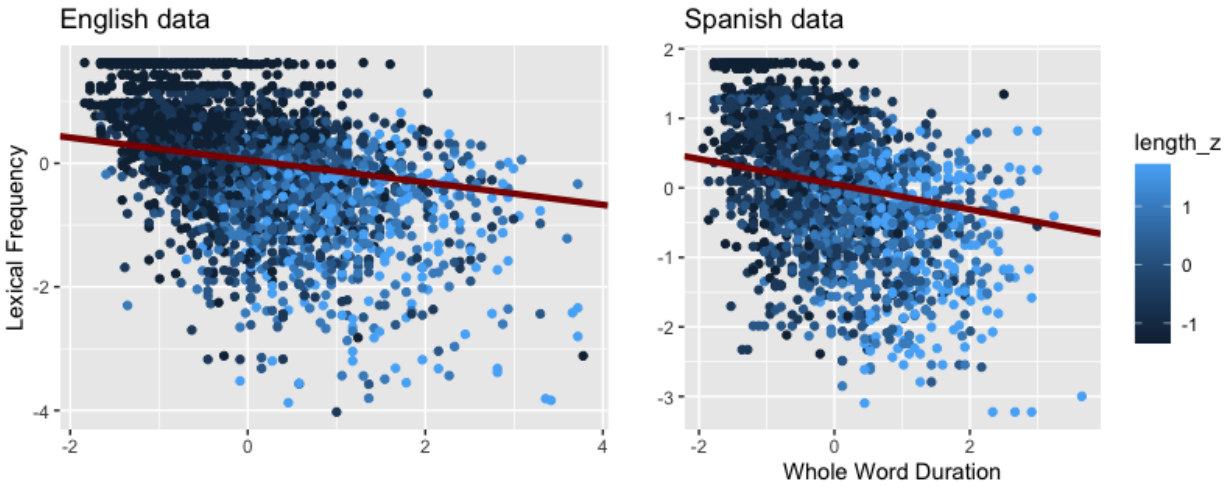
# The effect of lexical frequency on word duration: analyzing corpus data in Spanish and English

The present study investigates the effect of lexical frequency on word duration in Spanish. Previous studies have found that vowel duration in English is influenced by extra-linguistic factors, such as lexical frequency (e.g., Gahl, 2008, 2009; Lohman, 2017). The shortening of frequent forms relative to infrequent ones may correspond to articulatory routinization, which suggests that neuromotor routines become more reduced with repetition (Bybee, 2001; Newmeyer, 2006). However, evidence showing that in homophone pairs (e.g., *thyme – time*), the item with higher frequency is shorter than the infrequent one reveals that frequency effects on duration may not be due to repetition of a phonological form but to lemma frequency effects instead. In the case of Spanish, differences in vowel duration are not as prominent as in English. Although vowel duration in Spanish unstressed syllables tends to be shorter than in stressed syllables (Marín Gálvez, 1994), it is still unclear whether lemma frequency modulates duration in Spanish.

The present study aims to, first, replicate the frequency effect found in vowel duration in English (Gahl, 2008; 2009; Lohman, 2017) but in whole duration. Second, the present study explores the effect of lexical frequency on word duration in Spanish. This study analyzes English corpus data from the *Free ST American English Corpus* and Spanish corpus data from the *Open SLR Corpus*. The English data consisted of 2806 recordings of cellphone conversations, and the Spanish data consisted of 1928 recordings of XXX conversations?. Lexical frequency was measured using the XXX. The data was analyzed using two Bayesian linear regressions with duration as the outcome variable and speech rate and orthographic length as fixed predictors.

The results replicated the frequency effects previously found in English. English frequent words were found to be shorter than infrequent ones when orthographic length and speech rate were controlled for (Figure 1). In Spanish, results exhibited a negative linear relationship between lexical frequency and word duration (Figure 2). Frequent words were shorter than infrequent ones when orthographic length and speech rate were controlled for. The findings have implications for neo-generative (Levelt, 1980), exemplar (Johnson, 1997), and hybrid (Pierrehumbert, 2016) models of sound representation.




\begin{figure}
\caption{Whole word duration in English (Left Panel) and Spanish (Right Panel) as a function of lexical frequency and orthographic length (length_z) with the most plausible line of best fit.} \end{figure}

# References