Overview

The paper I chose investigated the relationship between iconicity, lexical category, and age of acquisition in English and Spanish. It accomplished this by conducting five experiments in which native speakers rated the iconicity of words. These ratings were then regressed (using linear mixed effects models) on lexical category, age of acquisition, and additional features such as frequency and concreteness. The subsequent analysis suggested that in both languages, words learned earlier tended to be more iconic.

In this project, I will reproduce this analysis by using the data included with the paper. This will entail the reproduction of the linear mixed effects models and the recreation of four figures. Finally, I will extend the analysis by investigating the relationship between systematicity and other features in the data, including morpheme count, phoneme count, and “babyness rating.” I discuss these steps below.

Data

The paper includes two datasets: one for English, and one for Spanish. The datasets contain similar variables. Here is a summary of both:

English

- Word: the word being rated
- Subject code: the ID of the subject who supplied the rating (used for random effects)
- Task: the experiment number (1, 2, or 3)
- Iconicity rating: the rating provided by the subject
- Lexical category: the lexical category of the word
- Age of acquisition: the proportion of children who produce the word at 30 months
- Total phonemes: the number of phonemes in the word
- Total morphemes: the number of morphemes in the word
- Frequency: the log frequency of the word
- Systematicity: the systematicity rating of the word
- Babyness: the average “babyness rating” of the word

Spanish

- Word: the word being rated
- Subject code: the ID of the subject who supplied the rating (used for random effects)
- Infinitive form: the infinite form of the word
- Conjugated form: the conjugated form of the word
- Task: the experiment number (4 or 5)
- Iconicity rating: the rating provided by the subject
- English word: the unique english translation of the word, if any
- Lexical category: the lexical category of the word
- Age of acquisition: the proportion of children who produce the word at 30 months
- Total phonemes: the number of phonemes in the word
- Total morphemes: the number of morphemes in the word
- Frequency: the log frequency of the word

Figures and Models

Figure 1

The first figure illustrates the iconicity ratings across the five experiments and their lexical categories. The bars' heights represent the mean ratings and their error bars represent the standard errors.

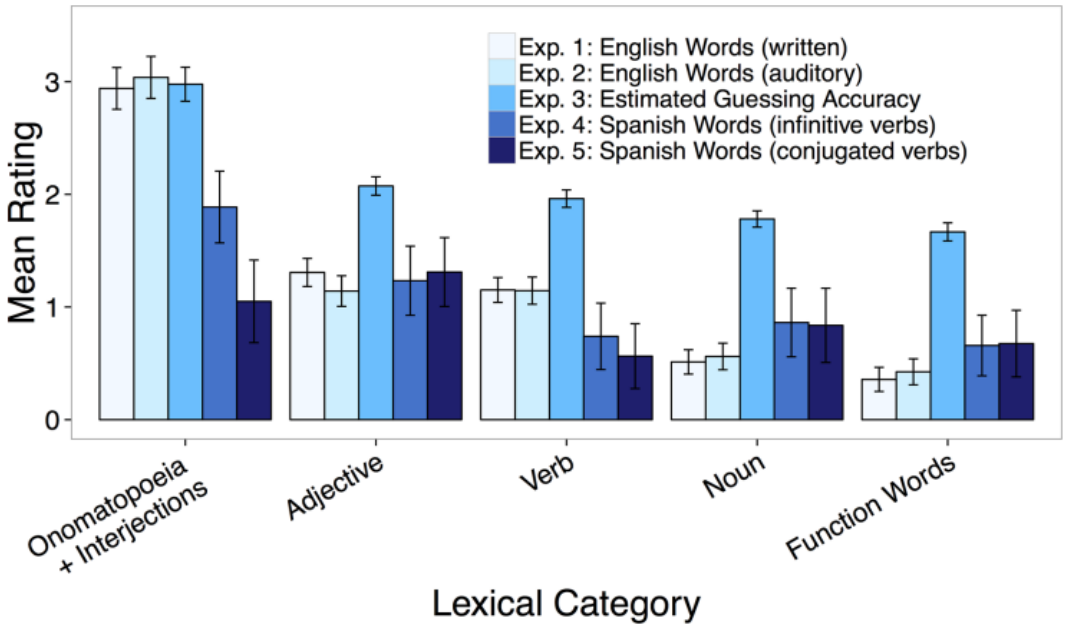


Figure 2

The second figure illustrates the relationship between age of acquisition and iconicity across the experiments. The bars' heights represent the estimated coefficients from linear mixed effects that regress residual iconicity on age of acquisition. The error bars represent a 95% confidence interval. The three groups along the x-axis correspond the three models: one that controls for all covariates excluding systematicity (so frequency, number of phonemes, number of morphemes, concreteness, and association with babies), one that also includes systematicity, and one that also excludes onomatopoeia & interjections. Reproducing this figure will entail computing the residuals produced by controlling for the original features and regressing these residuals on age of acquisition.

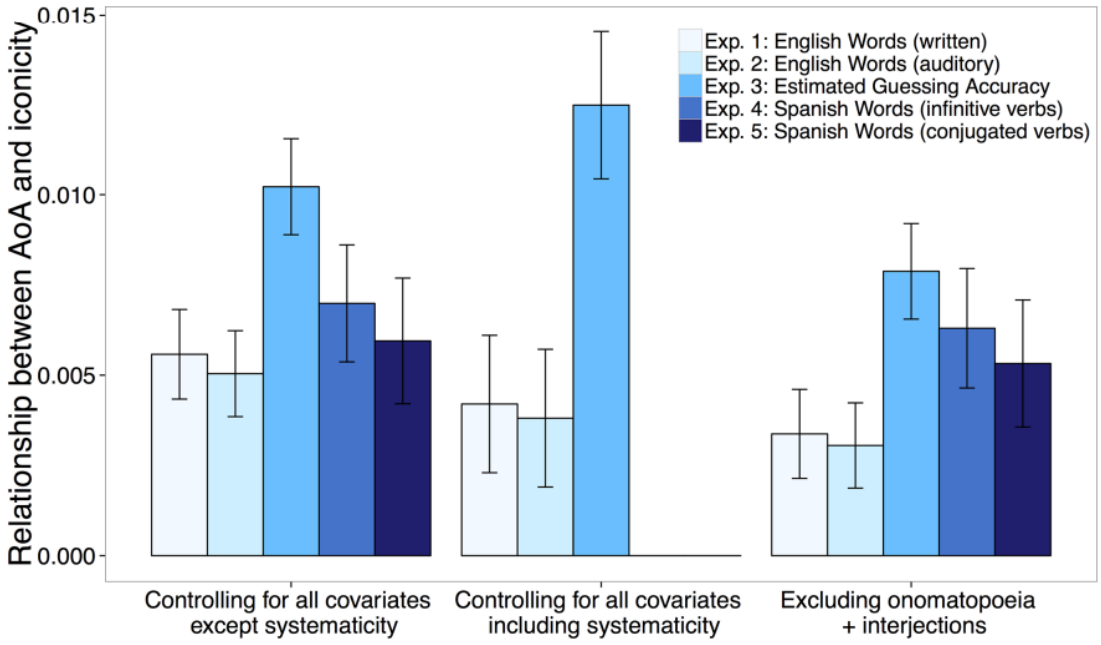


Figure 3

The third figure illustrates the relationship between iconicity and the age of acquisition for 412 english-to-spanish translation pairs. The bars' heights represent the coefficients yielded from regressing iconicity (via multiple regression) on the age of acquisition of each language. The error bars represent a 95% confidence interval for each coefficient.

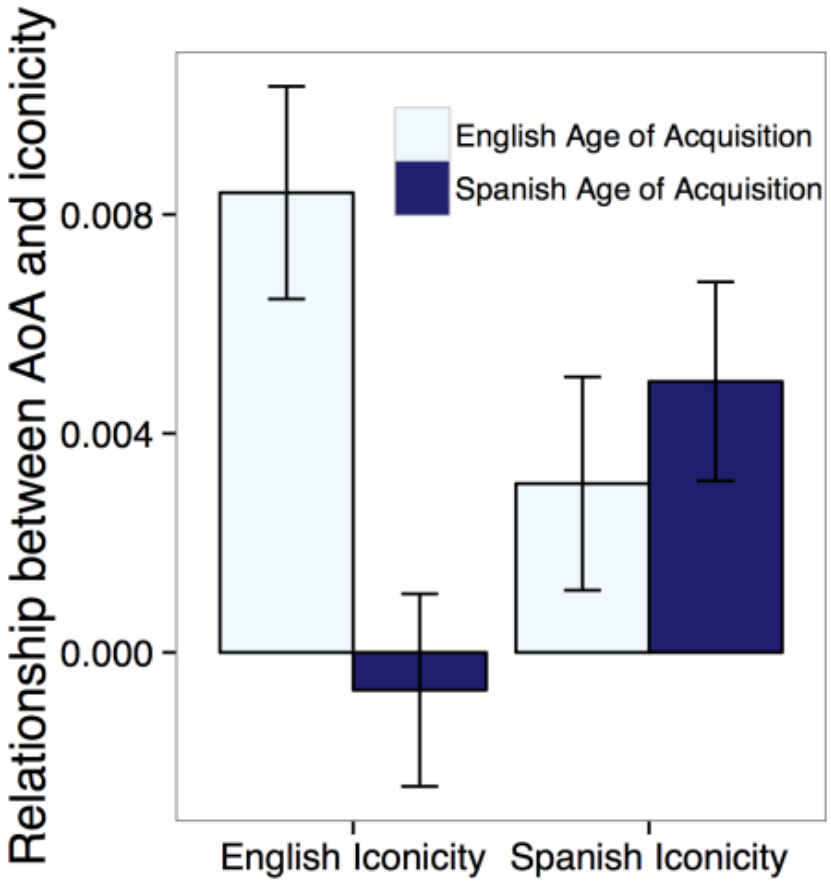
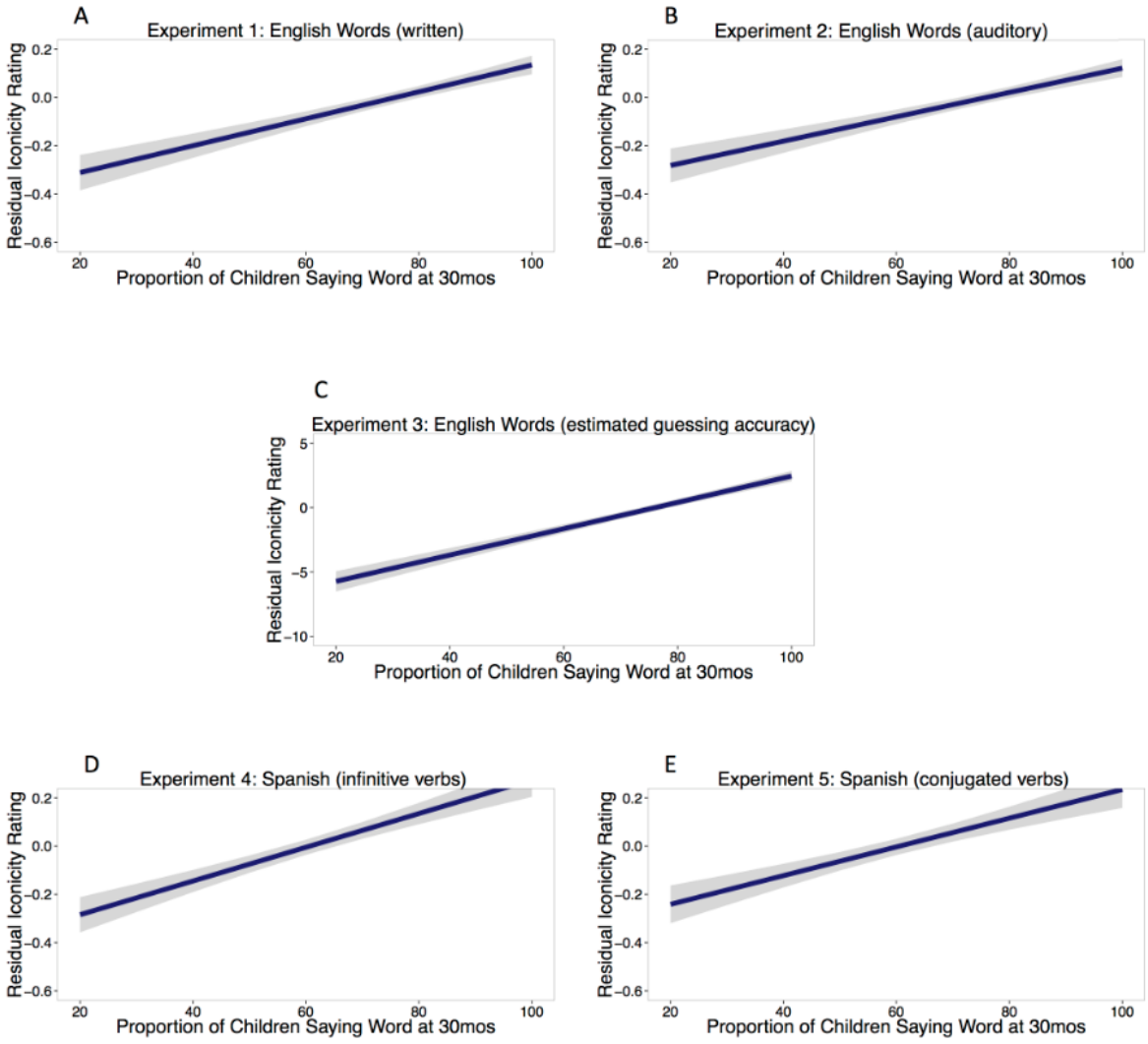


Figure 4

The fourth figure illustrates the models from Figure 2. (I am not sure which group the illustrations correspond to, but I will try each one.) The shaded regions represent a 95% confidence interval.



Extension

After reproducing these figures, I will extend the paper by investigating the relationship between systematicity, lexical category, and age of acquisition. I will do this by swapping the roles of iconicity and systematicity in the above analysis. Hence I will produce linear mixed effects models that regress systematicity on frequency, concreteness, phoneme count, morpheme count, “babyness rating,” iconicity, and age of acquisition. If this does not reveal anything notable, I will search for publicly available data that can better predict the included systematicity ratings.