The role of frequency in the deletion of intervocalic /d/ in Spanish first conjugation past participles

Sonia Barnes
The Ohio State University - barnes.523@osu.edu

**Introduction & Background**

Frequency has been observed to have a fundamental role in determining phonological patterns. A framework (Shannon 1948) that can account for the variation observed in Spanish, using deletion patterns.

**Types of frequency effects**

- **Absolute token frequency**
  - High frequency=models/simply access/Phonetic reduction (Bybee 2000, 2006).

- **Relative frequency**
  - Frequency of an element in a particular sequence
  - Distinct effect from that of absolute frequency in language processing (Hay 2001)

- **Neighborhood density**
  - Role of phonological similarity
  - Low density=Easy access
  - High density neighborhoods each segment contributes more to system entropy.
  - In low density neighborhoods segments are more likely to be redundant

**Intervocalic /d/ deletion in Spanish**

- Deletion is more advanced in the auto morpheme in past participles.
- No effect of token frequency was found within –ado participles.

**The goal of this study is to...**

- Discover whether absolute frequency or other types of frequency have an effect on /d/ deletion in Spanish first conjugation past participles.
- Provide a unified approach of the frequency effects within the information theory framework (Shannon 1948) that can account for the variation observed in Spanish, using the concepts of entropy and surprisal.

**Methodology**

- Corpus de Monterrey (Mexico) - transcription of 117 interviews
- 738 /ado/ participles: 246 /d/ deletion and 492 /d/ retention
- Frequency information: Corpus de Referencia del Espanol Actual (CREA)

**Absolute frequency**
- List of lexical frequencies in the CREA
- Overall use of lexical items in Spanish

**Frequency of construction**
- Searches for a particular construction in CREA
  - “ha estado”, “es que”, “cada vez que”, “alrededor de”,
  - “he cantado”, “este año pasado”, “un año pasado”...

**Neighborhood density**
- Number of lexical neighbors in CREA
- Addition, subtraction or change of one segment
- (absolute frequency – mean frequency of neighbors)
- Effect of frequency over neighborhood density

**Relative frequency in neighborhood**
- A generalized linear model using the independent variables described above was fitted to the data, using the glm function in R.
- The selection of the variables was done following a stepwise procedure.
- Nested models were compared using ANOVA

**Frequency, surprisal and entropy: Predictions**

- **P1** As the absolute frequency and the frequency of construction increase, the odds of deletion will also increase
- **P2** As the number of neighbors decreases, the odds of deletion will increase
- **P3** As relative frequency in the neighborhood increases, the odds of deletion will increase
- **P4** Interaction: as the relative frequency in the neighborhood increases, the effect of neighborhood density decreases

**Results & Discussion**

- The best fit model only includes absolute frequency and the frequency of construction as predictors of intervocalic /d/ deletion in first conjugation past participles.
  - **Prediction 1** was partially confirmed: as expected, when the frequency of a construction increases, the odds of /d/ being deleted also increase.
  - However, the opposite effect was found for absolute frequency: e.g., as frequency increases the rates of deletion decrease. This is an unexpected result.
  - **Predictions 2 – 4** were not confirmed.

- **Figure 1** shows the distribution of the two populations - tokens with /d/ deletion (1) and tokens without it (0) – with respect to frequency of construction.
- The median frequency of the deletion group is higher than the median of the maintenance group.
- The overlap of the notches indicates that the difference in the medians of the two groups is statistically significant.

**Absolute frequency and neighborhood density**

- The effect of absolute frequency stands even if frequency of construction is removed from the model
- A correlation (Figure 2) was found between absolute frequency and neighborhood density, which might explain the unexpected results for absolute frequency.
- As frequency increases, the number of neighbors also increases.
- Low frequency items are also in low density neighborhoods.

Further research is necessary to confirm the trends regarding neighborhood density observed in this study.

**Frequency vs. surprisal**

- When absolute frequency is removed from the model, relative frequency is no longer a significant predictor of deletion.
- Relative frequency alone is not an appropriate measure of surprisal.
- **Alternative analysis:**
  - Frequency of construction + Absolute frequency
  - Probability of occurrence of a particular construction

- The best fit model includes surprisal as the only predictor. The results show that as surprisal decreases the odds of observing deletion increase.

- **Figure 3** shows the distribution of the two populations with respect to surprisal.
- The median surprisal of the deletion group is lower than the maintenance group.
- The lack of overlap of the notches indicates that the difference in the medians of the two groups is statistically significant.

- **Conclusions**
  - This study examined the role that different measures of lexical frequency have on the deletion of intervocalic /d/ in Spanish first conjugation past participles.
  - The results obtained for absolute frequency contradict previous observations in which high frequency items are more likely to be phonetically reduced. This effect was reported by Bybee (2001) for intervocalic /d/ deletion in tokens other than participles in Spanish.
  - The fact that in this study high frequency items are more likely to show deletion can be explained by the positive correlation found between absolute frequency and neighborhood density. The trends observed suggest that the odds of deletion are higher in low density neighborhoods, when each segment in the participle form contributes less to system entropy and is more likely to be redundant.
  - In this study, considering absolute and relative frequency as separate predictors did not provide the best model. The frequency of a construction loses significance as the only predictor and no effects of frequency alone within past participle forms were found in previous studies (Bybee 2001). Instead, I propose a model in which surprisal constitutes the main effect.
  - According to the results obtained, as the surprisal of a particular constructions decreases, the rates of deletion increase. **Low surprisal** forms have greater expectedness and are unstable (Hume & Mailhot 211). Using the information-theoretic concepts of entropy and surprisal allows us to provide a unified model of the frequency effects on /d/ deletion in Spanish participles.