

# 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.

### Types of frequency effects

<b>Absolute token frequency</b>	High frequency → Easily accessed → Phonetic reduction (Bybee 2000, 2006)
<b>Relative frequency</b>	<ul style="list-style-type: none"> <li>Frequency of an element in a particular sequence</li> <li>Distinct effect from that of absolute frequency in language processing (Hay 2001)</li> </ul>
<b>Neighborhood density</b>	<ul style="list-style-type: none"> <li>Role of phonological similarity</li> <li>Low density → Easily accessed → Phonetic reduction (Luce 1986, Luce &amp; Pisoni 1998, Ussishkin &amp; Wedel 2009)</li> </ul>

### Intervocalic /d/ deletion in Spanish

Bybee 2001

Higher rates of deletion in high frequency lexical items.	Deletion is more advanced in the <b>-ado</b> morpheme in past participles	No effect of token frequency was found within <b>-ado</b> participles
---	---	---

*cantado, pasado, estado, amado*

### 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**:

**Entropy** → Measure of the uncertainty associated with selecting a possible outcome related to the message.

**Surprisal** → Measure of the information content associated with a particular outcome. Negative logarithm of its probability.

## Methodology

### Data

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

Independent variables	<b>Absolute frequency</b>	<ul style="list-style-type: none"> <li>List of lexical frequencies in the CREA</li> <li>Overall use of lexical items in Spanish</li> </ul>
	<b>Frequency of construction</b>	<ul style="list-style-type: none"> <li>Searches for a particular construction in CREA</li> <li>"he cantado", "estoy enojado", "año pasado"...</li> </ul>
	<b>Neighborhood density</b>	<ul style="list-style-type: none"> <li>Number of lexical neighbors in CREA</li> <li>Addition, subtraction or change of one segment</li> </ul>
	<b>Relative frequency in neighborhood</b>	<ul style="list-style-type: none"> <li>(absolute frequency – mean frequency of neighbors)</li> <li>Effect of frequency over neighborhood density (Ussishkin &amp; Wedel 2009)</li> </ul>

- 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.

Standard Frequency Indices (Luce 1986)

$$SFI = 40 + 10 \cdot (\log(\text{no. of occurrences} / \text{corpus size}) + 10)$$

## Frequency, surprisal and entropy: Predictions

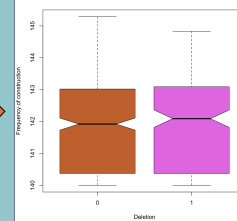
- High frequency is associated with **low surprisal**
  - Predictable and unstable elements (Hume & Mailhot 2011)
- **P1** As the absolute frequency and the frequency of construction **increase**, the odds of deletion will also **increase**
- In high density neighborhoods each segment contributes more to **system entropy**.
  - In low density neighborhoods segments are more likely to be **redundant**
- **P2** As the number of neighbors **decreases**, the odds of deletion will **increase**
- Tokens with **high relative frequency** within the neighborhood are less sensitive to neighborhood density (Ussishkin & Wedel 2009)
- **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; i.e. as frequency **increases** the rates of deletion **decrease**. This is an unexpected result.
- Predictions 2 – 4** were not confirmed.

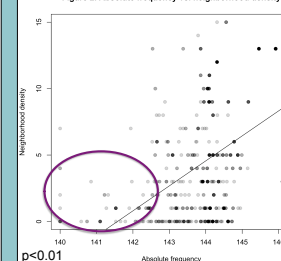
- Figure 1** shows the distribution of the two populations – tokens with /d/ deletion (1) and tokens without it (0) – with respect to **frequency of construction**.
- The median frequency of the deletion group is **higher** than the median of the maintenance group.
- The overlap of the notches indicates that the difference in the medians of the two groups is **not statistically significant** when frequency of construction is considered alone.

Figure 1. Effect of frequency of construction on /d/ deletion



## Absolute frequency and neighborhood density

Figure 2. Absolute frequency vs. neighborhood density

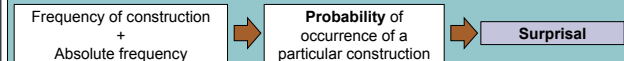


- The effect of absolute frequency stands even if frequency of construction is removed from the model
- A correlation (**Figure 2**) was found between absolute frequency and neighborhood density, which might explain the unexpected results for absolute frequency.
- As frequency increases, the number of 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:



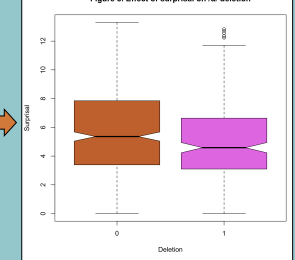
"ha estado" →  $P = \text{freq. of construction} / \text{absolute freq.} = 4833 / 100757 = 0.05$   
 $S = -\log_2(P) = -\log_2(4833 / 100757) \approx 4.38$

- In the example above, the surprisal of observing "ha estado" is 4.38
- Surprisal was added to the analysis in place of absolute frequency and frequency of 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 median of the maintenance group.
- The lack of overlap of the notches indicates that the difference in the medians of the two groups is **statistically significant**.

Figure 3. Effect of surprisal on /d/ deletion



- The results suggest that **surprisal** is a better predictor of intervocalic /d/ deletion in Spanish first conjugation past participles than considering absolute frequency and the frequency of a construction independently.

## Conclusion

- 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 low 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.