

Weight and Subject Position in Spanish: a Corpus Analysis

1. Introduction

Abundant research reports that weight affects the position of postverbal constituents (Hawkins 1994, 2004, Wasow 1997, 2002, Wasow and Arnold 2003). Our study analyzes whether preverbal positions are also affected by this processing constraint. In particular, we focus on the position of Spanish subjects with unaccusative verbs, because they may precede or follow the verb (Torrego 1989). We show that weight also affects preverbal positions. In addition, we investigate which is the most effective way of measuring weight: by words (Lohse, Hawkins and Wasow 2004), syllables (Gries 2003), or phonemes. All of these strategies prove weight to be statistically significant, although words were the optimal measure.

2. Methodology

We coded a sample of 410 sentences from the online corpus CREA (*Corpus de Referencia del Español Actual*). Using the statistical tool SPSS, we applied cross-tabulations and the χ^2 test to determine the significance of each independent variable. A value of $p=.05$ is the threshold for significance. The addition of a binary logistic regression analysis ensures the fitness of the model. The dependent variable was *subject position* (preverbal, postverbal), and the independent variable was *subject weight*, measured in words, syllables and phonemes. We wrote a Python software application to calculate weight automatically. Our software is even able to identify the number of syllables and phonemes of any data string written in Spanish.

3. Results

Subject weight is statistically significant regardless of the measuring method. Weight was partitioned into categories (Hawkins 2004). We divided constituents into three groups: for example, those with 1 or 2 words, 3 or 4, and 5 or more. We adopted similar divisions for syllables and phonemes. Table 1 shows that the percentage of postverbal subjects is higher as the number of words increases ($p=0.000$). Thus, heavier subjects tend to appear postverbally with Spanish unaccusatives. In table 2 a division by phonemes yields similar results ($p=0.000$). Syllabic weight also proved significant ($p=0.000$).

4. Conclusion

Our research shows that *weight* influences preverbal positions in Spanish. To our knowledge, this is a novel research finding in the weight literature. Our study also sheds light on the controversy around the most appropriate method to estimate weight (Wasow 2002, Rosenbach 2005). Our software provides a research tool for future corpus-based linguistic studies.

Table 1: Subject weight by words			Subject Weight			Total
			1-2 words	3-4 words	5 or more	
Position of Subject	Preverbal	count	101	43	66	210
		% column	69.7%	58.1%	34.7%	51.3%
	Postverbal	count	44	31	124	199
		% column	30.3%	41.9%	65.3%	48.7%
Total		count	145	74	190	409
		Value	Degrees of Freedom		Significance	
Pearson Chi Square		41.792 ^a	2		.000	

Table 2: Subject weight by phonemes			Subject Weight			Total
			1-10 phon.	11-20 phon.	21 or more	
Position of Subject	Preverbal	count	63	76	70	209
		% column	67.0%	63.3%	36.3%	51.4%
	Postverbal	count	31	44	123	198
		% column	33.0%	36.7%	63.7%	48.6%
Total		count	94	120	193	407
		Value	Degrees of Freedom		Significance	
Pearson Chi Square		33.709 ^a	2		.000	