

Noun to Verb ratio in Basque, Spanish, Catalan and Galician

All languages distinguish the syntactic categories Noun and Verb (Croft 2000). The noun-verb distinction relies on inflectional morphology, semantic correspondence and syntactic distribution (Baker 2003; Chung 2012). Based on the universality of nouns and verbs, some studies have observed correlations between the relative distribution of these two categories cross-linguistically and certain linguistic traits, and have tried to explain what factors account for them. This study tests the distribution of nouns vs. verbs in digital corpora in the languages of Spain (Basque, Spanish, Catalan and Galician) according to two different noun-verb ratio calculi: (a) Seifart (2011) and (b) Polinsky (2012).

Seifart (2011) investigates the distribution of nouns and verbs in five languages (Baure, Chintang, Bora, N|uu and Sri Lanka Malay). He uses DoBeS corpora that include texts from various narrative genres, conversations, procedural texts and some songs, and calculates the noun-verb ratio dividing the number of nouns by that same number plus the number of verbs (nouns / [nouns + verbs]). Results show that the noun-verb ratio correlates with the argument marking system of the language: languages with rich argument marking on verbs (agreement) have lower noun-verb ratios. Polinsky (2012) investigates the noun-verb ratio in 30 languages. She uses the available wordnet lexical databases (Miller et al. 1990) of each language to obtain the frequencies of nouns and verbs, and she calculates the noun-verb ratio by dividing the number of nouns by the number of verbs that are listed for each language (nouns / verbs). The results reveal a correlation between the noun-verb ratio and the headedness of the language (OV/VO). Head-initial languages (VO) have a lower noun-verb ratio than head-final languages (OV).

Here we present data from a cross-linguistic study of the noun-verb ratios in the languages spoken in Spain (Basque (non-Indo-European, OV), Spanish, Catalan and Galician (all Romance and VO)). To this end, we used digital corpora: the ETC corpus for Basque; AnCora corpus for Spanish and Catalan; and CORGA corpus for Galician. Each digital corpus contains about 500,000 words for each language, obtained from newspaper articles. These corpora, unlike those used by Polinsky (2012), reflect the usage frequency of nouns and verbs in each language, that is, it is not a list of nouns and verbs of each language, but a collection of texts/sentences. Our results reveal that the distribution of the noun-verb ratio of Basque, Spanish, Catalan and Galician converges with the results in Polinsky (2012).-There is a correlation between the noun-verb ratio and the headedness: languages with low noun-verb ratio are VO languages, and those with high noun-verb ratio OV languages. However, our results do not converge with the results in Seifart (2011). Basque, which has verb agreement with three arguments, shows a higher noun-verb ratio than Spanish, Catalan and Galician, which are subject-verb agreement languages.

References:

- Baker, M. C. (2003). Lexical categories and the nature of the grammar, *Lexical Categories: Verbs, Nouns and Adjectives* (pp. 264-302). Cambridge: Cambridge University Press.
- Chung, S. (2012). Are lexical categories universal? The view from Chamorro, *Theoretical Linguistics*, 38, 1-56.
- Croft, W. A. (2000). Parts of speech as typological universals and as language particular categories, In P. Vogel & B. Comrie (Eds.), *Approaches to the typology of word classes* (pp. 65-102). Berlin: Mouton de Gruyter.
- Miller, G., Beckwith, R., Fellbaum, C.D., Gross, D., & Miller, K. (1990). WordNet: An online lexical database, *International Journal of Lexicography*, 3, 235-244.
- Polinsky, M. (2012). Headedness, again. In T. Graf & D. Paperno & A. Szabolsci & J. Tellings (Eds.), *Theories of Everything: In Honor of Ed Keenan* (Vol. 17, pp. 348-359). Los Angeles: UCLA.
- Seifart, F. (2011). Cross-linguistic variation in the noun-to-verb ratio: the role of verb morphology and narrative strategies, *Association for Linguistic Typology 9th Biennial Conference*. University of Hong Kong, Japan.