Underdetermination of semantic attributes for the optimization of functional features. A study on Italian.

Francesca Franzon
Universitat Pompeu Fabra

Natural languages consistently convey through grammatical features a small set of semantic attributes concerning the referential world. An example is the nominal morphology of Romance languages, in which inflectional gender and number features can encode the animacy, sex, and numerosity of the referents. The extra-linguistic cognitive representation of these salient attributes, carried out by core knowledge systems (Spelke and Kinzler, 2007) is paramount for humans to engage in biologically successful behaviors, such as choosing to aggregate to the largest available group of social companions (Piantadosi and Cantlon, 2017). Therefore, the ability to transmit all the possible information concerning these semantic attributes seems a desirable property for languages.

Nevertheless, grammatical systems never display a one-to-one encoding of the information that can be processed by the core knowledge systems (Franzon et al., 2019). For instance, in Romance languages, singulars do not only encode numerosities equals to one, as in sentences denoting kinds or mass references. Similarly, morphological masculines are used to refer to male as well as to non-male individuals.

Some reasons of this apparently suboptimal encoding can be sought in the functional role of inflectional features in tracing agreement patterns. Besides this crucial role in the construction of sentences, the presence of agreement can enhance language processing by sustaining the prediction of upcoming words; e.g., hearing or reading a determiner in the singular feminine allows the receiver of a linguistic signal to expect an upcoming related noun in the singular feminine.

I will discuss how the semantic and the functional aspects compete and cooperate in shaping morphological systems. As a case study, I will present large scale data on the distribution of the nominal lexicon of Italian (Franzon and Zanini, 2022; Pescuma et al., 2021). By means of entropy metrics, we assessed the primarily functional purpose of gender and number features in the lexicon, observing a distribution of nouns that can optimally serve agreement-based parsing and prediction of words in sentences. The apparently suboptimal encoding of semantic attributes observed in morphology, consisting in the lack of one-to-one correspondences between semantic attributes and morphological features, is the hallmark of the particular type of information compression occurring in inflectional systems, namely underdetermination. Being realized at least in one value for any system, underdetermination is a foundational property of morphology. Besides unlocking scalability of inflection to novel referents and allowing speakers to avoid redundant signals, underdetermination has the fundamental role of removing semantic constraints to perform functional operations such as agreement.

References

Franzon, F. and Zanini, C. (2022). The entropy of morphological systems in natural languages is modulated by functional and semantic properties. *Journal of Quantitative Linguistics*, pages 1–25.

Franzon, F., Zanini, C., and Rugani, R. (2019). Do non-verbal number systems shape grammar? numerical cognition and number morphology compared. *Mind & Language*, 34(1):37–58.

Pescuma, V. N., Zanini, C., Crepaldi, D., and Franzon, F. (2021). Form and function: A study on the distribution of the inflectional endings in italian nouns and adjectives. *Frontiers in Psychology*, page 4422.

Piantadosi, S. T. and Cantlon, J. F. (2017). True numerical cognition in the wild. *Psychological science*, 28(4):462–469.

Spelke, E. S. and Kinzler, K. D. (2007). Core knowledge. *Developmental science*, 10(1):89–96.