Form, function, meaning. A study on the distribution of inflectional morphemes in Italian.

Francesca Franzon¹, Chiara Zanini² and Valentina N. Pescuma¹

¹ Neuroscience Area, International School for Advanced Studies (SISSA), Trieste, Italy
² Romanes Seminar (RoSe), Universität Zürich, Switzerland

Abstract: The current paper investigates the distribution of inflectional morphemes in Italian. The analysis is based on a 2 billion token corpus of the Italian National Corpus and the morphologically tagged corpus of the Italian Web of Information (itWaC). The study reveals that the distribution of inflectional morphemes varies across levels of processing and semantic interpretation.

Introduction

- Inflectional morphology: meaning, form and function
- Inflectional features in Italian: masculine/feminine (gender), singular/plural (number).
- Allomorphy and syncretism: inconsistency between form and feature.
- Some research discusses the role of transparency in morphological processing [1, 2].

Aim: To measure the extent of allomorphy and syncretism, in order to provide an initial ground for future exploration of research questions concerning how our cognitive system deals with such inconsistency (e.g., does it equally affect different levels of processing?).

Method

- Two sources merged: itWaC, a 2 billion token web-based corpus [3] and Morph-It, a list of morphologically tagged words [4].
- Exclusion of homographs with ambiguous inflectional features, such as ‘cameriere’, which can be both M-SG (waiter) and F-PL (waitresses).
- Final dataset: 22,638 morphologically tagged noun types – 209,942,221 tokens.

Outcomes and discussion

- In our dataset, noun types quite evenly distributed across inflectional features.
- Near-maximal entropy of type distribution optimization of inflectional features to reduce uncertainty in sentence processing.
- Computation of type and token frequency of nouns for each inflectional feature (Gender and Number): extraction of each word form’s last character (morpheme) → number of different morpheme types linked to each inflectional feature → type and token frequency of nouns ending in each of them.
- Computation of entropy values for the distribution of type and token frequencies.

Results

- Prevalence in the association of each feature with one morpheme; observed decrease in type frequency of the other associated morphemes.
- Persistence of lower type frequency morphemes possibly explained by their higher token frequency (cf. discriminative learning accounts; e.g., [6]).

- Association of the same morpheme with more than one feature: -a, typically feminine singular, frequently associated with other features as well; -e ambiguous in the singular.

References


The observed diversity of morpheme types within feature may result from communication and learning pressures [2].

Conversely, the association of the same morpheme with more than one feature seems undesirable, but it is far from uncommon.

Plurals display less types of morphemes than singulars, suggesting a more stable association between form and meaning.

Allomorphy appears to affect more evidently more functionalized features.