

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 Romanisches Seminar (RoSe), Universität Zürich, Switzerland

✉ ffranzon@sissa.it, chiara.zanini@uzh.ch, vpescuma@sissa.it



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.

Feminine singular	-a casa (house) -e volpe (fox) -i crisi (crisis) -o mano (hand) -à città (city)
-a	F-SG casa (house) M-SG pianeta (planet) F-PL ossa (bones) M-PL tosaerba (mower)

- 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?).

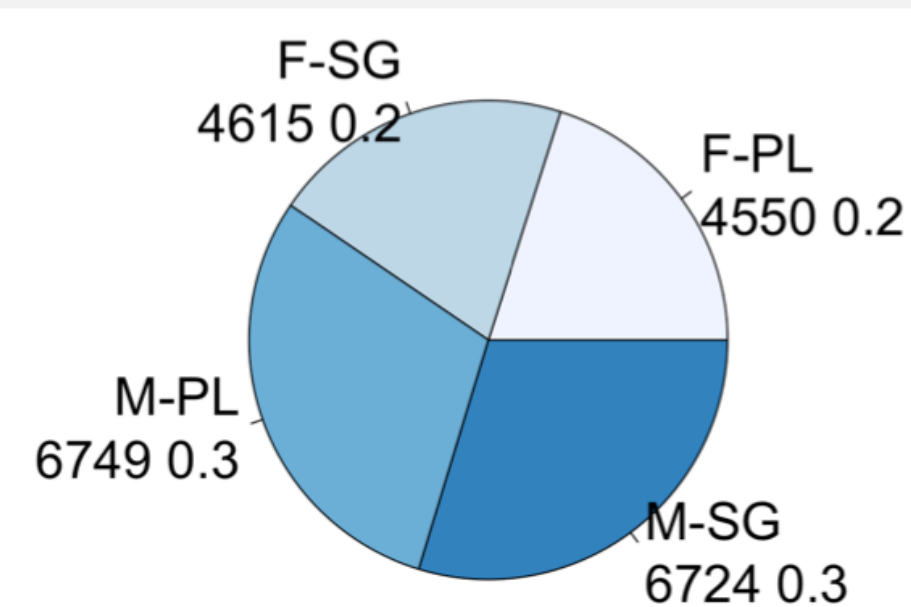
Method

- Two sources merged: **itWaC**, a 2 billion token web-based corpus [3] and **Morph-iti**, 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.**

- 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.

Outcome and discussion

- In our dataset, noun types quite evenly distributed across inflectional features.



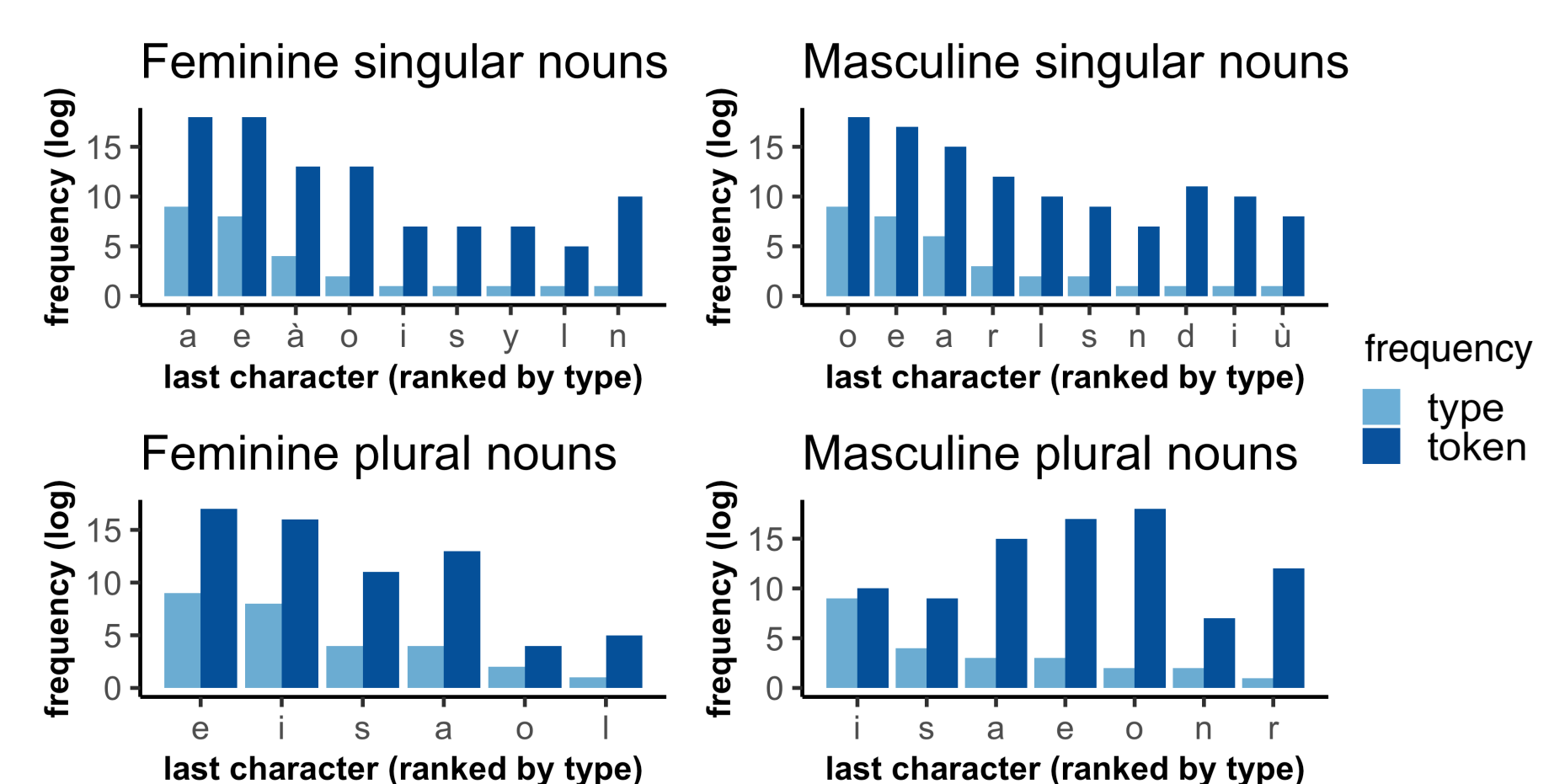
- **Near-maximal entropy** of **type** distribution → optimization of inflectional features to reduce uncertainty in sentence processing.

	M-SG	M-PL	F-SG	F-PL
Type	6724	6749	4615	4550
Frequency	0.297	0.298	0.204	0.201
H(Type)	1.974			

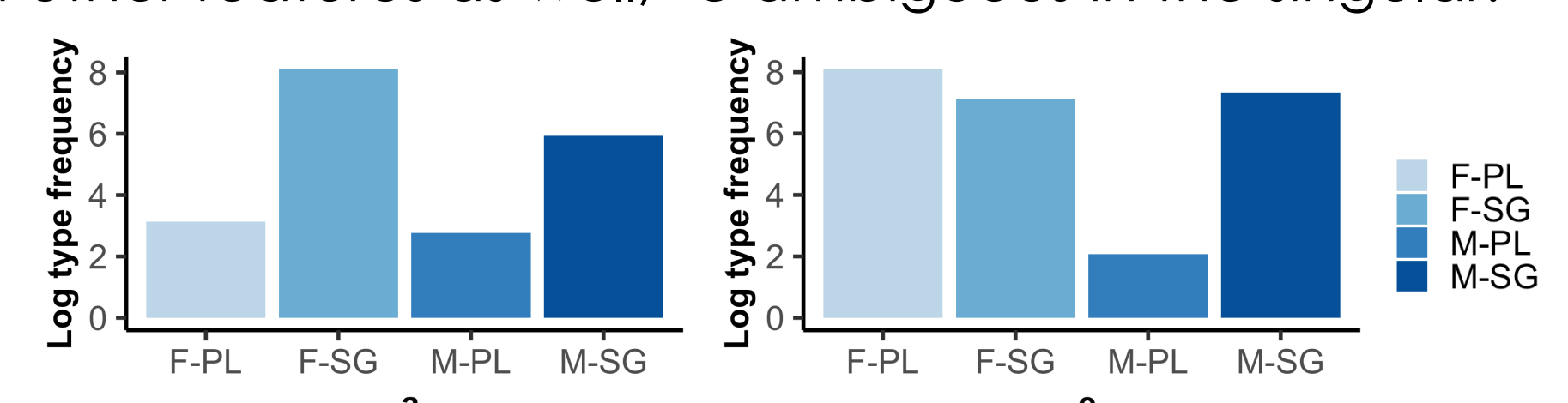
- **Information decreases** to 1.882 for **token** distribution, due to the increased proportion of singulars → **singulars** can be used as a **default** value yielding no semantic interpretation about numerosity, vs **plurals**, whereby a less ambiguous encoding of semantic interpretation about numerosity results in greater information [5].

	M-SG	M-PL	F-SG	F-PL
Token	74588931	36446318	71458795	27448177
Frequency	0.355	0.174	0.340	0.131
H(Token)	1.882			

- 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]).



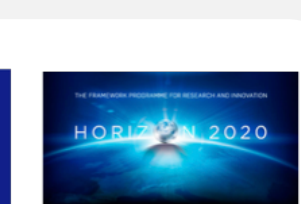
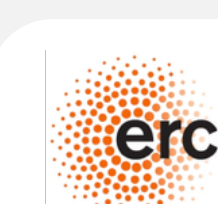
- Association of the same morpheme with more than one feature: **-a**, typically feminine singular, frequently associated with other features as well; **-e** ambiguous in the singular.



- 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**.

References

[1] De Bleser, R., & Luzzatti, C. (1994). Morphological processing in Italian agrammatic speakers syntactic implementation of inflectional morphology. *Brain and Language*, 46(1), 21–40. [2] Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of "mouses" in adult speech. *Language*, 760-793. [3] Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, 43(3), 209-226. [4] Zanchetta, E., & Baroni, M. (2005). Morph-iti: A free corpus-based morphological resource for the Italian language. In: *PROCEEDINGS OF CORPUS LINGUISTICS*, [HTTP://DEV.SSLMIT.UNIBO.IT/LINGUISTICS/MORPH-IT.PHP](http://dev.sslmit.unibo.it/linguistics/morph-iti.php). [5] Arcara, G., Franzon, F., Gastaldon, S., Brotto, S., Semenza, C., Peressotti, F., & Zanini, C. (2019). One can be some but some cannot be one: ERP correlates of numerosity incongruence are different for singular and plural. *Cortex*, 116, 104-121. [6] Blevins, J. P., Milin, P., & Ramscar, M. (2017). The Zipfian paradigm cell filling problem. In *Perspectives on Morphological Organization* (pp. 139-158). Brill.



This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme Grant Agreement No 679010 STATEARN-ERC-2015-STG/ERC-2015-STG.



Get PDF here