Applying multiple correspondence analysis to P-demotion domain

Mohammad Tavakoli
Adam Mickiewicz University in Poznan

Keywords: P-demotion, multiple correspondence analysis, multivariate analysis, dimension reduction, statistical analysis

Abstract

P-demotion is a family of voice and valency alternations distinguished by two key features: i. Syntactic demotion of the P-argument and ii. Unaffected argument structure, that is the agent remains agent and patient remains patient (Janic and Witzlack-Makarevich 2021; Zúñiga and Kittilä 2019). Therefore, P-demotion operation yields a variety of intransitive constructions discussed in the literature under different names such as antipassive (1), noun incorporation, conative, and A-labile constructions in which P-argument is either expressed as an oblique, incorporated, suppressed, or omitted (Heaton 2017; Janic 2020).

1)
a. John shot the lion.
b. John shot at the lion.

In this research, we aim to investigate the correlation between form and function in P-demotion domain using a statistical approach. To do so, we have coded formal features, including voice marking, flagging, and indexation, and functional features, including referentiality and affectedness of different types of P-demotion realizations within 55 languages from six macro-areas. We then consider the database as a high-dimensional matrix with more than 90 rows containing various P-demotion realizations and columns including relevant formal and functional features. To interpret this high-dimensional matrix and to capture the potential correlations between formal and functional properties we apply a multivariate analysis technique, which further help us establish if the language's syntactic structure is arbitrary or not when it comes to P-demotion.

We apply multiple correspondence analysis (MCA) from a family of dimension reduction techniques to the P-demotion domain to reduce the large number of dimensions in the original database to an interpretable small set of underlying dimensions. There are several reasons why MCA fits our purpose. Most importantly, this exploratory technique allows us to capture the frequency-based associations between form and function in our complex dataset as MCA attempts to capture as much variation as possible within the first few dimensions (Glynn 2014; Husson, Le, and Pagès 2010; Levshina 2015). MCA facilitates the interpretation of the associations between categories plotted on the same biplot, here our formal and functional features as represented in Fig. 1. For example, here the result (produced using an early version of database) confirms the correlation between oblique flagging and referentially of the object represented by their close distance on the plot. We also expect to capture the correlations for sparse features as with MCA one will still observe correlations that are likely to be due to small numbers of a given feature (Glynn 2014).

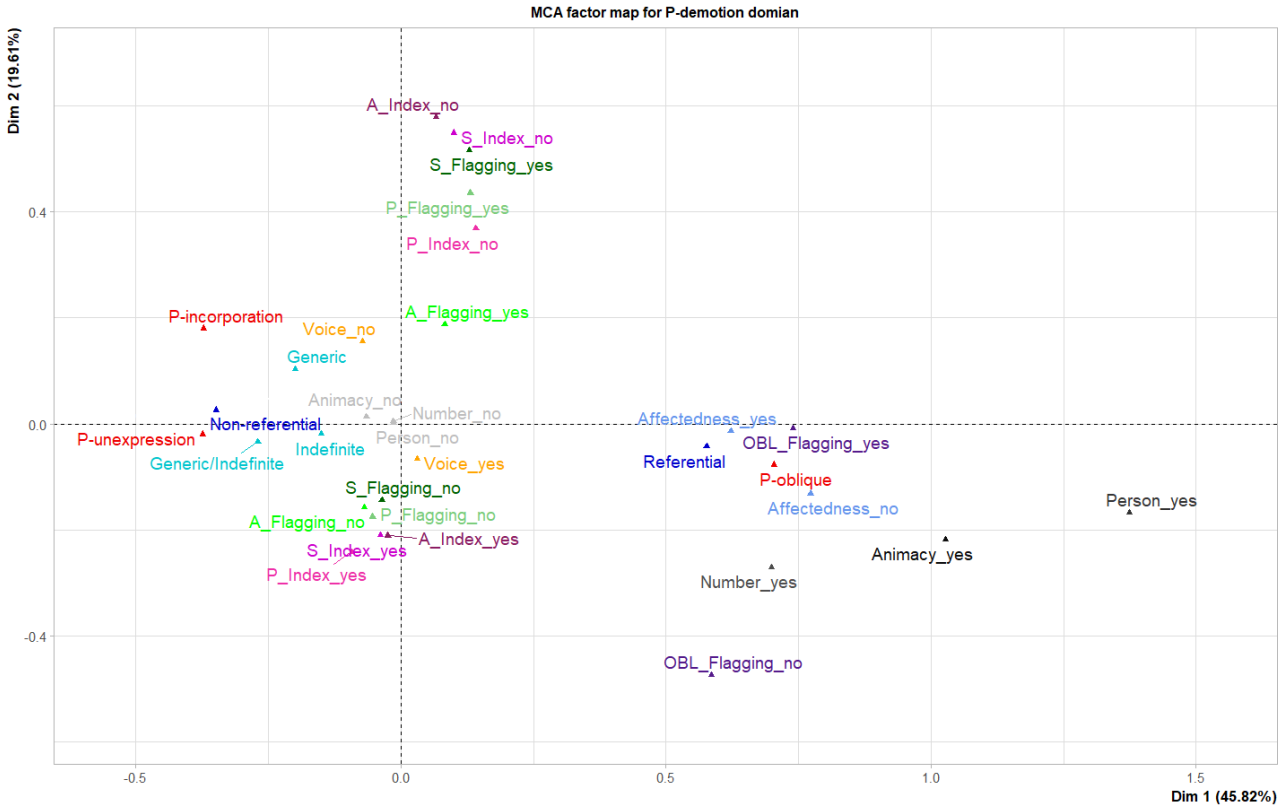

Figure 1. multiple correspondence analysis of P-demotion domain

This study is part of the project Polonez Bis co-funded by the National Science Centre in Poland and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie (grant agreement No 945339), headed by Katarzyna Janic as Principle Investigator.

References

Glynn, Dylan. 2014. "Correspondence Analysis: Exploring Data and Identifying Patterns." In *Quantitative Studies in Polysemy and Synonymy*, 443–86. Amsterdam/Philadelphia: John Benjamins.

Heaton, Raina. 2017. "A Typology of Antipassives, with Speciation Reference to Mayan." University of Hawaii at Manoa.

Husson, Francois, Sebastien Le, and Jérôme Pagès. 2010. *Exploratory Multivariate Analysis by Example Using R*. New York: CRC Press.

Janic, Katarzyna. 2020. "A Case Study of the P Demotion Operation." *Forum for Contemporary Issues in Language and Literature*, 56–92.

Janic, Katarzyna, and Alena Witzlack-Makarevich. 2021. *Antipassive: Typology, Diachrony, and Related Constructions*. Amsterdam: John Benjamins.

Levshina, Natalia. 2015. *How to Do Linguistics with R*. Vol. 10. Amsterdam: John Benjamins.

Zúñiga, Fernando, and Seppo Kittilä. 2019. *Grammatical Voice*. Cambridge: Cambridge University Press.