Open Library of Humanities

EDITORIAL

Empirical Approaches to Portuguese Linguistics – New insights from studies in various areas of grammar

Irene Fally, University of Vienna, AT, irene.fally@univie.ac.at

Pamela Goryczka, University of Vienna, AT, pamela.goryczka@univie.ac.at

Anna Kocher, Catholic University of Eichstaett-Ingolstadt, DE, anna.kocher@ku.de

Over the last two decades, empirical methods have become increasingly prominent in linguistics. Even traditionally more introspective theoretical frameworks have begun to embrace empirical methods to test their hypotheses. This special collection unites works applying state of the art empirical methods to the study of Portuguese and its varieties, to highlight the applicability of these methods to Portuguese linguistics. In doing so, this collection aims to show the potential and the possible limitations of these methods. This special collection brings together contributions on different varieties of Portuguese from the subfields of phonology, morpho-syntax, syntax and semantics, using a variety of empirical methods: The methods range from classic corpus linguistics and sociolinguistic field methods to psycholinguistic experiments. The aim of this collection is to illustrate how these methods can be applied to the study of Portuguese in all linguistic subfields. By explicitly addressing how empirical methods contribute to furthering our understanding of Portuguese linguistics, we want to increase awareness for empirical methods and their applicability to the study of Portuguese, and ultimately encourage researchers to make greater use of them.

Understanding the language system, its representations, changes, and the dynamics within ultimately requires the observation of language data as a manifestation of linguistic behavior. The question of what counts as evidence in linguistics and how a theory of language can be tested, continually sparks discussion to what degree linguistics is to be considered an empirical branch of science (cf. Kepser & Reis 2008). This is partly due to the different lines of research that have developed within linguistics over the years. Particularly with the advent of Generative Grammar, discussions of how certain claims can be justified have occupied many a linguist in the field (cf. e.g., Bever, Fodor & Weksel, 1965; Linell, 1976).

Over the last two decades, empirical techniques have become increasingly prominent in linguistics (cf. e.g., Sampson, 2002). Even traditionally more introspective theoretical frameworks have begun to embrace experimental methods to test their hypotheses (Phillips & Lasnik, 2003). This special collection unites works applying state of the art empirical methods to the study of Portuguese to highlight the applicability of these methods in Portuguese linguistics. In doing so, this collection aims to show the potential and the possible limitations of these methods.

This special collection is the result of the first Vienna Workshop on Portuguese Linguistics, which focused on *Empirical Research on Portuguese*, showcasing a variety of empirical methods and their application on Portuguese language data. The goal of the workshop was to probe the empirical bases of (Portuguese) linguistics and to discuss methodological issues when working with linguistic data: As Grieve (2021) points out, the study of language has evolved greatly over the last decades, resulting in advances in the way linguistic behavior can be measured and quantified, while at the same time making the issue of generalizability and reproducibility more pressing than ever.

Following Grieve's (2021) call for a combination of experimental and observational methods in linguistics, this special collection brings together contributions on different varieties of Portuguese focusing on phonology, morpho-syntax, syntax and semantics and using a variety of empirical methods. Techniques of data collection comprise classic corpus-based analyses (Nkollo, Balla-Johnson) and corpus-driven computational approaches (Santos & Simões) as well as observational sociolinguistic field work (Henriques Pestana) and psycholinguistic experiments (Garica & Guzzo, Flores et al., Wall). The aim of this collection is to illustrate how these methods can be applied to the study of Portuguese in all linguistic subfields. By explicitly addressing how empirical methods contribute to furthering our understanding of Portuguese linguistics, we want to increase awareness for empirical methods and their applicability to the study of Portuguese, and ultimately encourage researchers to make greater use of them.

In what follows, we briefly introduce and contextualize the contributions to this collection.

Portuguese corpus linguistics has been a growing field since the early 2000s (cf. Davies, 2008). By today, a variety of resources are available for corpus-based analyses ranging from

very large web-crawled corpora to reference corpora and more specialized dialectal and variety specific corpora.

In his article *Nasal epenthesis in preverbal accusative clitic pronouns. A variationist study of present-day dialectal European Portuguese,* Mikołaj Nkollo investigates an external sandhi phenomenon in dialectal European Portuguese in which the vowel-initial preverbal clitic pronouns *a, o, os, as* are realized with a nasal epenthesis as *na, no, nos, nas*. The study is based on data from the dialectal corpus CORDIAL-SIN. Nkollo argues in favor of the assumption of distinct grammars (prosody-syntax mappings) that systematically differentiate with respect to whether nasal epenthesis only applies to preverbal clitic pronouns or also affects other monosyllabic clitics (definite articles and demonstrative pronouns). The author explores the role phonology triggers, morphosyntax and proclisis triggers have in conditioning the realization of the nasal epenthesis. He furthermore discusses the historical development of this sandhi phenomenon and identifies similarities between modern dialectal European Portuguese and earlier stages of the language.

William Balla-Johnson's article *The Portuguese Pluperfect: Development and the effect of auxiliary* ter *generalization* combines corpus data analysis with linguistic theorizing in diachrony. The author discusses the diachronic development of the Portuguese pluperfect and proposes that the generalization of the auxiliary *ter* over *haver* is the primary reason for the prolonged survival of the Latin synthetic pluperfect form in Portuguese as opposed to the pattern usually found in other Romance languages, such as French and Spanish. The observation is empirically supported by quantitative and qualitative analyses of data from the *Corpus do Português*. Balla-Johnson's contribution highlights the ongoing relevance of traditional corpus-based methods in diachronic studies.

In their paper *Clustering emotions in Portuguese*, Diana Santos and Alberto Simões explore linguistic expressions of emotions in Portuguese from a computational point of view. The authors study different statistical methods to evaluate which can be used to create and refine semantic embeddings in a large textual database of Portuguese. They make use of a semi-automatically annotated corpus of emotions in Portuguese based on the resources from *Linguateca*, to understand how co-occurrences and word-embeddings can help form semantic clusters based on emotions. In this way they illustrate how different computational linguistic techniques can be employed on Portuguese data and highlight some crucial issues that arise from the application of the individual measures.

Observing language from a sociolinguistic perspective has always played a crucial role in the study of Portuguese due to its wide range of regional and social varieties (cf. Barbosa et al., 2017). Analyzing language use in real-life situations gives insight into how language is shaped by social variables (e.g., age, gender, education) and, crucially, how it changes over time and in different social contexts. Sociolinguistic interviews are still the most common method of

eliciting (more or less) natural speaking situations, which provide rich data on the social factors influencing language use.

In her article *On the presence and absence of definite articles with anthroponyms in rural varieties of Madeiran Portuguese*, Yoselin Henriques Pestana examines the variable use of definite articles preceding anthroponyms based on a corpus of semi-directed interviews and free conversations between elderly speakers of rural varieties of Madeiran Portuguese. To account for the variation, the author proposes a more fine-grained categorization of intersubjective proximity considering different types of social relations and kinship. The distinction proves essential since the nature of the relationship between speaker and interlocutor appears to facilitate the production of anthroponyms in general. Additionally, shared knowledge between speech act participants seems to influence the use of definite articles. On the whole, the study thus also contributes to the discussion of biases of different methods in data collection.

To gain quantitative evidence in the study of linguistic representations, (psycho-)linguistic experiments allow for administering tasks in a controlled environment (Hemforth, 2013). Over the last decade, experimental approaches have been used more widely for the study of Portuguese (Leal & Gupton, 2021). Experimental approaches allow to tap into language processing directly, obtain meta-linguistic judgements and collect language production data.

In *Lexical access in Portuguese stress*, Guilherme Duarte Garcia and Natália Brambatti Guzzo tackle the issue of lexical stress in Brazilian Portuguese from an experimental perspective. They question the validity of modeling lexical stress as a categorical distinction between regular and irregular stress, arguing for the adoption of a probabilistic model of lexical stress. Garcia & Guzzo use a lexical decision task – an established experimental method in lexical processing – to explore the role of stress patterns in lexical retrieval, combining the experimental technique with exploratory Bayesian analyses. The authors explore both classical and novel approaches to stress in Portuguese and combine this with experimental analyses to call into question the usefulness of categorical distinctions in lexical stress patterns.

Number-neutral indefinite objects in Brazilian Portuguese as a case of semantic incorporation by Albert Wall deals with an understudied phenomenon in Brazilian Portuguese. He investigates cases of indefinite objects that, in combination with a particular group of verbs, are interpreted as number-neutral despite the presence of an article (for instance, *assistir uma televisão* 'watch (a) television'). Wall proposes to treat these indefinite objects on par with bare objects as semantically incorporated structures with a non-quantified atelic activity reading. He bases his argumentation on a broad set of experimental tasks that offer a first systemic approach to the properties of this construction. Wall also addresses the pragmatic effects of the construction in the context of related phenomena in Brazilian Portuguese and English.

Cristina Flores, Esther Rinke, Jacopo Torregrossa and Daniel Weingärtner make use of written language productions to investigate the development of the verbal domain in Portuguese-German bilingual children. In their article *Language separation and stable syntactic knowledge: Verbs and verb phrases in bilingual children's narratives*, they tackle a core question of heritage language research, namely to what extent language competence develops in the heritage language and how this relates to the environmental language. Employing a cloze test, they tested productive abilities of verbs, in terms of lexical richness, (correct) tense morphology and word order. In their analysis, Flores et al. consider language-specific properties of verbs and verb phrases, as well as contextual variables, such as age-of-onset of acquisition and participants' age at the time of testing. With this work, the authors challenge deficit-oriented accounts of bilingual language acquisition.

The contributions in this Special Issue *Empirical Approaches to Portuguese Linguistics – New insights from studies in various areas of grammar* may differ with regard to the varieties of Portuguese under investigation as well as the phenomena discussed, yet they demonstrate quite clearly that linguistic analysis essentially thrives on empirical data, whether based on more traditional corpus and sociolinguistic approaches or on experimental techniques of data elicitation.

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Barbosa, P., de Paiva, M. da C., & **Rodrigues, C.** (Eds.) (2017). *Studies on Variation in Portuguese*. Amsterdam: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/ihll.14

Bever, T. G., Fodor, J. A., & **Weksel, W.** (1965). Is linguistics empirical? *Psychological Review*. *American Psychological Association, 72*(6), 493–500. DOI: https://doi.org/10.1037/h0022698

Davies, M. (2008). New Directions in Spanish and Portuguese Corpus Linguistics. *Studies in Hispanic and Lusophone Linguistics, 1*(1), 149–186. DOI: https://doi.org/10.1515/shll-2008-1009

Grieve, J. (2021). Observation, experimentation, and replication in linguistics. *Linguistics, 59*(5), 1343–1356. DOI: https://doi.org/10.1515/ling-2021-0094

Hemforth, B. (2013). Experimental Linguistics. In *Linguistics*. Oxford University Press. DOI: https://doi.org/10.1093/obo/9780199772810-0112

Kepser, S., & **Reis, M.** (2008). Evidence in Linguistics. In S. Kepser & M. Reis (Eds.), *Linguistic Evidence* (pp. 1–6). Berlin, New York: De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110197549.1

Leal, T., & **Gupton, T.** (2021). Acceptability Experiments in Romance Languages. In G. Goodall (Ed.), *The Cambridge Handbook of Experimental Syntax* (Cambridge Handbooks in Language and Linguistics), 448–476. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/9781108569620.018

Linell, P. (1976). Is Linguistics an Empirical Science? *Studia Linguistica, 30*(1). 77–94. DOI: https://doi.org/10.1111/j.1467-9582.1976.tb00637.x

Phillips, C., & **Lasnik, H.** (2003). Linguistics and empirical evidence: Reply to Edelman and Christiansen. *Trends in Cognitive Sciences, 7*(2), 61–62. DOI: https://doi.org/10.1016/S1364-6613(02)00046-3

Sampson, G. (2002). *Empirical Linguistics*. London & New York: Continuum.