Challenges of modeling Portuguese morphology in a computational grammar of Portuguese in the HPSG formalism

Alexandre Rademaker[1], Leonel Figueiredo de Alencar[2], and Ana Luiza Nunes[3]

[1] *School of Applied Mathematics of the Getúlio Vargas Foundation and IBM Research, Brazil,* alexandre.rademaker@fgv.br
[2] *Graduate Program in Linguistics, Federal University of Ceará, Brazil,* leonel.de.alencar@ufc.br
[3] *School of Applied Mathematics of the Getúlio Vargas Foundation. Brazil*

Keywords: deep syntactic parsing, grammar engineering, computational morphology, evaluative suffixes

A comprehensive lexical database is an indispensable component of a linguistic-based computational grammar. This grammar type strives to mathematically model a fluent speaker's knowledge about the structures of their natural language. It has proven to be very useful in natural-language understanding tasks [6, 7]. For languages with a rich inflectional system and productive word-formation processes like Portuguese, one cannot simply limit oneself to the entries of a standard dictionary as a lexical knowledge source. It is mandatory to model the combinatorics of stems, inflectional, and derivational affixes, as well as the orthographic and morphophonological that apply to the output of these combinations [3]. In this talk, we report on our experience in developing PorGram [1], an open-source, free software computational grammar of Portuguese in the HPSG formalism, encoded in the Type Description Language (TDL) [5]. The construction of PorGram so far has combined the Grammar Matrix [4], which automatically generates TDL code from choices in a customization questionnaire, and hand-coding of TDL specifications. PorGram's ultimate goal is to parse unrestricted text. Therefore, it needs to incorporate a vast amount of lexical knowledge. We discuss the solutions implemented to handle inflectional and derivational morphology and the envisaged strategies to tackle problems that remain unsolved. These strategies draw on existing resources, such as MorphoBr [2], a large full-form lexicon, and the UD_Portuguese-Bosque treebank [8]. The focus of the presentation will be the implementation of evaluative suffixation. A hallmark of Portuguese morphology, it comprises diminutive, augmentative, and absolute superlative suffixes, which are extremely productive, potentially applicable to any noun and/or adjective [9, 10]. Since they involve intricate interactions between inflectional and derivational processes, their implementation is an ideal test case for assessing the strengths and weaknesses of a framework for computational morphology.

Bibliography

[1] L. F. de Alencar, L. R. Coutinho, W. J. L. da Silva, A. L. Nunes, A. Rademaker, *Extracting valences from a dependency treebank for populating the verb lexicon of a Portuguese HPSG Grammar*, in Vladia Pinheiro et al., editors, *Computational Processing of the Portuguese Language: PROPOR 2022*, Lecture Notes in Computer Science 13208, pp. 365-375, Springer, Cham, 2022.

[2] L. F. de Alencar, B. Cuconato, A. Rademaker, *Morphobr: An open source large-coverage full-form lexicon for morphological analysis of Portuguese.* Texto Livre: Linguagem e Tecnologia, 11, 3, 1-25, 2018.

[3] K. R. BEESLEY, L. KARTTUNEN, *Finite State Morphology*, CSLI, Stanford, 2003.

[4] E. M. BENDER, D. FLICKINGER, S. OEPEN, *The Grammar Matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars*, in J. CARROLL, N. OOSTDIJK, R. SUTCLIFFE, editors, *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, pp. 8-14, Taiwan, 2012.

[5] ANNE COPESTAKE, *Implementing typed feature structure grammars*, CSLI, Stanford, 2002.

[6] D. FERRUCCI ET AL., *Building Watson: An overview of the DeepQA project.* AI Magazine, 31, 3, 59-79, 2010.

[7] M. C. MCCORD, J. W. MURDOCK, B. K. BOGURAEV, *Deep parsing in Watson.* IBM Journal of Research and Development, 56, 3.4, 3:1-3:15, 2012.

[8] A. RADEMAKER ET AL., *Universal Dependencies for Portuguese*, in *Proceedings of the Fourth International Conference on Dependency Linguistics (Depling)*, pp. 197-206, Pisa, 2017.

[9] G. M. RIO-TORTO ET AL., *Gramática derivacional do português*, Coimbra University Press, Coimbra, 2016.

[10] A. VILLALVA, J. P. SILVESTRE, *Introdução ao estudo do léxico: descrição e análise do português*, Petrópolis, Vozes, 2014.