Lexicon-Grammar Web

Jorge Baptista[1,2], **David Antunes**[1], **Nuno Mamede**[1], **and Eugénio Ribeiro**[1,3]
[1] INESC-ID Lisboa, Portugal
[2] Faculdade de Ciências Humanas e Sociais, Universidade do Algarve, Portugal
[3] Instituto Universitário de Lisboa (ISCTE-IUL), Portugal
{jorge.baptista,david.f.l.antunes,nuno.mamede,eugenio.ribeiro}@inesc-id.pt

Abstract

This demo showcases a web-based interface that provides open, interactive access to a large-scale grammatical database of European Portuguese verbal constructions. Through a unified search and exploration environment, users can query, inspect, and compare more than 7,000 distributionally free verbal constructions and over 2,700 verbal idioms (frozen constructions), grounded in long-standing Lexicon–Grammar descriptions. For each construction, the interface exposes core linguistic properties such as argument structure, distributional constraints, semantic roles, major syntactic transformations, and curated usage examples with English translations. The demo illustrates how detailed, manually validated grammatical knowledge can be explored dynamically via the web, supporting linguistic research, language teaching, and NLP development. To the best of our knowledge, this is the largest publicly accessible, web-based grammatical resource dedicated to European Portuguese verbal constructions.

1 Overview: Web Interface and Resources

The resources presented in this demo originate from long-term research of European Portuguese within the Lexicon–Grammar framework (Gross, 1996), with a strong orientation toward computational processing. The description of distributionally free verbal constructions stems from the systematic development of a Lexicon–Grammar of EP verbs, leading to the creation of the VIPER database (Baptista, 2013). This work resulted in a large, manually curated inventory of verb constructions, later consolidated in the *Dicionário Gramatical de Verbos do Português Europeu* (Baptista and Mamede, 2020). The corresponding web-accessible resource is made available through the STRING NLP chain (Mamede et al., 2012)[1] and the PORTULAN/CLARIN research infrastructure[2].

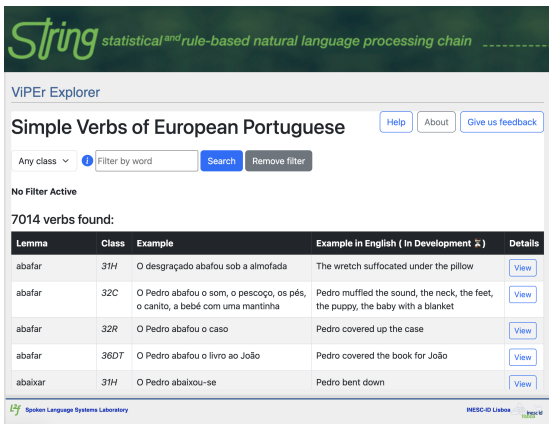
The VIPER resource encodes the +7,000 most frequently used EP verbal constructions. The information represented includes: (i) basic syntactic structure (e.g. impersonal, intransitive, transitive, ditransitive, and transitive–predicative patterns), covering more than 70 formal classes; (ii) explicit argument structure and distributional constraints, including human vs. non-human noun distinctions and over 70 semantic features corresponding to semantic prototypes (Bick, 2009); (iii) semantic roles selected from a set of approximately 50 highly reproducible constructs (Talhadas et al., 2013); (iv) constraints on subclause modality (indicative, subjunctive, factive, and *se*-interrogative) and coreference; and (v) major syntactic transformations, including several passive types, pronominalization and restructuring processes, symmetry constructions (Baptista, 2005), intrinsically reflexive constructions (e.g. *queixar-se* 'complain'), and *verba dicendi* constructions (Baptista, 2010). Each entry is illustrated with a canonical example and its English translation, complemented by manually curated corpus and web-based examples. Fig. 1 illustrates (a) the Search view of the VIPER and (b) the entry view of verb *agradar* 'please' (class 05).

The VIPER database has supported several NLP tools and linguistic studies integrated into the STRING system, including work on EP verb sense disambiguation (Pires, 2016), a survey on EP communication predicates (Reis et al., 2021), a study on transitive–predicative constructions (Baptista, 2021), and contrastive studies of locative verbs in European and Brazilian Portuguese (Rodrigues et al., 2015). It also serves as the catalogue of verbal senses for a lexicalized Abstract Meaning Representation initiative (Baptista et al., 2024). Together, these studies highlight the relevance of de-

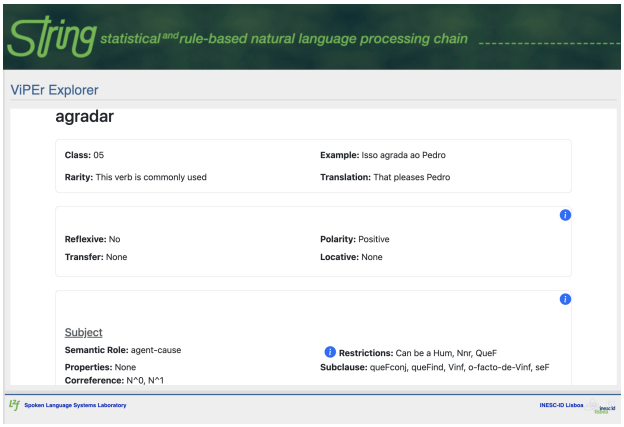
[1] https://string.hlt.inesc-id.pt/

[2] https://hdl.handle.net/21.11129/0000-000D-F91E-A

Proceedings of the 17th International Conference on Computational Processing of Portuguese (PROPOR 2026)
Vol. 2, pages 8–10
April 13-16, 2026



(a) Search view



(b) Entry view

Figure 1: VIPER web interface.

tailed lexical–syntactic descriptions for both theoretical investigation and applied NLP research, now made openly accessible through the web interface presented in this demo.

While grammatical descriptions of verbal constructions exist for both major varieties of Portuguese, many are available only in hard-copy form (Busse, 1994; Cançado et al., 2013) or as downloadable datasets primarily targeting Brazilian Portuguese, such as VERBNET.BR (Scarton, 2011) or VaLexPB[3]. In contrast, few resources for European Portuguese are manually curated, natively developed for EP, and provide comparably fine-grained linguistic descriptions in a web-accessible format.

In parallel, a second resource (VIDIOM) focusing on +2,700 verbal idioms has been developed. Verbal idioms are elementary sentences in which the verb and at least one of its arguments are distributionally frozen and the overall meaning is often non-compositional, that is, it is different from the meanings of the individual elements taken in isolation (e.g., *bater a bota*, lit. 'beat the boot', meaning 'to kick the bucket / die'). Fig. 2 shows the (a) the Search view, for a query with verb *bater* 'beat', and (b) the Entry view of the verbal idiom *bater a bota* 'die'. This work builds upon earlier, formal linguistic descriptions of frozen sentences in Portuguese, and has been progressively extended to support computational processing (Baptista et al., 2014). Recent developments include the integration of verbal idioms into rule-based parsers (Galvão et al., 2019), the creation of a corpus annotated for verbal idioms (Antunes et al., 2025b)[4], and dedicated methods for their automatic processing (Antunes et al., 2025a).

2 System Design

The system is implemented as a lightweight web application with a clear separation between data storage, backend logic, and user interface. The data layer relies on a SQLite relational database populated from structured spreadsheet sources and accessed through SQLAlchemy. The backend is implemented in Python using Flask, with modular routing via blueprints and server-side rendering through Jinja2 templates. The frontend consists of server-rendered HTML enhanced with Bootstrap components, enabling responsive layouts and interactive elements without reliance on a heavy client-side framework.

3 What the demo shows

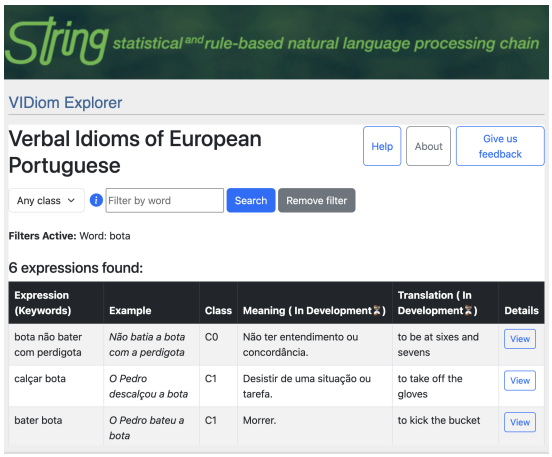
The demo provides interactive access to the Lexicon–Grammar resources via a unified web interface. Users can search and browse verbal constructions, examine their syntactic and semantic properties, and compare distributionally free constructions with verbal idioms, illustrated by curated examples and English translations to support analysis.

Acknowledgments

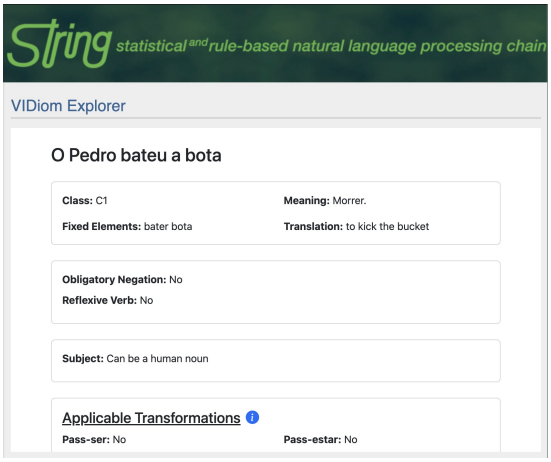
Work supported by national funds through Fundação para a Ciência e a Tecnologia, I.P. (FCT) (DOI: https://doi.org/10.54499/UID/50021/2025 and https://doi.org/10.54499/UID/PRR/50021/2025).

[3] https://github.com/jessemourao/VaLexPB/

[4] https://portulanclarin.net/repository/browse/vidiom-pt/



(a) Search view



(b) Entry view

Figure 2: VIDIOM web interface.

References

David Antunes, Jorge Baptista, and Nuno Mamede. 2025a. Processamento automático de expressões idiomáticas do português europeu. *Linguamática*, 17(1):55–73.

David Antunes, Jorge Baptista, and Nuno J. Mamede. 2025b. A European Portuguese corpus annotated for verbal idioms. In *Workshop on Multiword Expressions (MWE 2025)*, pages 58–66.

Jorge Baptista. 2005. Construções simétricas. In *Estudos de homenagem a Mário Vilela*, pages 353–367. FLUP.

Jorge Baptista. 2010. *Verba dicendi*: A structure looking for verbs. In *Les Tables. La grammaire du français par le menu. Mélanges en hommage à Christian Leclère*, pages 11–20. CENTAL/Presses Universitaires de Louvain.

Jorge Baptista. 2013. ViPEr: uma base de dados de construções léxico-sintáticas de verbos do Português europeu. In *Actas do XXVIII Encontro da APL - Textos Selecionados*, pages 111–129, Lisboa.

Jorge Baptista. 2021. Construções de verbos transitivos-predicativos em português: uma perspetiva transformacional. *Rev. da Assoc. Port. Linguística*, 8:10–25.

Jorge Baptista and Nuno Mamede. 2020. *Dicionário Gramatical de Verbos do Português Europeu*. U.Algarve Editora.

Jorge Baptista, Nuno Mamede, and Ilia Markov. 2014. Integrating verbal idioms into an NLP system. In *PROPOR 2014*, pages 250–255. Springer Verlag.

Jorge Baptista, Sónia Reis, João Dias, and Pedro Santos. 2024. Lexicalized Meaning Representation (LMR). In *5th Int. Workshop on Designing Meaning Representations @ LREC-COLING 2024*, pages 101–111.

Eckhard Bick. 2009. Semantic prototype tags for nouns. Online technical description.

Winfried Busse. 1994. *Dicionário Sintáctico de Verbos*. Almedina, Coimbra.

Márcia Cançado, L. Godoy, and L. Amaral. 2013. *Catálogo de Verbos do Português Brasileiro. Verbos de Mudança*. Editora UFMG.

Ana Galvão, Jorge Baptista, and Nuno J. Mamede. 2019. Processing European Portuguese Verbal Idioms: From the Lexicon-Grammar to a Rule-based Parser. In *3rd Int. Conf. Computational and Corpus-based Phraseology (EUROPHRAS 2019)*, pages 70–77.

Maurice Gross. 1996. Lexicon-grammar. In *Concise Encyclopedia of Syntactic Theories*, pages 244–259. Pergamon.

Nuno Mamede, Jorge Baptista, Cláudio Diniz, and Vera Cabarrão. 2012. STRING - A Hybrid Statistical and Rule-Based Natural Language Processing Chain for Portuguese. In *PROPOR 2012*, volume Demo Session, Coimbra, Portugal. PROPOR.

Ricardo Pires. 2016. Verb Sense Disambiguation in STRING. Master's thesis, Instituto Superior Técnico, Univ. Lisboa.

Sónia Reis, Nuno Mamede, and Jorge Baptista. 2021. Predicados de Comunicação em Português Europeu. *Rev. da APL*, 8:237–259.

Roana Rodrigues, Jorge Baptista, and Oto Araújo Vale. 2015. Análise contrastiva dos verbos locativos. In *Symposium in Information and Human Language Technology & IV Jornada de Descrição do Português*, pages 233–240.

Carolina Scarton. 2011. VerbNet.Br: construção de um léxico de verbos. In *Brazilian Symposium in Information and Human Language Technology*.

Rui Talhadas, Nuno Mamede, and Jorge Baptista. 2013. Semantic Roles for Portuguese Verbs. In *Int. Conf. on Lexis and Grammar*, pages 127–132.