

VerbNet.Br 1.0 – User guide

VerbNet.Br (Scarton 2011; Scarton and Aluisio, 2012; Scarton, 2013; Scarton et al., to appear) is a lexical resource based on VerbNet (Kipper, 2005). VerbNet deals with the syntactic-semantic interface of verbs to group them into classes. It is based on Levin's verb classes (Levin, 1993) and follows the hypothesis that verbs that share syntactic behavior should present argument structures with the same semantic roles. For example, the verbs “to spray” and “to load” in sentences (1) and (2) illustrate the locative alternation (extracted from (Levin, 1993) , p. 2)):

1. a. Sharon sprayed water on the plants.
b. Sharon sprayed the plants with water.
2. a. The farmer loaded apples into the cart.
b. The farmer loaded the cart with apples.

In (1) and (2), the syntactic patterns (or subcategorization frames - SCFs) of (1.a) and (2.a) are similar: NP V NP PP, in which NP is a noun phrase, V is the verb and PP is a prepositional phrase with preposition “on” in (1.a) and “into” in (2.a) (despite the fact that the prepositions are different, they both introduced a place). The same happens for (1.b) and (2.b) in which the common SCF is NP V NP PP (the preposition “with” appeared in both cases).

Each class in VerbNet is described by its members (verbs that belongs to the class); by the thematic roles (roles that the verb arguments can assume, e.g. “Agent”, “Theme”); by the selectional restrictions (restrictions for the thematic roles, e.g. “+animate”, “-region”); by the syntactic frames (SCFs), and by semantic predicates, that are related to the SCFs, and describe the event that the verb members perform. VerbNet also has mappings to other computational lexical resource (CLRs) like WordNet, FrameNet and PropBank (Loper et al., 2007).

To build the VerbNet.Br, three existing CLRs are used: VerbNet, WordNet (Fellbaum, 1998) and WordNet.Br (Dias-da-Silva, 2010). From VerbNet, the classes and theory were extended to VerbNet.Br. From WordNet.Br, the BP verbs were selected as candidate members of VerbNet.Br. WordNet was only used to link VerbNet and WordNet.Br.

Our method is divided in four stages. Stage 1 consisted of the manual translation of SCFs in VerbNet from English into Portuguese. It is worth emphasizing that we did not translate the examples and only validate those syntactic structures which are fully acceptable in Portuguese, or requires only the substitution of the preposition of the preposition phrase (PP) structures. Stage 2 consisted of the automatic search of SCFs in Brazilian Portuguese corpora. This search was done by using Zanette et al. (2012) SCF extraction tool. Stage 3 used the links among VerbNet, WordNet and WordNet.Br to select candidate verbs that fit into VerbNet.Br classes. Finally, stage 4 validated the candidate members defined in stage 3 based on the information of stages 1 and 2. For a detailed explanation of all stages and also evaluation results, refer to Scarton (2011), Scarton and Aluisio (2012), Scarton (2013) and Scarton et al. (to appear).

1. Instalation Prerequisites

MySQL server is a prerequisite tool for using the database of VerbNet.Br. The installation guidelines of MySQL can be found here:

<http://dev.mysql.com/doc/refman/5.1/en/installing.html>.

This user guide will show how to use VerbNet.Br database considering MySQL is already installed and it is configured in the PATH variable. We will show command lines that can be used in Linux. They should work in Windows/MacOS similarly.

SQL knowledge is also required. For an introduction on SQL refer to:
<http://www.w3schools.com/sql/default.asp>.

2. Creating the *verbnetbr* database with the VerbNet.Br 1.0 data

First, you need to download the SQL script of VerbNet.Br (*verbnet.sql*). After that, open a terminal, go to the directory where the script was downloaded and type:

```
$ mysql -u <mysql_user> -p < verbnetbr.sql
```

This will create a database in your local MySQL server called *verbnetbr* with three tables.

3. Using the *verbnetbr* database

a. Table *candidatos*

Table *candidatos* is the main table of VerbNet.Br database. It contains 10 columns:

- *verbo*: contains the verbs of VerbNet.Br
- *classe*: stores the classes for each verb
- *vingles*: defines which English verb was responsible to bring the Portuguese verb into the database
- *ili*: defines the ILI where the verb appears in WordNet.Br
- *alinhamento*: the kind of alignment between the synsets of WordNet and WordNet.Br
- *exp1*: binary attribute that defines whether a given verb was selected to figure in VerbNet.Br, according to EXP1 of Scarton et al. (to appear) and Scarton (2013)
- *exp2*: binary attribute that defines whether a given verb was selected to figure in VerbNet.Br, according to EXP2 of Scarton et al. (to appear) and Scarton (2013)
- *exp3*: binary attribute that defines whether a given verb was selected to figure in VerbNet.Br, according to EXP3 of Scarton et al. (to appear) and Scarton (2013)
- *exp4*: binary attribute that defines whether a given verb was selected to figure in VerbNet.Br, according to EXP4 of Scarton et al. (to appear) and Scarton (2013)
- *freq*: frequency of a verb in three corpora defined in Scarton et al. (2014), Scarton et al. (to appear) and Scarton (2013)

In order to recover all the verbs that figure in VerbNet.Br according to EXP1, one can type on mysql terminal:

```
mysql> select * from candidatos where exp1=1;
```

Results of EXP2, EXP3 and EXP4 can be recovered similarly. To recover results for EXP0, consider frequency higher than 10:

```
mysql> select * from candidatos where freq>10;
```

b. Table *classes*

This table contains information about the classes of VerbNet.Br and the translated SCFs (Stage 1). It has 7 columns:

- *classe*: class of VerbNet.Br
- *alternancia*: SCF parameterized by prepositions
- *alternancia_pure*: SCF non-parameterized by prepositions
- *ordem*: number of SCFs for a given class (SCFs with preposition variation and same meaning will have the same number)
- *pingles*: link with English SCF
- *singles*: link with English SCF
- *alter_ingles*: number of SCFs in the English VerbNet

c. Table *papeis*

This table contains the Thematic Roles of each class. There are 2 columns:

- *classe*: class of VerbNet.Br
- *papel*: Thematic Roles expected for each class (inherited automatically from VerbNet)

License

Please, read the license document before use VerbNet.Br. This document can be download here:

http://www.nilc.icmc.usp.br/portlex/images/arquivos/verbnnetbr/verbnnet.br_license.pdf

Contact

Carolina Scarton: carol@icmc.usp.br

How to cite

To cite VerbNet.Br, please, refer the following references:

Scarton, C. (2011): VerbNet.Br: construção semiautomática de um léxico computacional de verbos para o português do Brasil. In the Proceedins of 8th Brazilian Symposium in Information and Human Language Technology (STIL 2011), Cuiabá-MT, Brazil.

Scarton, C. and Aluisio, S. (2012): Towards a cross-linguistic VerbNet-style lexicon to Brazilian Portuguese. In Proceedings of the LREC 2012 Workshop on Creating Cross-language Resources for Disconnected Languages and Styles (CREDISLAS 2012). Istambul, Turquia, pp 11-18.

Scarton, C. E. (2013). VerbNet.Br: construção semiautomática de um léxico verbal online e independente de domínio para o português do Brasil. Master's dissertation, Instituto de Ciências Matemáticas e de Computação, University of São Paulo, São Carlos. Accessed in 2014-09-29, at <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-19042013-160640/>

Scarton, C., Duran, M. S. and Aluisio, S. M. (to appear): Using Cross-linguistic Knowledge to Build VerbNet-style Lexicons: Results for a (Brazilian) Portuguese VerbNet. Accepted to appear in the proceedings of the International Conference on Computational Processing of Portuguese (PROPOR 2014), October 6-9, Sao Carlos-SP, Brazil

Acknowledgements

This work was supported by FAPESP/Brazil (No. 2010/03785-0 and No. 2011/22882-0). We also thank to EXPERT (EU Marie Curie ITN No. 317471) project.

References

Dias-da-Silva, B. C. (2010): Brazilian Portuguese WordNet: A Computational Linguistic Exercise of Encoding Bilingual Relational Lexicons. *International Journal of Computational Linguistics and Applications*, New Delhi, v.1, n. 1-2, p.137 - 150.

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. MIT Press. Cambridge, Massachusetts.

Kipper, K. (2005): *Verbnet: A broad coverage, comprehensive verb lexicon*. 146f. Ph.D. Thesis (Philosophy) - University of Pennsylvania, USA.

Levin. B. (1993). *English Verb Classes and Alternation, A Preliminary Investigation*. The University of Chicago Press.

Loper, E., Yi, S-T and Palmer, M. (2007): Combining Lexical Resources: Mapping between PropBank and VerbNet. In *Proceedings of the 7th International Workshop on Computational Linguistics*. Tilburg, Países Baixos.

Scarton, C. (2011): VerbNet.Br: construção semiautomática de um léxico computacional de verbos para o português do Brasil. In the Proceedings of 8th Brazilian Symposium in Information and Human Language Technology (STIL 2011), Cuiabá-MT, Brazil.

Scarton, C. and Aluisio, S. (2012): Towards a cross-linguistic VerbNet-style lexicon to Brazilian Portuguese. In *Proceedings of the LREC 2012 Workshop on Creating Cross-language Resources for Disconnected Languages and Styles (CREDISLAS 2012)*. Istambul, Turquia, pp 11-18.

Scarton, C. E. (2013). VerbNet.Br: construção semiautomática de um léxico verbal online e independente de domínio para o português do Brasil. Master's dissertation, Instituto de Ciências Matemáticas e de Computação, University of São Paulo, São Carlos. Accessed in 2014-09-29, at <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-19042013-160640/>

Scarton, C., Sun, L., Kipper-Schuler, K., Duran, M., Palmer, M. and Korhonen, A. (2014): Verb Clustering for Brazilian Portuguese. In the Proceedings of 15th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2014), Katmandu, Nepal, pp. 25-39.

Scarton, C., Duran, M. S. and Aluisio, S. M. (to appear): Using Cross-linguistic Knowledge to Build VerbNet-styleLexicons: Results for a (Brazilian) Portuguese VerbNet. Accepted to appear in the proceedings of the International Conference on Computational Processing of Portuguese (PROPOR 2014), October 6-9, Sao Carlos-SP, Brazil.

Zanette, A., Scarton, C. and Zilio, L. (2012): Automatic extraction of subcategorization frames from corpora: an approach to Portuguese. In *Demonstration Session of the International Conference on Computational Processing of Portuguese Language (PROPOR 2012)*. Coimbra, Portugal.